Energy-Efficient Trajectory Optimization for Aerial Video Surveillance under QoS Constraints

Cheng Zhan*, Han Hu†, Shiwen Mao‡, Jing Wang§
*School of Computer and Information Science, Southwest University, Chongqing, China
†School of Information and Electronics, Beijing Institute of Technology, Beijing, China
‡Department of Electrical and Computer Engineering, Auburn University, Auburn, USA
§Capital Normal University, Beijing, China
Email: zhanc@swu.edu.cn, hhu@bit.edu.cn, smao@ieee.org, jwang_ruc@163.com

***Abstract*—Surveillance drones are unmanned aerial vehicles (UAVs) that are utilized to collect video recordings of targets. In this paper, we propose a novel design framework for aerial video surveillance in urban areas, where a cellular-connected UAV captures and transmits videos to the cellular network that services users. Fundamental challenges arise due to the limited onboard energy and quality of service (QoS) requirements over environment-dependent air-to-ground cellular links, where UAVs are usually served by the sidelobes of base stations (BSs). We aim to minimize the energy consumption of the UAV by jointly optimizing the mission completion time and UAV trajectory as well as transmission scheduling and association, subject to QoS constraints. The problem is formulated as a mixed-integer non-linear programming (MINLP) problem by taking into account building blockage and BS antenna patterns. We first consider the average performance for uncertain local environments, and obtain an efficient sub-optimal solution by employing graph theory and convex optimization techniques. Next, we investigate the site-specific performance for specific urban local environments. By reformulating the problem as a Markov decision process (MDP), a deep reinforcement learning (DRL) algorithm is proposed by employing a dueling deep Q-network (DQN) neural network model with only local observations of sampled rate measurements. Simulation results show that the proposed solutions achieve significant performance gains over baseline schemes.**

***Index Terms*—Unmanned aerial vehicle (UAV), video surveillance, quality of service (QoS), energy consumption, deep reinforcement learning (DRL).**

I. Introduction

With the advantages of low cost and high mobility, unmanned aerial vehicles (UAVs), a.k.a. drones, have experienced a rapid evolution from primarily military applications to various civilian applications, such as food delivery, emergency response, and aerial monitoring [1], [2]. Equipped with cameras, UAVs have also been integrated into aerial surveillance systems to track and monitor various targets, such as people, vehicles, or specific areas. Different from the conventional ground video surveillance which employs cameras installed at pre-configured, fixed locations, aerial surveillance can drastically improve the surveillance performance since the UAV with high mobility makes it possible to flexibly avoid any blockage in the view path, and even to actively track the surveillance targets [3], [4]. To enable the high-performance aerial surveillance, air-to-ground (A2G) communication is a crucial challenge, which requires ultra reliable low latency communication for video streaming. Conventional UAV-assisted wireless communications, e.g., aerial base stations (BSs), simply rely on point-to-point, short-range communications over unlicensed spectrum [5], [6], which are not sufficient for beyond visual line-of-sight (BVLoS) communications, especially for aerial surveillance which requires wide-area connectivity. To tackle such issue, cellular-connected UAV communications can be adopted which utilizes the existing cellular network to provide reliable control and high-throughput communications [7]. In this case, surveillance drones are treated as *aerial users* and high-capacity backhaul links in cellular networks are exploited to provide BVLoS communications and performance enhancement in terms of reliability, security and throughput.

However, there still exists several fundamental challenges for aerial surveillance in urban area with cellular-connected UAVs. First, ground BSs usually tilt their antennas downward to enhance the desired signal of terrestrial users. Consequently, surveillance drones are usually served by the BS's sidelobes and therefore might suffer significant performance degradation [8], [9]. Second, the blockage by buildings in urban areas depends on the local environment where the UAVs are actually deployed, which leads to environment-dependent A2G channels. Due to the dynamic and environment-dependent A2G channels, how to guarantee the quality of service (QoS) becomes a challenging task as video surveillance requires low latency and low jitter [10]. Another challenge for aerial surveillance is the limited onboard energy. Different from pre-configured terrestrial cameras, UAV incurs additional propulsion energy consumption to support its movement and hovering [11], [12], which should be efficiently utilized to extend the endurance of surveillance. Thus, the joint design of UAV trajectory and transmission scheduling for aerial surveillance with cellular-connected UAV should balance the tradeoff between QoS provisioning and energy consumption, which is the main focus of our paper.

In this paper, we propose a cellular-connected UAV enabled video surveillance design framework in the urban area. More flexible aerial surveillance services to the wide area can be provided by dynamically adjusting the UAV's position and UAV-BS transmission scheduling and association. Our aim

978-1-6654-5822-1/22/$31.00 ©2022 IEEE

is to minimize the energy consumption of the UAV through jointly optimizing the mission completion time and UAV trajectory as well as transmission scheduling and association, while the QoS requirements for video streaming should be satisfied in the urban area with realistic channel models. The main contributions of this paper are summarized as follows.

- First, a novel design framework is proposed for aerial surveillance with cellular-connected UAV capturing and transmitting videos to the cellular network. Realistic A2G propagation channels in urban environments are taken into account. A mixed-integer non-linear programming (MINLP) optimization problem is formulated to minimize the energy consumption of the UAV with QoS requirements, which is difficult to tackle due to the environment-dependent channel.
- Second, we consider the average performance for uncertain local environments and analyze the optimal structure, based on which we reformulate the problem and examine its feasibility by employing a graph theory based method. An efficient iterative algorithm is proposed by leveraging alternating optimization and Golden-section (GS) search as well as successive convex approximation (SCA) techniques, which converge to a solution satisfying the Karush-Kuhn-Tucker (KKT) conditions.
- Third, we investigate the site-specific performance for a specific local environment. By reformulating the problem as a Markov decision process (MDP), a deep reinforcement learning (DRL) algorithm is proposed with a dueling deep Q-network (DQN) model, where the UAV intelligently makes the most advantageous decisions of flying and transmission scheduling by leveraging only the local observations of sampled rate measurements.
- Finally, extensive simulations are conducted to evaluate the effectiveness of our proposed algorithms, and results show that the proposed algorithms outperform several benchmark schemes, where the trade-off between energy consumption and QoS provisioning can be balanced.

II. Related Work

An important component of the beyond 5G and 6G networks is the use of UAVs, and we have seen some promising research results focusing on UAV-enabled video streaming. The work in [6] and [13] studied multiuser downlink adaptive video streaming with UAV BS and UAV relay over fading channels, where video streaming is adjusted dynamically based on the transmission rate. In [14], the authors proposed a UAV-enabled pseudo-analog wireless video broadcast system to maximize the minimum peak signal-to-noise ratio of ground users' video reconstruction quality. Joint resource allocation and trajectory design for UAV-enabled video streaming was studied in [15] to maximize the video utility and minimize the UAV operation time. Machine learning methods were employed in [16] and [17] to support UAV-enabled omnidirectional video streaming and provide UAV anti-jamming video services to combat jamming attacks. Trajectory planning was investigated in [18] for a UAV conducting covert video surveillance, where visual covertness and energy efficiency were taken into account. A solar-powered UAV was employed in [19] for covert video surveillance by the joint design of power management and trajectory optimization. The work in [4] proposed a multi-set space-time shift keying for UAV-aided video surveillance, where throughput can be improved with near-flawless video quality. In summary, the above works mainly studied UAV-assisted wireless communications where UAVs were employed as aerial communication platforms to assist terrestrial networks. However, these works did not consider cellular-connected UAVs, where UAVs are treated as aerial users of existing cellular networks.

Cellular-connected UAV supports BVLoS communications, which is more suitable for aerial surveillance that requires wide-area connectivity. The use of coordinated multi-point transmission for providing seamless connectivity to aerial users was investigated in [20]. The work in [21] proposed a cooperative beamforming technique to mitigate the co-channel terrestrial interference transmissions to a cellular-connected UAV. In [22], the authors studied the uplink transmission from a UAV to cellular BSs, where the trade-off between inter-cell interference coordination and air-ground performance are balanced. The work in [23] maximized the achievable rate for cellular-connected multi-UAV communications by jointly designing user association, beamforming, and UAV-height control. A radio map based path planning framework was proposed in [24] for a cellular-connected UAV to minimize its flying distance with signal-to-interference-plus-noise ratio (SINR) requirement. Power-domain aerial-terrestrial non-orthogonal multiple access (NOMA) was employed in [25] and [26] to enable the concurrent uplink transmissions of both aerial users and terrestrial users. However, the above works relied on the assumption that the UAV has enough energy, which is impractical due to the limited energy budget.

III. System Model and Problem Statement

Consider a UAV-aided video surveillance system, where a rotary-wing UAV equipped with a camera is employed to capture and transmit videos from an initial location to a destination. For example, the initial location may correspond to the target point for surveillance while the final location may correspond to the charging station. Through A2G links, the video captured is transmitted to the cellular network, which consists of a set $\mathcal{M} = \{s_1, s_2, \ldots, s_M\}$ of M terrestrial BSs. The cellular network will then stream the video contents to the users via high capacity wireless backhaul links. Dynamic adaptive streaming over HTTP (DASH) is adopted for the video streaming [27], [28]. Note that the design framework can be extended to the scenarios with multiple UAVs with different initial/final locations. For ease of exposition, we assume that all the terrestrial BSs have the same altitude of H_G, and the UAV flies at a fixed altitude $H_U \gg H_G$, since frequent aircraft ascending and descending are energy-inefficient. The horizontal coordinate of the BS s_m is denoted by $\mathbf{g}_m \in \mathbb{R}^{2\times 1}$. The UAV's trajectory projected onto the horizontal plane is denoted as $\{\mathbf{u}(t)|\mathbf{u}(t) \in \mathbb{R}^{2\times 1}, t \in [0, T]\}$, with T denoting

the mission completion time, which is a design variable in this paper. The instantaneous velocity of the UAV at time instant t is denoted by $\mathbf{v}(t) \triangleq \dot{\mathbf{u}}(t)$, which is constrained by the maximum speed $V_{\max}$, i.e., $\|\mathbf{v}(t)\| \leq V_{\max}, \forall t \in [0,T]$. The distance between the UAV and BS s_m at time t is given as $d_m(t) = \sqrt{(H_U - H_G)^2 + \|\mathbf{u}(t) - \mathbf{g}_m\|^2}$.

A. Channel Model

Denote $h_m(t)$ as the baseband equivalent channel between the UAV and s_m at time t, which depends on the BS antenna model, large-scale fading effects, and small-scale fading. The BS antenna model follows the current LTE standard [29], where each BS is equipped with a uniform linear array (ULA) consisting of $N_0 = 8$ antenna elements with half-wavelength spacings, which are tilted downwards with elevation angle $\theta_D = 10°$. The overall antenna gain between the UAV and s_m is specified in [30], i.e., $G_m(t) = \frac{(10^{\frac{G_e}{10}})\sin^2\left(\frac{N_0\pi}{2}(\sin\theta_m(t) - \sin\theta_D)\right)}{N_0\sin^2\left(\frac{\pi}{2}(\sin\theta_m(t) - \sin\theta_D)\right)}$, where $G_e \triangleq -\min\left(12\left(\frac{\theta_m(t)}{HPBW_v}\right)^2, G_0\right)$ is the element power gain, with $\theta_m(t) \triangleq \frac{180}{\pi}\sin^{-1}(\frac{H_G - H_U}{d_m(t)})$ denoting the elevation angle at time t, G_0 denotes the antenna nulls threshold, and $HPBW_v$ is the half power beamwidth.

The large-scale channel coefficient between the UAV and BS s_m at time t can be modeled by $\beta_m(t) = \beta_0 d_m(t)^{-\alpha}$ for the line-of-sight (LoS) link and $\beta_m(t) = \mu\beta_0 d_m(t)^{-\alpha}$ for a non-LoS (NLoS) link, where β_0 is the channel power gain at 1 m, $\alpha \geq 2$ denotes the path loss exponent, $\mu < 1$ represents the additional signal attenuation factor due to NLoS propagation. The LoS probability between the UAV and s_m at time t is given by $\mathbb{P}_m^L(t) = \frac{1}{1 + a\exp(-b[|\theta_m(t)| - a])}$, where a and b are constant parameters depending on the propagation environment [31]. As a result, we have

$$h_m(t) = \sqrt{G_m(t)\beta_m(t)}\tilde{h}_m(t), \tag{1}$$

where $\tilde{h}_m(t)$ is a random variable with $\mathbb{E}[|\tilde{h}_m(t)|^2] = 1$ accounting for the small-scale fading. In general, $\tilde{h}_m(t)$ can be modeled as Rician fading with Rician factor K_c for LoS links, and Rayleigh fading for NLoS links.

B. Video Streaming Model

Denote a binary variable $z_m(t) \in \{0,1\}$ as the transmission scheduling indicator at time t, where $z_m(t) = 1$ indicates that the UAV transmits video to s_m and $z_m(t) = 0$ otherwise. At each time t, we assume that only one BS is scheduled for transmission, then we have $\sum_{m=1}^{M} z_m(t) \leq 1, t \in [0,T]$. Note that $z_m(t)$ indicates not only the transmission scheduling but also the association between the UAV and ground BSs at time t. If scheduled, the maximum achievable rate from the UAV to s_m at time t, denoted by $R_m(t)$, is given by

$$R_m(t) = B\log_2\left(1 + \frac{P|h_m(t)|^2}{\sigma^2}\right), \tag{2}$$

where B denotes bandwidth, σ^2 denotes noise power, and P corresponds to the maximum transmit power of the UAV.

We consider the video streaming between the surveillance UAV and cellular network, which requires QoS aware communications between the UAV and ground BSs. Let $R_{\min}$ denote the minimum required data rate for video streaming, then we have $\sum_{m=1}^{M} z_m(t)R_m(t) \geq R_{\min}$, for all t. For example, $R_{\min}$ may correspond to the data rate that the video users demand. However, the quality of A2G channels is highly susceptible to the channel randomness due to the random small-scale fading and random occurrence of LoS and NLoS links, and a *rate outage event* may occur at time t if $\sum_{m=1}^{M} z_m(t)R_m(t) < R_{\min}$, where the video streaming may be interrupted. To tackle such issue, for each time instant $t \in [0,T]$, we define function $L(t)$ as the latest time that no rate outage event occurs, i.e.,

$$L(t) \triangleq \max_{\tau}\{\tau | \sum_{m=1}^{M} z_m(\tau)R_m(\tau) \geq R_{\min}, \tau \in [0,t]\}, \tag{3}$$

where $L(t) \leq t$. If $L(t) = t$, then there is no rate outage up to time t. If $L(t) < t$, then rate outage occurs from time $L(t)$ to t with a finite duration $t - L(t)$, which is defined as the *rate outage duration* at time t. Note that if the rate outage duration is too long for video streaming, then the playout buffers at video users will be empty and video stall (i.e., rebuffering) occurs, when the video playing will pause until enough video chunks are fully downloaded. In this case, the video streaming is interrupted and the experience of video users will degrade significantly. As a result, the QoS aware communications for aerial video surveillance between the UAV and ground BSs should be designed to ensure that the maximum rate outage duration is no larger than a tolerable threshold δ, i.e., $\max_{t\in[0,T]}(t - L(t)) \leq \delta$. The selection of δ is based on the size and management strategy of users' playout buffers. In practice, the aerial surveillance system can provide more QoS provisioning for video users with larger $R_{\min}$ and smaller δ.

C. Problem Formulation

The energy consumption of the UAV consists of both communication-related energy and propulsion energy. In general, the communication-related energy is smaller than propulsion energy by two orders of magnitude [11], which is thus ignored in this paper. The propulsion energy consumption of the UAV can be calculated by $E_u = \int_{t=0}^{T} P_u(t)dt$, where the propulsion power $P_u(t)$ of the rotary-wing UAV at time t is approximated as in [12], i.e.,

$$P_u(t) \approx P_0\left(1 + \frac{3\|\mathbf{v}(t)\|^2}{U_{tip}^2}\right) + \frac{1}{2}d_0\rho s A\|\mathbf{v}(t)\|^3 + P_i\left(\sqrt{1 + \frac{\|\mathbf{v}(t)\|^4}{4o^4}} - \frac{\|\mathbf{v}(t)\|^2}{2o^2}\right)^{1/2}. \tag{4}$$

In (4), P_0 and P_i represent blade profile and induced power in hovering status, A and s denote the disc area and rotor solidity, respectively, d_0 is the fuselage drag ratio, ρ denote the air density, o denotes the mean rotor induced velocity in hover, while U_{tip} denotes the tip speed of the rotor blade.

In this paper, we aim to minimize the energy consumption of the UAV by optimizing the mission completion time T and UAV trajectory $\{\mathbf{u}(t)\}$ as well as transmission scheduling and association $\{z_k(t)\}$, subject to the UAV's mechanical constraints and rate outage duration constraint for video streaming. The optimization problem can be written as

$$\begin{aligned}
(\text{P1}): \quad & \min_{T,\{\mathbf{u}(t)\},\{z_m(t)\}} \int_{t=0}^{T} P_u(t)dt \\
\text{s.t.} \quad & z_m(t) \in \{0,1\}, \forall m, t, && (5)\\
& \sum_{m=1}^{M} z_m(t) \le 1, \forall t, && (6)\\
& \max_{t\in[0,T]} \left(t - L(t)\right) \le \delta, && (7)\\
& \|\mathbf{v}(t)\| \le V_{\max}, \forall t, && (8)\\
& \mathbf{u}(0) = \mathbf{u}_I, \mathbf{u}(T) = \mathbf{u}_F, && (9)
\end{aligned}$$

where $\mathbf{u}_I, \mathbf{u}_F \in \mathbb{R}^{2\times 1}$ are the initial/final locations of the UAV, respectively, and the values of $R_{\min}$ and δ are predetermined based on the specific QoS requirement. Note that problem (P1) is difficult to solve directly since the optimization variables $\{\mathbf{u}(t)\}$ and $\{z_m(t)\}$ are continuous time series. In addition, (P1) is an MINLP problem due to the binary constraints in (5) and the complicated objective function, which contains an integral whose upper limit is also a design variable T. Furthermore, the left-hand-side (LHS) of (7) contains complicated expression of $R_m(t)$ which is a random variable due to the randomness of $h_m(t)$.

IV. Graph based Solution

In this section, we consider the approximate average communication performance for uncertain local environments.

A. Problem Reformulation

Note that $h_m(t)$ is a random variable, we first focus on the expected communication rate as in [32], denoted by $\mathbb{E}[R_m(t)]$. Since $\mathbb{E}[|\tilde{h}_m(t)|^2] = 1$, we have $\mathbb{E}[|h_m(t)|^2] = \mathbb{P}_m^L(t)\beta_0 d_m(t)^{-\alpha} G_m(t) + (1-\mathbb{P}_m^L(t))\mu\beta_0 d_m(t)^{-\alpha} G_m(t)$. By applying the Jensen's inequality, we have

$$\begin{aligned}
\mathbb{E}[R_m(t)] &\le B\log_2\left(1 + \frac{P\mathbb{E}[|h_m(t)|^2]}{\sigma^2}\right) \\
&= B\log_2\left(1 + \frac{P\beta_0 \hat{P}_m^L(t) G_m(t)}{\sigma^2\left((H_U - H_G)^2 + \|\mathbf{u}(t) - \mathbf{g}_m\|^2\right)^{\alpha/2}}\right) \quad (10)
\end{aligned}$$

where $\hat{P}_m^L(t) \triangleq \mathbb{P}_m^L(t) + (1 - \mathbb{P}_m^L(t))\mu$ can be regarded as a regularized LoS probability. However, the expression in (10) is still difficult to handle since $G_m(t)$ and $\hat{P}_m^L(t)$ are complicated functions with respect to elevation angle $\theta_m(t)$, which depends on the UAV trajectory $\mathbf{u}(t)$. To tackle such difficulty, we employ the homogeneous approximation approach as in [11]. Specifically, we let $G_m(t) \approx \bar{G}_m(t), \hat{P}_m^L(t) \approx \bar{P}_m^L(t)$, where $\bar{G}_m(t)$ and $\bar{P}_m^L(t)$ are set as the average value corresponding to certain heuristic UAV trajectory, e.g., the straight-line flight route from $\mathbf{u}_I$ to $\mathbf{u}_F$. It is verified in [11] that satisfactory accuracy can be obtained with such approximation, and we adopt $\mathbb{E}[R_m(t)] \approx \bar{R}_m(t) \triangleq B\log_2\left(1 + \frac{\gamma_m}{\left((H_U-H_G)^2 + \|\mathbf{u}(t)-\mathbf{g}_m\|^2\right)^{\alpha/2}}\right)$ in the UAV trajectory design, where $\gamma_m \triangleq \frac{P\beta_0 \bar{G}_m(t)\bar{P}_m^L(t)}{\sigma^2}$. As a result, the QoS constraint in (P1) can be approximated as

$$\max_{t\in[0,T]} \left(t - \bar{L}(t)\right) \le \delta, \quad (11)$$

and $\bar{L}(t) \triangleq \max\{\tau | \tau \in [0,t], \sum_{m=1}^{M} z_m(\tau)\bar{R}_m(\tau) \ge R_{\min}\}$.

To ensure (11), one optimal transmission scheduling and association policy can be obtained as $z_{m^*}(t) = 1$ and $z_{m'}(t) = 0, \forall m' \ne m^*$, where $m^* = \arg\max_{m\in\{1,2,\ldots,M\}} \bar{R}_m(t)$. As such, $\bar{L}(t)$ can be expressed as $\bar{L}(t) = \max\{\tau | \tau \in [0,t], \max_m \bar{R}_m(\tau) \ge R_{\min}\}$. In other words, $\bar{L}(t) = \max\{\tau | \tau \in [0,t], \min_m \|\mathbf{u}(t) - \mathbf{g}_m\| \le \bar{d}_m\}$, where $\bar{d}_m \triangleq \sqrt{\left(\frac{\gamma_m}{2^{R_{\min}/B}-1}\right)^{\frac{2}{\alpha}} - (H_U - H_G)^2}$ can be regarded as the minimum horizontal communication radius for video streaming to BS s_m with minimum rate requirement. Define *coverage area* $\mathcal{D}_m \triangleq \{\mathbf{u} | \mathbf{u} \in \mathbb{R}^{2\times 1}, \|\mathbf{u} - \mathbf{g}_m\| \le \bar{d}_m\}$ for BS s_m, which is a disk region with radius $\bar{d}_m$ centered at $\mathbf{g}_m$, $1 \le m \le M$. If $\mathbf{u}(t) \in \mathcal{D}_m, 1 \le m \le M$, then no rate outage will occur. Otherwise, if $\mathbf{u}(t) \notin \bigcup_{m=1}^{M} \mathcal{D}_m$, then rate outage event occurs. In the following, we refer to the event that the UAV enters $\mathcal{D}_m$ as UAV *visits* s_m.

Without loss of generality, we denote the set of visited BSs during the UAV flight as $\mathcal{K} = \{s_{\omega_1}, s_{\omega_1}, \ldots, s_{\omega_K}\} \subseteq \mathcal{M}$ with cardinality $K \triangleq |\mathcal{K}|$, where $\omega_i \in \{1,\ldots,M\}$ is the index of the visited BSs in $\mathcal{M}$, $1 \le i \le K$. Define $t_{\omega_i}^I$ as the time instant that the UAV starts to visit BS s_{ω_i}, then $t_{\omega_i}^I = \min\{t | t \in [0,T], \mathbf{u}(t) \in \mathcal{D}_{\omega_i}\}$. By re-arranging $\{\omega_i\}$ with the increasing order of $t_{\omega_i}^I$ we obtain the visiting order of $\mathcal{K}$, i.e., $\boldsymbol{\pi} = (\pi_1, \pi_2, \ldots, \pi_K)$, which is a permutation of $(\omega_1, \omega_2, \ldots, \omega_K)$. By following the visiting order $\boldsymbol{\pi}$, we define $t_{\pi_i}^O$ as the time instant that the UAV stops visiting BS s_{π_i}, then $t_{\pi_i}^O = \max\{t | t \in [t_{\pi_i}^I, t_{\pi_{i+1}}^I], \mathbf{u}(t) \in \mathcal{D}_{\pi_i}\}$ with $t_{\pi_{K+1}}^I \triangleq T$, $1 \le i \le K$. In this case, we have $t_{\pi_i}^I \le t_{\pi_i}^O \le t_{\pi_{i+1}}^I \le t_{\pi_{i+1}}^O, 1 \le i \le K-1$, and $\|\mathbf{u}(t) - \mathbf{g}_{\pi_i}\| \le \bar{d}_{\pi_i}, t \in [t_{\pi_i}^I, t_{\pi_i}^O]$. In addition, the outage duration between visiting BS s_{π_i} and BS $s_{\pi_{i+1}}$ can be calculated as $t_{\pi_{i+1}}^I - t_{\pi_i}^O$. As a result, the QoS constraint (11) can be expressed as $\max_{i=0,\ldots,K}\{t_{\pi_{i+1}}^I - t_{\pi_i}^O\} \le \delta$ with $t_{\pi_0}^O \triangleq 0$. Based on the aforementioned discussions, the optimization problem (P1) can be reformulated as

$$\begin{aligned}
(\text{P2}): \quad & \min_{T,\{\mathbf{u}(t)\},\mathcal{K},\boldsymbol{\pi},\{t_{\pi_i}^I\},\{t_{\pi_i}^O\}} \int_{t=0}^{T} P_u(t)dt \\
\text{s.t.} \quad & \mathcal{K} \subseteq \mathcal{M}, && (12)\\
& \max_{i=0,\ldots,K}\{t_{\pi_{i+1}}^I - t_{\pi_i}^O\} \le \delta, && (13)\\
& \|\mathbf{u}(t) - \mathbf{g}_{\pi_i}\| \le \bar{d}_{\pi_i}, t \in [t_{\pi_i}^I, t_{\pi_i}^O], i = 1,\ldots,K, && (14)\\
& t_{\pi_i}^I \le t_{\pi_i}^O \le t_{\pi_{i+1}}^I \le t_{\pi_{i+1}}^O, 1 \le i \le K-1, && (15)\\
& t_{\pi_0}^O = 0, t_{\pi_{K+1}}^I = T, && (16)\\
& (8), (9).
\end{aligned}$$

Let $\mathbf{u}_{\pi_i}^I \triangleq \mathbf{u}(t_{\pi_i}^I)$ and $\mathbf{u}_{\pi_i}^O \triangleq \mathbf{u}(t_{\pi_i}^O)$ denote the input waypoint and output waypoint of coverage area $\mathcal{D}_{\omega_i}$, respectively. Without loss of optimality to problem (P2), it is not difficult to prove that the UAV trajectory can be assumed to consist of line segments connecting the $2(K+1)$ waypoints $\mathbf{W} \triangleq \{\mathbf{u}_{\pi_0}^O, \mathbf{u}_{\pi_1}^I, \mathbf{u}_{\pi_1}^O, \ldots, \mathbf{u}_{\pi_K}^I, \mathbf{u}_{\pi_K}^O, \mathbf{u}_{\pi_{K+1}}^I\}$, since we can always replace the curves connecting the above waypoints with line segments which consume less energy. Therefore, the UAV trajectory in (P2) can be represented by $\mathbf{W}$ and time durations along the $2K+1$ line segments, denoted by $\mathbf{T} \triangleq \{\tau(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I), 0 \le i \le K\} \bigcup \{\tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O), 1 \le i \le K\}$. The average UAV velocity along line segment between $\mathbf{u}_{\pi_i}^I$ and $\mathbf{u}_{\pi_i}^O$ can be calculated as $\mathbf{v}(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) = \frac{\mathbf{u}_{\pi_i}^O - \mathbf{u}_{\pi_i}^I}{\tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)}$, and the energy consumption along such line segment can be approximated as $E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) = P_u(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)\tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)$, where $P_u(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)$ is a function with $\left\|\mathbf{v}(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)\right\|$ as in (4), and $E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I)$ can be similarly obtained. As a result, (P2) can be written as

$$\text{(P3)}: \min_{\mathcal{K}, \boldsymbol{\pi}, \mathbf{W}, \mathbf{T}} \left(\sum_{i=0}^{K} E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) + \sum_{i=1}^{K} E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) \right)$$

$$\text{s.t.} \quad \mathcal{K} \subseteq \mathcal{M}, \tag{17}$$

$$\tau(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) \le \delta, 0 \le i \le K, \tag{18}$$

$$\left\|\mathbf{u}_{\pi_i}^I - \mathbf{g}_{\pi_i}\right\| \le \bar{d}_{\pi_i}, \left\|\mathbf{u}_{\pi_i}^O - \mathbf{g}_{\pi_i}\right\| \le \bar{d}_{\pi_i}, i = 1, \ldots, K, \tag{19}$$

$$\mathbf{u}_{\pi_0}^O = \mathbf{u}_I, \mathbf{u}_{\pi_{K+1}}^I = \mathbf{u}_F, \tag{20}$$

$$\left\|\mathbf{u}_{\pi_i}^O - \mathbf{u}_{\pi_i}^I\right\| \le \tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) V_{\max}, 1 \le i \le K, \tag{21}$$

$$\left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\| \le \tau(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) V_{\max}, 0 \le i \le K. \tag{22}$$

B. Proposed Solution

Note that when δ is sufficiently small, constraints (18) and (22) may not be met simultaneously. Thus, we should first check the feasibility of (P3). Considering the special case of (P3) that the UAV flies with the maximum speed, then minimizing UAV's energy consumption is equivalent to minimizing flying distance, which can be written as

$$\text{(P4)}: \min_{\mathcal{K}, \boldsymbol{\pi}, \mathbf{W}, \mathbf{T}} \left(\sum_{i=0}^{K} \left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\| + \sum_{i=1}^{K} \left\|\mathbf{u}_{\pi_i}^O - \mathbf{u}_{\pi_i}^I\right\| \right)$$

$$\text{s.t.} \quad (17) - (20),$$

$$\left\|\mathbf{u}_{\pi_i}^O - \mathbf{u}_{\pi_i}^I\right\| = \tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) V_{\max}, 1 \le i \le K, \tag{23}$$

$$\left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\| = \tau(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) V_{\max}, 0 \le i \le K. \tag{24}$$

It can be verified that (P3) is feasible if and only if (P4) is feasible. By applying the triangle inequality, we have $\left\|\mathbf{g}_{\pi_{i+1}} - \mathbf{g}_{\pi_i}\right\| = \left\|(\mathbf{g}_{\pi_{i+1}} - \mathbf{u}_{\pi_{i+1}}^I) + (\mathbf{u}_{\pi_i}^O - \mathbf{g}_{\pi_i}) + (\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O)\right\| \le \left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{g}_{\pi_{i+1}}\right\| + \left\|\mathbf{u}_{\pi_i}^O - \mathbf{g}_{\pi_i}\right\| + \left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\|$. Due to (19) and non-negativeness of norm function, we have $\left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\| \ge \max\{0, \left\|\mathbf{g}_{\pi_{i+1}} - \mathbf{g}_{\pi_i}\right\| - (\bar{d}_{\pi_i} + \bar{d}_{\pi_{i+1}})\}, 1 \le i \le K-1$. Similarly, we can obtain $\left\|\mathbf{u}_{\pi_1}^I - \mathbf{u}_{\pi_0}^O\right\| \ge \max\{0, \left\|\mathbf{g}_{\pi_1} - \mathbf{u}_I\right\| - \bar{d}_{\pi_1}\}$ and $\left\|\mathbf{u}_{\pi_{K+1}}^I - \mathbf{u}_{\pi_K}^O\right\| \ge \max\{0, \left\|\mathbf{u}_F - \mathbf{g}_{\pi_K}\right\| - \bar{d}_{\pi_K}\}$.

Based on the above analysis, we can construct a weighted graph $G(V, E)$ as follows:

Definition 1. *With given locations of BSs $\{\mathbf{g}_m\}$ and coverage radius $\{\bar{d}_m\}$, a weighted graph $G(V, E, w)$ is defined as:*

- *$V \triangleq \{v_0, v_1, \ldots, v_M, v_{M+1}\}$, where v_o and v_{M+1} correspond to $\mathbf{u}_I$ and $\mathbf{u}_F$, respectively, and v_m corresponds to BS s_m, $1 \le m \le M$.*
- *$E \triangleq \{(v_0, v_m) | \left\|\mathbf{g}_m - \mathbf{u}_I\right\| - \bar{d}_m \le V_{\max}\delta, s_m \in \mathcal{M}\} \cup \{(v_m, v_{M+1}) | \left\|\mathbf{u}_F - \mathbf{g}_m\right\| - \bar{d}_m \le V_{\max}\delta, s_m \in \mathcal{M}\} \cup \{(v_i, v_j) | \left\|\mathbf{g}_i - \mathbf{g}_j\right\| - (\bar{d}_i + \bar{d}_j) \le V_{\max}\delta, s_i, s_j \in \mathcal{M}, s_i \ne s_j\}$.*
- *$w : E \to \mathbf{R}^+$, where w is a weight function. Specifically, $w(v_0, v_m) = \left\|\mathbf{g}_m - \mathbf{u}_I\right\|, \forall s_m \in \mathcal{M}$, $w(v_m, v_{M+1}) = \left\|\mathbf{u}_F - \mathbf{g}_m\right\|, \forall s_m \in \mathcal{M}$, $w(v_i, v_j) = \left\|\mathbf{g}_i - \mathbf{g}_j\right\|, s_i \ne s_j, s_i, s_j \in \mathcal{M}$.*

In other words, (v_i, v_j) exists if and only if the the UAV can fly from the coverage area of s_i to that of s_j within δ. Weight $w(v_i, v_j)$ is assigned to edge (v_i, v_j) to represent the distance between BS s_i and s_j. Due to Definition 1, (P4) is feasible if and only if there exists a connected path in G from v_0 to v_{M+1}. By applying the triangle inequality, the objective function of (P4) is upper bounded by $\left\|\mathbf{g}_{\pi_1} - \mathbf{u}_I\right\| + \left\|\mathbf{g}_{\pi_K} - \mathbf{u}_F\right\| + \sum_{k=1}^{K-1} \left\|\mathbf{g}_{\pi_k} - \mathbf{g}_{\pi_{k+1}}\right\|$. As a result, (P4) can be approximately solved by finding the shortest path over weighted graph $G(V, E, w)$, which can be efficiently obtained by utilizing Dijkstra Algorithm with complexity $O(M^2)$ [33]. Furthermore, the feasibility of (P4) can also be checked by applying Dijkstra Algorithm on $G(V, E, w)$. If no path is returned, then (P4) is infeasible. Otherwise, we obtain a shortest path $\vec{p}^* = (v_0, v_{\pi_1}, \ldots, v_{\pi_{K^*}}, v_{M+1})$, from which the visited BS set $\mathcal{K}^* = \{s_{\pi_k} | v_{\pi_k} \in \vec{p}^*\}$ with $K^* = |\mathcal{K}^*|$ and visiting order $\boldsymbol{\pi}^* = (s_{\pi_1}, \ldots, s_{\pi_{K^*}})$ are determined.

With given $\mathcal{K}^*$ and $\boldsymbol{\pi}^*$, (P3) is reduced to

$$\text{(P5)}: \min_{\mathbf{W}, \mathbf{T}} \left(\sum_{i=0}^{K^*} E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) + \sum_{i=1}^{K^*} E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) \right)$$

$$\text{s.t.} \quad (18) - (22).$$

Problem (P5) is a non-convex optimization problem with coupled variables. In the following, we decompose (P5) into two sub-problems (P6) and (P7), based on which an efficient iterative algorithm is proposed by leveraging alternating optimization and GS search as well as SCA techniques. Specifically, with given $\mathbf{W}$, (P5) is reduced to

$$\text{(P6)}: \min_{\mathbf{T}} \left(\sum_{i=0}^{K^*} E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) + \sum_{i=1}^{K^*} E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) \right)$$

$$\text{s.t.} \quad (18), (21), (22),$$

where $\mathbf{T}$ are the only design variables. Thus, (P6) can be solved by tackling $2K^* + 1$ parallel optimization sub-problems, i.e., minimizing $E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I)$ with $\tau(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) \in \left[\frac{\left\|\mathbf{u}_{\pi_{i+1}}^I - \mathbf{u}_{\pi_i}^O\right\|}{V_{\max}}, \delta\right], 0 \le i \le K^*$ and minimizing $E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O)$ with $\tau(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) \ge \frac{\left\|\mathbf{u}_{\pi_i}^O - \mathbf{u}_{\pi_i}^I\right\|}{V_{\max}}, 1 \le i \le$

K^*. In this case, each sub-problem can be solved by one-dimensional search, e.g., the efficient GS search [34] with complexity $O\left(\log(1/\psi)\right)$, where ψ is the target tolerance.

On the other hand, with given $\mathbf{T}$, (P5) reduces to

$$\text{(P7)}: \min_{\mathbf{W}} \left(\sum_{i=0}^{K^*} E(\mathbf{u}_{\pi_i}^O, \mathbf{u}_{\pi_{i+1}}^I) + \sum_{i=1}^{K^*} E(\mathbf{u}_{\pi_i}^I, \mathbf{u}_{\pi_i}^O) \right) \quad \text{s.t.} \quad (19)-(22).$$

Note that there are $N^* \triangleq 2K^*+1$ line segments in (P7), we reformulate (P7) as (P8) to simplify the use of notations.

$$\text{(P8)}: \min_{\{\mathbf{u}_n\},\{\mathbf{v}_n\}} \sum_{n=1}^{N*} \tau_n P_n^u$$

$$\text{s.t.} \quad \mathbf{u}_1 = \mathbf{u}_I, \mathbf{u}_{N^*+1} = \mathbf{u}_F, \tag{25}$$

$$\|\mathbf{u}_{2i} - \mathbf{g}_{\pi_i}\| \le \bar{d}_{\pi_i}, \|\mathbf{u}_{2i+1} - \mathbf{g}_{\pi_i}\| \le \bar{d}_{\pi_i}, i = 1, \ldots, K^*, \tag{26}$$

$$\|\mathbf{u}_{n+1} - \mathbf{u}_n\| \le \tau_n V_{\max}, 1 \le n \le N^*, \tag{27}$$

where τ_n and $\mathbf{v}_n$ are the time duration and average velocity along the n-th line segment between waypoints $\mathbf{u}_n$ and $\mathbf{u}_{n+1}$, $1 \le n \le N^*$, and $P_n^u \triangleq P_0\left(1+\frac{3\|\mathbf{v}_n\|^2}{U_{tip}^2}\right) + \frac{1}{2}d_0\rho s A\|\mathbf{v}_n\|^3 + P_i\omega_n$ with $\mathbf{v}_n \triangleq \frac{\mathbf{u}_{n+1}-\mathbf{u}_n}{\tau_n}$ and $\omega_n \triangleq \left(\sqrt{1+\frac{\|\mathbf{v}_n\|^4}{4o^4}} - \frac{\|\mathbf{v}_n\|^2}{2o^2}\right)^{1/2}$. Then, equality $\omega_n^2 + \frac{\|\mathbf{v}_n\|^2}{o^2} = \frac{1}{\omega_n^2}$ holds and we have the following problem,

$$\text{(P9)}: \min_{\{\mathbf{u}_n\},\{\mathbf{v}_n\},\{\omega_n\}} \sum_{n=1}^{N^*} \tau_n P_n^u$$

$$\text{s.t.} \quad (25)-(27),$$

$$\omega_n^2 + \frac{\|\mathbf{v}_n\|^2}{o^2} \ge \frac{1}{\omega_n^2}, 1 \le n \le N^*. \tag{28}$$

It can be verified that (P9) is equivalent to (P8) since we can always decrease ω_n in (P9) to achieve the equality in (28) and the objective value also decreases. With the SCA technique [11], by apply the first-order Taylor expansion on the LHS of (28) with given local point $\omega_n^r, \mathbf{v}_n^r$, we have $\omega_n^2 + \frac{\|\mathbf{v}_n\|^2}{o^2} \ge (\omega_n^r)^2 + 2\omega_n^r(\omega_n - \omega_n^r) + \frac{\|\mathbf{v}_n\|^2}{o^2} + \frac{2}{o^2}(\mathbf{v}_n^r)^T(\mathbf{v}_n - \mathbf{v}_n^r) \triangleq z_n^{lb}$. Then, (P9) can be approximated as

$$\text{(P10)}: \min_{\{\mathbf{u}_n\},\{\mathbf{v}_n\},\{\omega_n\}} \sum_{n=1}^{N^*} \tau_n P_n^u$$

$$\text{s.t.} \quad (25)-(27),$$

$$z_n^{lb} \ge \frac{1}{\omega_n^2}, 1 \le n \le N^*. \tag{29}$$

Problem (P10) is a standard convex optimization problem and can be efficiently solved by CVX solvers [35]. With the alternating optimization technique, (P6) and (P10) are optimized alternately with the local points updated until convergence is achieved. The details of the proposed algorithm for (P3) is summarized in Algorithm 1. With the alternating optimization and SCA techniques, Algorithm 1 is guaranteed to converge to a KKT solution [11] and the computation complexity is given by $O\left(M^{3.5}\log(1/\psi)\right)$.

Algorithm 1 Graph based Optimization Algorithm for Problem (P3)

1: Construct graph $G(V, E, w)$ as in Definition 1;
2: Find the shortest path $\bar{p}^*$ in G with Dijkstra Algorithm;
3: **if** $\bar{p}^* = \emptyset$ **then**
4: **Output**: Infeasible.
5: **else**
6: Obtain visited BS set $\mathcal{K}^*$ with $K^* = |\mathcal{K}^*|$ and visiting order $\boldsymbol{\pi}^*$ based on $\bar{p}^*$;
7: Set $r = 0$, and initialize UAV waypoints $\mathbf{W}^r$;
8: **repeat**
9: With given $\mathbf{W}^r$, solve (P6) to obtain optimal $\mathbf{T}^{r+1}$ by GS search;
10: With given $\mathbf{T}^{r+1}$ and $\mathbf{W}^r$, solve (P10) with CVX solver to obtain optimal UAV waypoints $\mathbf{W}^{r+1}$;
11: $r = r + 1$;
12: **until** Convergence is achieved;
13: $\mathbf{W}^* = \mathbf{W}^r$, $\mathbf{T}^* = \mathbf{T}^r$;
14: **end if**
15: **Output**: $\{\mathcal{K}^*, \boldsymbol{\pi}^*, \mathbf{W}^*, \mathbf{T}^*\}$.

V. DRL BASED SOLUTION

In this section, we considerd the site-specific performance for specific urban local environments.

A. Problem Reformulation

Similar to Section IV-A, an optimal transmission scheduling policy for (P1) can be given by

$$z_m(t) = \begin{cases} 1, & m = \arg\max_{m' \in \{1,2,\ldots,M\}} R_{m'}(t), \\ 0, & \text{Otherwise.} \end{cases} \tag{30}$$

As such, $L(t)$ can be expressed as $L(t) = \max\{\tau | \tau \in [0,t], \max_m R_m(\tau) \ge R_{\min}\}$. Then, (P1) can be written as

$$\text{(P11)}: \min_{T,\{\mathbf{u}(t)\}} \int_{t=0}^{T} P_u(t)dt \quad \text{s.t.} \quad (7)-(9).$$

By employing the path discretization approach [11], we discretize the UAV path into N line segments each with length Δ, where Δ is small enough such that within each line segment, the distances between the UAV and all BSs are approximately unchanged. As such, the UAV trajectory $\mathbf{u}(t)$ can be approximated by $N+1$ waypoints $\{\mathbf{u}_n, 1 \le n \le N+1\}$ and time durations along the line segments $\{\tau_n, 1 \le n \le N\}$, where $\mathbf{u}_1 = \mathbf{u}_I$ and $\mathbf{u}_{N+1} = \mathbf{u}_F$. In this case, $\mathbf{v}_n = \frac{\mathbf{u}_{n+1}-\mathbf{u}_n}{\tau_n}$ denotes the average velocity along n-th line segment. The distance between the UAV and s_m at the n-th line segment is $d_{m,n} = \sqrt{(H_U - H_G)^2 + \|\mathbf{u}_n - \mathbf{g}_m\|^2}$. Similarly, the discretized form of $\theta_m(t)$, $G_m(t)$, $\beta_m(t)$, $\tilde{h}_m(t)$, and $R_m(t)$ are denoted as $\theta_{m,n}$, $G_{m,n}$, $\beta_{m,n}$, $\tilde{h}_{m,n}$, and $R_{m,n}$, respectively, where $R_{m,n} = B\log_2\left(1+\frac{PG_{m,n}\beta_{m,n}|\tilde{h}_{m,n}|^2}{\sigma^2}\right)$. In addition, the UAV's energy consumption along the n-th line

segment is $E_n = P_n^u t_u$, where P_n^u is a function of $\|\mathbf{v}_n\|$ as given in (4).

Note that the maximum allowed outage duration is δ. Then we can check whether the QoS requirement is satisfied in every δ time units. Let $\tau_n \leq \delta$, for all n. Then if $\max_m R_m[n] \geq R_{\min}$, for all n, the maximum outage duration will be limited to δ. As a result, (P11) can be approximated as

$$\begin{aligned}
(\text{P12}): \quad & \min_{N,\{\mathbf{u}_n\},\{\tau_n\}} \sum_{n=1}^{N} E_n \\
\text{s.t.} \quad & \max_m R_{m,n} \geq R_{\min}, \forall n, && (31)\\
& \frac{\Delta}{V_{\max}} \leq \tau_n \leq \delta, \forall n, && (32)\\
& \mathbf{u}_{n+1} - \mathbf{u}_n = \mathbf{v}_n \tau_n, \forall n = 1, \ldots, N, && (33)\\
& \|\mathbf{u}_{n+1} - \mathbf{u}_n\| = \Delta, \forall n = 1, \ldots, N, && (34)\\
& \mathbf{u}_1 = \mathbf{u}_I, \mathbf{u}_{N+1} = \mathbf{u}_F. && (35)
\end{aligned}$$

Note that with given Δ, E_n is only relevant to τ_n, and then the optimal τ_n^* can be solved by GS search on τ_n over interval $[\frac{\Delta}{V_{\max}}, \delta]$. Thus, $\|\mathbf{v}_n\|^* = \frac{\Delta}{\tau_n^*}$, based on which E_n^* is obtained. Denote $\vec{\mathbf{v}}_n \triangleq \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|^*}$ as the UAV flying direction, where $\|\vec{\mathbf{v}}_n\| = 1$. Then problem (P12) can be written as

$$\begin{aligned}
(\text{P13}): \quad & \max_{N,\{\mathbf{u}_n\},\{\vec{\mathbf{v}}_n\}} -\sum_{n=1}^{N} E_n^* \\
\text{s.t.} \quad & \mathbf{u}_{n+1} = \mathbf{u}_n + \Delta \vec{\mathbf{v}}_n, \forall n = 1, \ldots, N, && (36)\\
& (31), (35).
\end{aligned}$$

Note that $R_{m,n}$ in (31) depends on antenna gain $G_{m,n}$ and large-scale channel gain $\beta_{m,n}$, as well as small-scale fading $\tilde{h}_{m,n}$. With given $\mathbf{u}_n$, the LoS/NLoS states are exactly determined by checking whether there are obstacles between the UAV and BSs in the specific local environment, and $G_{m,n}$ and $\beta_{m,n}$ are exactly determined. In addition, $\tilde{h}_{m,n}$ can be measured by leveraging the existing handover mechanisms with continuous reference signal received power (RSRP) measurements [36]. Therefore, $R_{m,n}$ can be measured at $\mathbf{u}_n$, and (P13) can be modeled as an MDP $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R})$ as follows:

- *State*: $\mathcal{S} \subseteq \mathbb{R}^{2\times 1}$ is the continuous state space which consists of all possible UAV horizonal locations. State $\mathbf{u}_n \in \mathcal{S}$ corresponds to the UAV location at time step n.
- *Action*: $\mathcal{A} \triangleq \{\vec{\mathbf{v}} | \|\vec{\mathbf{v}}\| = 1\}$ is the action space which consists of all possible flying directions. Action $\vec{\mathbf{v}}_n \in \mathcal{A}$ corresponds to the flying direction at time step n. The set of actions is discretized into κ values uniformly such that the action space is finite, i.e., $\mathcal{A} = \{\vec{\mathbf{v}}^{(1)}, \ldots, \vec{\mathbf{v}}^{(\kappa)}\}$.
- *State Transition*: The state transition probability is deterministically computed as in (36).
- *Reward*: The reward function $\mathcal{R} \triangleq \{\hat{R}_n\}$ is designed to award the UAV for reaching its destination, and penalize the UAV for energy consuming or violating the QoS requirement, i.e.,

$$\hat{R}_n = \begin{cases} \hat{R}_{des} - \lambda E_n^* - I_{n+1}, & \mathbf{u}_{n+1} = \mathbf{u}_F, \\ -\lambda E_n^* - I_{n+1}, & \text{otherwise}, \end{cases} \quad (37)$$

where $I_{n+1} = 1$ when $\max_m R_{m,n+1} < R_{\min}$, and $I_{n+1} = 0$ otherwise. λ is a normalization factor for the energy cost.

Note that traditional tabular Q-learning method needs to store every possible state-action value, which is impractical for problem (P13) due to the continuous state space $\mathcal{S}$. In the following, we propose a DRL based algorithm based on DQN with a dueling network architecture and double Q-learning.

B. Proposed Solution

In DRL, deep learning is employed to learn the state-action function with neural networks instead of the Q-table. A DQN with dueling network architecture, also named dueling DQN [37], is a multi-layered neural network that is used to approximate state-action value $Q(\mathbf{u}, \vec{\mathbf{v}})$ with network parameters $\boldsymbol{\theta}$, such that the training process can be stabilized. Compared to the standard DQN, there are two separate estimators in dueling DQN, one is used to estimate the state values and the other is used to represent the state-dependent action advantages, and the output of the two estimators are combined in a smart way to produce output Q-functions $\hat{Q}(\mathbf{u}, \vec{\mathbf{v}}; \boldsymbol{\theta})$. In this case, the dueling DQN can learn the state-value function more efficiently, and is also more robust to the approximation errors between the output $\hat{Q}(\mathbf{u}, \vec{\mathbf{v}}; \boldsymbol{\theta})$ and the true value of $Q(\mathbf{u}, \vec{\mathbf{v}})$.

Two significant ingredients to stabilize the training of dueling DQN are the experience replay and the dual network, which contains a Q network and a target network. The target network with parameters $\boldsymbol{\theta}^-$, is a copy of the Q network with parameters $\boldsymbol{\theta}$, where $\boldsymbol{\theta}^-$ are copied from $\boldsymbol{\theta}$ every Ω steps. As for experience replay, the state-action-reward-nextState transitions $(\mathbf{u}_n, \vec{\mathbf{v}}_n, \hat{R}_n, \mathbf{u}_{n+1})$ are stored in a buffer and a mini-batch of experiences from the replay buffer are randomly sampled to perform the updates during learning, where the random sampling breaks the correlation between sequential samples and stabilizes the training process. In addition, the double Q-learning technique [38] is employed to reduce the overestimations of action values in the training process, which decouples the action selection from action evaluation of bootstrapping by using two separate networks. To be specific, the Q network with $\boldsymbol{\theta}$ is used for action selection while the target network with $\boldsymbol{\theta}^-$ is used for action evaluation of bootstrapping. Therefore, the update target of training algorithm for dueling DQN with double Q-learning is given by

$$y_n = \hat{R}_n + \gamma \hat{Q}(\mathbf{u}_{n+1}, \arg\max_{\vec{\mathbf{v}}' \in \mathcal{A}} \hat{Q}(\mathbf{u}_{n+1}, \vec{\mathbf{v}}'; \boldsymbol{\theta}); \boldsymbol{\theta}^-), \quad (38)$$

where γ is the discount factor and the training algorithm aims to minimize the loss given by $(y_n - \hat{Q}(\mathbf{u}_n, \vec{\mathbf{v}}_n; \boldsymbol{\theta}))^2$. The UAV selects action $\vec{\mathbf{v}}_n$ at state $\mathbf{u}_n$ according to the ϵ-greedy policy [39]. The details of the DRL algorithm for problem (P13) is summarized in Algorithm 2.

In Algorithm 2, $\Psi^{\max}$ and $N^{\max}$ denote the maximum number of episodes and the maximum number of time steps for each episode. Adam optimizer is utilized to guide the updating process of dueling DQN for minimization of the

Algorithm 2 DRL Algorithm for Problem (P13)

1: Initialize dueling DQN with parameters $\boldsymbol{\theta}$ and the target network with parameters $\boldsymbol{\theta}^- = \boldsymbol{\theta}$. Initialize the experience replay buffer $\mathcal{C}$;
2: **for** episode $= 1$ to $\Psi^{\max}$ **do**
3: Set the initial state $\mathbf{u}_1 = \mathbf{u}_I$. Set $n = 1$;
4: **while** $n \leq N^{\max}$ and $\mathbf{u}_n \neq \mathbf{u}_F$ **do**
5: Choose action $\vec{\mathbf{v}}_n$ according to ϵ-greedy policy;
6: Take action $\vec{\mathbf{v}}_n$ and obtain state $\mathbf{u}_{n+1}$, measure $R_{m,n+1}$ at $\mathbf{q}_{n+1}$. Obtain E_n^*, τ_n^* as well as $\|\mathbf{v}_n\|^*$ by GS search, and obtain the immediate reward $\hat{R}_n$ based on (37);
7: Store the transition $(\mathbf{u}_n, \vec{\mathbf{v}}_n, \hat{R}_n, \mathbf{u}_{n+1})$ in $\mathcal{C}$;
8: Sample a random minibatch of transitions $(\mathbf{u}_j, \vec{\mathbf{v}}_j, R_j, \mathbf{u}_{j+1})$ from $\mathcal{C}$;
9: Let $\vec{\mathbf{v}}^* = \arg\max_{\vec{\mathbf{v}} \in \mathcal{A}} \hat{Q}(\mathbf{u}_{j+1}, \vec{\mathbf{v}}; \boldsymbol{\theta})$, and set

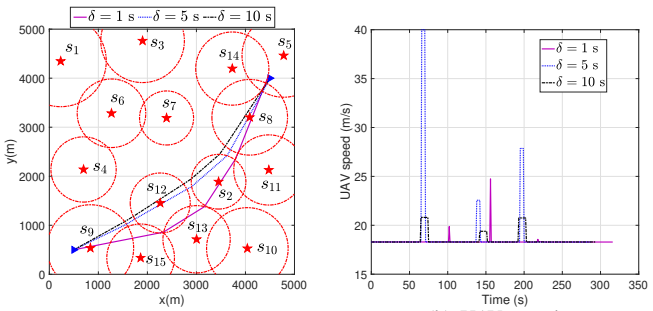
$$y_j = \begin{cases} R_j, & \text{if } \mathbf{u}_{j+1} = \mathbf{u}_F, \\ R_j + \gamma \hat{Q}(\mathbf{u}_{j+1}, \vec{\mathbf{v}}^*; \boldsymbol{\theta}^-), & \text{otherwise;} \end{cases} \quad (39)$$

10: Perform an Adam gradient descent step on $\left(y_j - \hat{Q}(\mathbf{q}_j, \vec{\mathbf{v}}_j; \boldsymbol{\theta})\right)^2$ with respect to $\boldsymbol{\theta}$;
11: $n = n + 1$;
12: **end while**
13: Update the target network parameters $\boldsymbol{\theta}^- = \boldsymbol{\theta}$ every Ω episodes, periodically;
14: **end for**

mean square error loss at Step 10. Algorithm 2 adopts a dueling DQN consisting of five hidden layers, where the number of neurons for the first four layers are 512, 256, 128 and 128, respectively. The last hidden layer is the dueling layer with $\kappa + 1$ neurons. The activation function for all the hidden layers is the ReLu function [39]. The proposed learning Algorithm 2 is guaranteed to converge [37].

VI. Simulation Study

We consider an aerial surveillance system with a cellular-connected UAV, and there are $M = 15$ ground BSs, which are uniformly distributed in a square area with width 5.0 km, as shown in Fig. 1(a). The building locations and heights of the urban area are generated based on one realization of the statistical model given by the International Telecommunication Union (ITU) [40]. The parameters related to the rotary-wing UAV are set as in [12]: $d_0 = 0.6$, $U_{tip} = 120$, $A = 0.503$, $\rho = 1.225$, $s = 0.05$, $o = 4.03$, $P_0 = 79.8563$, $P_i = 88.6279$, $V_{\max} = 50$ m/s. The parameters related to the antenna patterns are set as in [41]: $HPBW_v = 65°$, $\theta_D = 10°$, $G_m = 30$ dB, $N_0 = 8$. The parameters related to A2G channels are set as: $\beta_0 = -60$ dB, $\sigma^2 = -110$ dBm, $\alpha = 2.2$, $a = 10$, $b = 0.6$, $\mu = 0.2$, $K_c = 15$ dB as in [42]. For the DRL algorithm, we set $\Psi^{\max} = 5000$, $N^{\max} = 800$, $\kappa = 8$, $\epsilon = 0.5$, $\Omega = 5$, and $\lambda = 10^{-3}$. Multi-step learning is employed with step size equal to 30 to enable faster learning [43]. Unless otherwise stated, the other parameters related to the simulations are set as follows: $H = 100$ m, $H_G = 25$ m, $P = 0.1$ W, $B = 1$ MHz, $R_{\min} = 0.4$ Mbps, $\Delta = 20$ m, and $\delta = 1$ s.



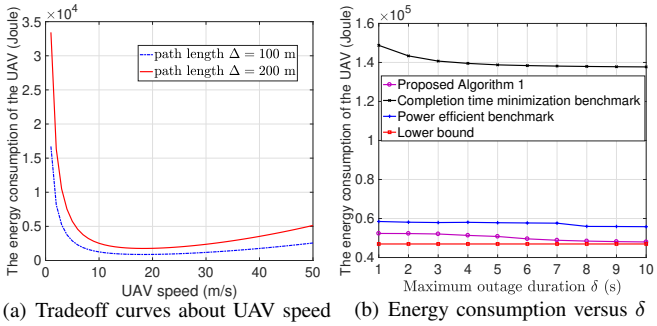
(a) UAV trajectory (b) UAV speed

Fig. 1. Optimized trajectories and UAV speeds with different δ ($R_{\min} = 0.4$ Mbps). The red stars denote the horizonal locations of ground BSs. The blue triangles denote the initial and final UAV locations.

A. Average Performance with the Graph Based Algorithm

We first consider the average communication performance for uncertain local environments. In Fig. 1(a), the circle around BS s_m denotes the coverage area D_m. The UAV sequentially enters and leaves out the coverage areas of BSs, where the rate outage duration is upper bounded by δ. As δ increases, the number of visited BSs decreases as more freedom can be employed and the UAV tends to fly directly to the destination, which leads to less energy consumption of the UAV. From Fig. 1(b) we can see that the UAV flies with the energy minimization speed (e.g., 18m/s in our setup) in the coverage areas of BSs, and will increase the speed when it is out of these coverage areas. The energy minimization speed should balance the tradeoff between decreasing the propulsion power and decreasing the flying time. With given path length Δ, the UAV's energy consumption versus UAV speed is plotted in Fig. 2(a). It is firstly observed that with the increase of UAV speed, the rotary-wing UAV's energy consumption first decreases and then increases, and the minimum value can be efficiently solved by GS search, which is in accordance with Fig. 1(b). When the UAV is out of the coverage areas, the speed should balance the tradeoff between decreasing the rate outage duration and decreasing the energy consumption, which is efficiently obtained by our proposed Algorithm 1.

We compare Algorithm 1 with three benchmark schemes, which are referred to as the *Completion time minimization benchmark* and the *Power efficient benchmark*. In the completion time minimization benchmark, the UAV flies with the maximum speed by following the visiting order determined by graph algorithm in [44]. In the power efficient benchmark, the UAV flies with the power efficient speed in [12] and follows the visiting order determined by our algorithm. In addition, by ignoring the QoS constraints in (5)-(7), we obtain a lower bound of UAV's energy consumption by *energy minimization straight-line flight* from $\mathbf{u}_I$ to $\mathbf{u}_F$ with optimal speed determined by GS search. It is observed in Fig. 2(b) that Algorithm 1 consumes less energy than other benchmark schemes while satisfying the QoS requirement, and the performance gains are more pronounced when δ increases, as expected. When δ



(a) Tradeoff curves about UAV speed (b) Energy consumption versus δ

Fig. 2. Performance comparisons with different schemes for average communication performance.

is large enough, the UAV's energy consumption obtained by Algorithm 1 achieves the lower bound, since QoS requirement is no longer the system bottleneck. Algorithm 1 performs better than the power efficient benchmark since the energy minimization speed is different from the power efficient speed, which is in accordance with Fig. 2(a). The performance gap between the completion time minimization benchmark and Algorithm 1 demonstrates the additional gain of the joint design of visiting order and UAV waypoints as well as time duration along line segments.

B. Site-specific Performance with the DRL Based Algorithm

Next, we consider the site-specific performance for the given specific local environment when $R_{\min} = 0.3$ Mbps and $\delta = 1$ s. Fig. 3 shows the coverage area map, which specifies whether $R_{\min}$ is achieved at any location for real rate measurement. The circles around BSs denote the coverage areas obtained by the statistical model as in Section VI-A. It is observed that the statistical model based coverage areas are similar to that obtained by site-specific performance with minor differences. This demonstrates the approximation accuracy of average performance and the characteristic of site-specific performance, which is influenced by the actual radio propagation environment with buildings and BS antenna patterns. From Fig. 3, we can see that with DRL based Algorithm 2, the UAV tries to fly to the destination by avoiding coverage holes or staying in coverage holes for no longer than δ, such that the QoS requirement can be satisfied. The other schemes may violate the QoS requirement since they are only based on the statistical model for average performance without knowledge of the specific local environment.

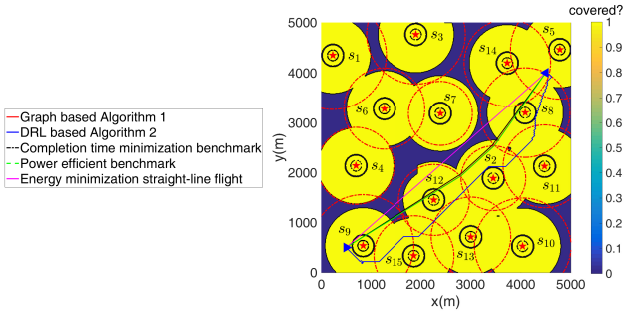
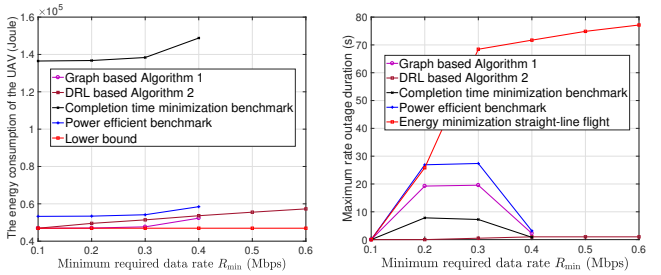


Fig. 3. Optimized trajectory for site-specific performance ($R_{\min} = 0.3$ Mbps, $\delta = 1$ s).

With different $R_{\min}$ values, Fig. 4(a) shows the UAV's energy consumption while Fig. 4(b) shows the maximum rate outage duration for the given specific local environment when $\delta = 1$ s. When $R_{\min}$ increases, the coverage area shrinks and thus more energy is required to fly to the destination while satisfying the QoS requirement. It is observed that no results are plotted in Fig. 4 when $R_{\min} > 0.4$ Mbps for statistical model based schemes, since these schemes are based on graph model constructing by the disk coverage area without any knowledge about the specific local environment. When $R_{\min} > 0.4$ Mbps, the disk coverage area shrinks and there exists no feasible path in the constructed graph, and thus no feasible solution can be obtained. With DRL based Algorithm 2, the UAV learns to avoid coverage holes from accumulated experience even when $R_{\min} > 0.4$ Mbps by interacting with the specific local environment, and the maximum rate outage duration is below the threshold with a comparable energy consumption as Algorithm 2 and the lower bound, which demonstrates the effectiveness of the DRL based solution. Although the energy minimization straight-line flight achieves the lower bound of UAV energy consumption, it leads to larger rate outage durations since the QoS requirement is ignored.



(a) The energy consumption of UAV versus $R_{\min}$ (b) The maximum rate duration along the UAV flight versus $R_{\min}$

Fig. 4. Performance comparisons with different schemes for site-specific performance ($\delta = 1$ s).

VII. Conclusion

In this paper, we considered cellular-connected UAV enabled aerial surveillance and aimed to minimize the energy consumption while satisfying the QoS requirement for video streaming. To tackle the environment-aware MINLP problem, we first proposed a graph based algorithm by employing graph theory and convex optimization techniques to obtain a suboptimal solution for average communication performance. Then, in order to investigate site-specific performance for specific urban local environment, we proposed a DRL based algorithm by employing a dueling DQN neural network model with only local observations of sampled rate measurements. Simulation results showed that the proposed algorithms outperformed several baseline schemes and provided interesting insights for the aerial surveillance system.

Acknowledgement

Han Hu is corresponding author for this paper. This work was supported in part by the National Natural Science Foundation of China under Grant 62172339 and Grant 61971457, in part by the NSF under Grant ECCS-1923163.

REFERENCES

[1] M. Zolanvari, R. Jain, and T. Salman, "Potential data link candidates for civilian unmanned aircraft systems: A survey," *IEEE Commun. Surveys & Tutorials*, vol. 22, no. 1, pp. 292-319, 2020.

[2] 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.

[3] M. Chen, W. Saad, and C. Yin, "Echo-liquid state deep learning for 360° content transmission and caching in wireless VR networks with cellular connected UAVs," *IEEE Trans. Commun.*, vol. 67, no. 9, pp. 6386-6400, Sep. 2019.

[4] Y. Zhang, C. Xu, I. A. Hemadeh, M. El-Hajjar, and L. Hanzo, "Near-instantaneously adaptive multi-set space-time shift keying for UAV-aided video surveillance," *IEEE Trans. Veh. Techn.*, vol. 69, no. 11, pp. 12843-12856, Nov. 2020.

[5] M. Mozaffari, W. Saad, M. Bennis, Y.-H. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys & Tutorials*, vol. 21, pp. 2334-2360, 2019.

[6] C. Zhan, H. Hu, Z. Wang, R. Fan, and D. Niyato, "Unmanned aircraft system aided adaptive video streaming: A joint optimization approach," *IEEE Trans. Multimedia*, vol. 22, no. 3, pp. 795-807, Mar. 2020.

[7] Y. Zeng, J. Lyu, and R. Zhang, "Cellular-connected UAV: Potential, challenges, and promising technologies," *IEEE Wireless Commun.*, vol. 26, no. 1, pp. 120-127, Feb. 2019.

[8] M. M. Azari, F. Rosas, and S. Pollin, "Cellular connectivity for UAVs: Network modeling, performance analysis, and design guidelines," *IEEE Trans. Wireless Commun.*, vol. 18, no. 7, pp. 3366-3381, Jul. 2019.

[9] C. Zhan and Y. Zeng, "Energy-efficient data uploading for cellular-connected UAV systems," *IEEE Wireless Commun.*, vol. 19, no. 11, pp. 7279-7292, Nov. 2020.

[10] X. Xiao, W. Wang, T. Chen, Y. Cao, T. Jiang, and Q. Zhang, "Sensor augmented neural adaptive bitrate video streaming on UAVs," *IEEE Trans. Multimedia*, vol. 22, no. 6, pp. 1567-1576, Jun. 2020.

[11] Y. Zeng and R. Zhang, "Energy-efficient UAV communication with trajectory optimization," *IEEE Trans. Wireless Commun.*, vol. 16, no. 6, pp. 3747-3760, Jun. 2017.

[12] Y. Zeng, J. Xu, and R. Zhang, "Energy minimization for wireless communication with rotary-wing UAV," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2329-2345, Apr. 2019.

[13] Y. Chen, H. Zhang, and Y. Hu, "Optimal power and bandwidth allocation for multiuser video streaming in UAV relay networks," *IEEE Trans. Veh. Techn*, vol. 69, no. 6, pp. 6644-6655, Jun. 2020.

[14] X. Tang, X. Huang, and F. Hu, "QoE-driven UAV-enabled pseudo-analog wireless video broadcast: A joint optimization of power and trajectory," *IEEE Trans. Multimedia*, vol. 23, pp. 2398-2412, 2021.

[15] C. Zhan, H. Hu, X. Sui, Z. Liu, J. Wang, and H. Wang, "Joint resource allocation and 3D aerial trajectory design for video streaming in UAV communication systems," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 8, pp. 3227-3241, Aug. 2021.

[16] I.-S. Comsa, G.-M. Muntean, and R. Trestian, "An innovative machinelearning-based scheduling solution for improving live UHD video streaming quality in highly dynamic network environments," *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 212-224, Mar. 2021.

[17] L. Xiao, Y. Ding, J. Huang, S. Liu, Y. Tang, and H. Dai, "UAV anti-jamming video transmissions with QoE guarantee: A reinforcement learning-based approach," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5933-5947, Sep. 2021.

[18] H. Huang, A. V. Savkin, and W. Ni, "Online UAV trajectory planning for covert video surveillance of mobile targets," to appear in *IEEE Trans. Autom. Sci. Eng.*.

[19] S. Hu, W. Ni, X. Wang, A. Jamalipour, and D. Ta, "Joint optimization of trajectory, propulsion, and thrust powers for covert UAV-on-UAV video tracking and surveillance," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 1959-1972, 2021.

[20] R. Amer, W. Saad, and N. Marchetti, "Mobility in the sky: Performance and mobility analysis for cellular-connected UAVs," *IEEE Trans. Commun.*, vol. 68, no. 5, pp. 3229-3246, May 2020.

[21] W. Mei and R. Zhang, "Cooperative downlink interference transmission and cancellation for cellular-Connected UAV: A divide-and-conquer approach," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1297-1311, Feb. 2020.

[22] W. Mei, Q. Wu, and R. Zhang, "Cellular-connected UAV: Uplink association, power control and interference coordination," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5380-5393, Nov. 2019.

[23] J. Hou, Y. Deng, and M. Shikh-Bahaei, "Joint beamforming, user association, and height control for cellular-enabled UAV communications," *IEEE Trans. Commun.*, vol. 69, no. 6, pp. 3598-3613, Jun. 2021.

[24] S. Zhang and R. Zhang, "Radio map-based 3D path planning for cellular-connected UAV," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1975-1989, Mar. 2021.

[25] N. Senadhira, S. Durrani, X. Zhou, N. Yang, and M. Ding, "Uplink NOMA for cellular-connected UAV: Impact of UAV trajectories and altitude," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5242-5258, Aug. 2020.

[26] X. Pang et al., "Uplink precoding optimization for NOMA cellular-connected UAV networks," *IEEE Trans. Commun.*, vol. 68, no. 2, pp. 1271-1283, Feb. 2020.

[27] C. Zhan and R. Huang, "Energy efficient adaptive video streaming with rotary-wing UAV," *IEEE Trans. Veh. Techn.*, vol. 69, no. 7, pp. 8040-8044, Jul. 2020.

[28] H. Hu, C. Zhan, J. An, and Y. Wen, "Optimization for HTTP adaptive video streaming in UAV-enabled relaying system," in Proc. *IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1-6.

[29] *Study on 3D Channel Model for LTE*, document TR 36.873 V12.7.0, 3GPP Dec. 2017.

[30] X. Xu and Y. Zeng, "Cellular-connected UAV: Performance analysis with 3D antenna modelling," in Proc. *IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2019, pp. 1-6.

[31] A. Al-Hourani, S. Kandeepan, and S. Lardner, "Optimal LAP altitude for maximum coverage," *IEEE Wireless Commun. Lett.*, vol. 3, no. 6, pp. 569-572, Dec. 2014.

[32] Y. Sun, D. Xu, D. W. K. Ng, L. Dai, and R. Schober, "Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems," *IEEE Trans. Commun.*, vol. 67, no. 6, pp. 4281-4298, Jun. 2019.

[33] D. B. West, *Introduction to Graph Theory*. Prentice Hall, 2001.

[34] Z. Zhang, Z. Ma, Y. Xiao, M. Xiao, G. K. Karagiannidis, and P. Fan, "Non-orthogonal multiple access for cooperative multicast millimeter wave wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1794-1808, Aug. 2017.

[35] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," 2016. [Online] Available: http://cvxr.com/cvx.

[36] Y. Zeng, X. Xu, S. Jin, and R. Zhang, "Simultaneous navigation and radio mapping for cellular-connected UAV with deep reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4205-4220, Jul. 2021.

[37] Z. Wang, et al., "Dueling network architectures for deep reinforcement learning," in Proc. of *the 33rd Int. Conf. Machine Learning (ICML)*, Jun. 2016, pp. 1995-2003.

[38] H. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. AAAI*, 2016, pp. 2094-2100.

[39] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016, http://www.deeplearningbook.org.

[40] *Propagation Data and Prediction Methods Required for the Design of Terrestrial Broadband Radio Access Systems Operating in a Frequency Range From 3 to 60 GHz*, document ITU-R, Rec. 1410-5, Feb. 2012.

[41] *Technical Specification Group Radio Access Network: Study on Enhanced LTE Support for Aerial Vehicles*, document 3GPP TR 36.777 V15.0.0, Dec. 2017.

[42] F. Ono, H. Ochiai, and R. Miura, "A wireless relay network based on unmanned aircraft system with rate optimization," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7699-7708, Nov. 2016.

[43] M. Hessel, et al., "Rainbow: Combining improvements in deep reinforcement learning," in Proc. of *AAAI*, 2018.

[44] S. Zhang and R. Zhang, "Trajectory design for cellular-connected UAV under outage duration constraint," in Proc. *IEEE International Conference on Communications (ICC)*, 2019, pp. 1-6.