Aerial Video Streaming Over 3D Cellular Networks: An Environment and Channel Knowledge Map Approach

Cheng Zhan, *Member, IEEE*, Han Hu, *Member, IEEE*, Zhi Liu, *Senior Member, IEEE*, Jing Wang, *Member, IEEE*, Nan Cheng, *Senior Member, IEEE*, and Shiwen Mao, *Fellow, IEEE*

***Abstract*— Aerial video streaming is a promising application of unmanned aerial vehicles (UAVs), which extends video service from ground to three-dimensional (3D) airspaces. However, high data rates and smooth transmission are required along with ubiquitous and environment-aware communications. To this end, we study the quality of experience (QoE) maximization problem in this paper for aerial video streaming over 3D cellular networks in urban environments with building avoidance. Different from the typical channel model based optimization in prior works, we tackle the joint design of 3D UAV trajectory and transmission scheduling as well as playback rate adaption with an environment and channel knowledge map (ECKM) approach, which provides rich information about the location-specific channel for enabling environment-aware communications. Specifically, we first consider the scenario with perfect ECKM, and propose efficient algorithms to obtain suboptimal solutions by utilizing two graph models and the iterative parameter-enabled block coordinate descent method. For the scenario without such map information, we propose a dueling Deep Q-learning (DQL) solution with map construction such that the learning process can be facilitated for path planning. Simulation results are provided to demonstrate the improvement in QoE by the proposed solutions over baseline schemes, as well as a tradeoff between video quality and rate variation.**

***Index Terms*— Aerial video streaming, unmanned aerial vehicle (UAV), quality of experience (QoE), environment and channel knowledge map, dueling deep Q-learning.**

Manuscript received 4 February 2023; revised 15 May 2023; accepted 21 June 2023. Date of publication 5 July 2023; date of current version 13 February 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62172339 and Grant 61971457 and in part by the Natural Science Foundation of Chongqing under Grant CSTB2022NSCQ-MSX0338. The associate editor coordinating the review of this article and approving it for publication was L.-C. Wang. *(Corresponding author: Han Hu.)*

Cheng Zhan is with the School of Computer and Information Science, Southwest University, Chongqing 400715, China (e-mail: zhanc@swu.edu.cn).

Han Hu is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: hhu@bit.edu.cn).

Zhi Liu is with the Graduate School of Informatics and Engineering, The University of Electro-Communications, Tokyo 182-8585, Japan (e-mail: liu@ieee.org).

Jing Wang is with the School of Information, Renmin University of China, Beijing 100872, China (e-mail: jwang_ruc@163.com).

Nan Cheng is with the State Key Laboratory of ISN and the School of Telecommunications Engineering, Xidian University, Xi'an 710071, China (e-mail: dr.nan.cheng@ieee.org).

Shiwen Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: smao@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3289501.

Digital Object Identifier 10.1109/TWC.2023.3289501

I. Introduction

In RECENT years, mobile video streaming has gained rapid development, which encompass a wide variety of applications, especially video services [1]. The rapid increase in high-quality video services impose great pressure on the current cellular network infrastructure, which becomes a bottleneck for quality of experience (QoE) provisioning to users. It is expected that the sixth-generation (6G) networks will provide wide coverage areas, seamless connection, and low latency, where integrated aerial and terrestrial architectures becomes an important component [2]. Due to their high flexibility and low deployment cost, unmanned aerial vehicles (UAVs) equipped with sensors and cameras have been widely used in urban environments for various surveillance tasks, such as security monitoring, aerial filming, aerial virtual reality and target tracking through aerial video streaming [3], [4], [5]. However, UAV enabled remote sensing and surveillance may not be effectively supported by the existing short range wireless technologies. To this end, cellular connected UAVs with advanced cellular technologies have a great potential to achieve beyond visual line of sight (BVLOS) UAV operations [6], [7], [8], [9]. Meanwhile, cellular networks can also provide ubiquitous coverage and low latency communications from base stations (BSs) with wireless backhauls [10], [11], [12]. In fact, even at the high attitude, High-definition (HD) video collected by UAVs can be readily transmitted to BSs within few milliseconds, by leveraging the 5G high-speed wireless infrastructure [6].

Despite the growing interest in cellular-connected UAVs, there are still many challenges for aerial video streaming. First, higher and smoother data rates are required to improve the QoE of users. Transmission failure may occur with high quality video due to the dramatic fluctuation of the air-to-ground (A2G) channel quality during the flight. In [13], a testbed was developed to measure the practical performance of video streaming from a UAV to a ground BS through Long Term Evolution (LTE). To tackle the dynamic fluctuation of channel quality, Deep Reinforcement Learning (DRL) and the additive variation bitrate method were adopted in [14] for dynamic video resolution selection based on network observations and video playback states. The authors in [15] and [16] considered freezing time and video rate for UAV

1536-1276 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

enabled video streaming to maintain the user QoE, where video smoothness are ignored in the quality measure.

Second, intelligent path planning is an important issue which provides a new degree of freedom for UAVs to improve the streaming performance, especially for more complex BVLOS remote surveillance scenarios. Note that many conventional trajectory designs of cellular-connected UAVs operate in the two-dimensional (2D) space with the assumption that the UAV flies at a sufficiently high altitude [17], [18], [19], which may not be practical since the altitude affect both the channel quality and camera coverage. In complex environments, such as urban areas, flexible three-dimensional (3D) trajectory design with obstacle avoidance is critical for efficient communication and safe UAV operations.

Third, practical environment-aware UAV communications are required. Most existing studies rely on strong assumptions such as an isotropic BS antenna pattern and probabilistic line-of-sight (LoS) channel models, or even the simple LoS links [19], [20], [21], where conventional convex optimization techniques are then utilized to tackle the problem. However, such over-simplified antenna and statistical channel assumptions only depict the A2G channel in an average sense, and the practical communication performance may be degraded due to the lack of site- or location-specific radio propagation information. In fact, the UAV may easily observe that the existence of an LoS link with environment-aware devices (e.g., Radar or camera) by actually checking the blockage situation by buildings.

To address the above challenges, in this paper, we investigate the QoE maximization problem for aerial video streaming over 3D cellular networks with building avoidance in complex urban environments. The measure of QoE is improved by introducing video smoothness, where a tradeoff between video quality and rate variation is captured. In addition, the vertical trajectory of the UAV is exploited in building avoidance to further improve the streaming performance. Different from the typical statistical channel based optimization, we store and utilize the physical environment information in the form of, e.g., a channel knowledge map (CKM) [22] or radio map [23], [24], which provides a site-specific database consisting of rich information about location-specific channels (e.g., channel gains) for enabling environment-aware communications. Such information can be acquired by offline simulation or online/offline measurements in advance [22]. The main contributions of this paper are summarized as follows:

- First, we develop a framework for aerial video streaming over 3D cellular networks, where a UAV delivers videos to ground BSs in an urban environment. The total QoE for video streaming is maximized by jointly optimizing the 3D UAV trajectory, transmission scheduling, and video playback rate adaption, where building avoidance and environment-aware communications are considered.
- Second, unlike the typical statistical model based optimization in prior works, we first assume a scenario when a perfect environment and channel knowledge map (ECKM) is available. Based on the map, graph algorithms are proposed to obtain the UAV path while an iterative parameter-enabled block coordinate descent (PBCD) algorithm is proposed to determine the playback rate and time allocation along the UAV path.
- Third, we consider the more challenging scenario without a perfect ECKM, and propose a dueling Deep Q-learning (DQL) solution for UAV path design. Specifically, a dueling deep Q-network (DQN) model is utilized and we only require information about rate measurement and sensing the physical environment for efficient path learning, where ECKM related information is also constructed for simulated experience generation to facilitate the learning process.
- Finally, extensive simulation results are provided to validate the effectiveness of the proposed algorithms. Significant performance gains over baseline schemes are achieved and the tradeoff between video quality and rate variation is also demonstrated.

The remainder of this paper is organized as follows. Section II presents the related work. Section III presents the system model and problem formulation. Section IV and Section V describe the details of the ECKM based solution and the DQL solution without a perfect ECKM, respectively. Section VI presents simulation results and Section VII concludes this paper.

II. Related Work

In this section, we introduce the related work on UAV-enabled video streaming and cellular-connected UAVs as well as radio map.

A. UAV-Enabled Video Streaming

In the literature, there are a flurry of recent works on UAV enabled video streaming. In [25], the minimum peak signal-to-noise ratio (PSNR) among users was maximized with UAVs, which served as aerial BSs in a pseudo-analog video broadcast system. By leveraging UAVs for providing computation and communication services, the authors in [26] proposed a UAV assisted mobile edge computing network for 360° video virtual reality (VR) applications, where the QoE for mobile VR users were maximized. By using reinforcement learning (RL), a UAV anti-jamming video transmission scheme with video compression selection was proposed in [27]. The work in [28] proposed a machine learning based scheduling solution to support UAV-based live omnidirectional (i.e., 360°) video streaming. A multi-group cell-free broadcast network was developed in [29] for real-time VR video transmission in a sports event, such that the experience of VR users can be enhanced. The authors in [15] studied the problem of bandwidth and transmit power allocation in UAV relay networks for multiuser video streaming, where the total long-term QoE of users are maximized.

The unstable transmission and heavy load problem were investigated for UAV-based streaming media transmission in [30], where a cross-layer design of routing, load balancing, and multi-link concurrent transmission were studied. The work in [31] employed a UAV as an aerial BS for multiuser video streaming, where the 3D aerial trajectory design and

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

resource allocation were jointly optimized. A statistical error-rate and delay bounded quality of service (QoS) provisioning scheme was proposed in [32] for capacity maximization over UAV-enabled 6G wireless networks. In summary, UAVs were leveraged as aerial communication platforms in the above work, such as relays or BSs [33], [34], to serve ground video users. Another emerging scenario is that UAVs are regarded as aerial users of cellular networks for providing aerial video services, such as video surveillance with cellular-connected UAVs [35].

B. Cellular-Connected UAVs

Cellular-connected UAVs have gained increasingly interest due to their BVLOS operations over wide areas. Furthermore, with the existing authentication mechanisms and high capacity cellular networks, cellular-connected UAVs are expected to achieve better security, reliability, and higher throughput for A2G communications. In [18], an integrated scheduling method of control, communication, and sensing was proposed for mmWave communications from UAV to ground BSs. The performance of two-hop 3D cellular networks was analyzed in [36] with realistic antenna patterns, where both UAVs and BSs coexist to serve ground users. A novel closed-loop control scheme was developed in [37] for open Radio Access Networks (RANs) to support UAV enabled video streaming over commercial 5G cellular networks. The authors in [38] developed an aerial video surveillance system in a noisy environment to perform monitoring of a set of points of interest at the city scale with multi-UAV relays.

Trajectory planning is another important research topic for aerial video services. The work in [39] investigated UAV enabled covert video surveillance with trajectory planning, where energy efficiency and visual covertness were considered. The UAV's energy consumption was minimized in [11] for aerial video surveillance under QoS constraints. The work in [40] studied the coverage-aware navigation problem with cellular-connected UAVs by minimizing the weighted sum of communication outage duration and completion time. However, only 2D trajectory design was considered in the above work where the UAV always flies at a sufficiently high altitude. The authors in [41] jointly optimized the data collection scheme and 3D flight trajectory over UAV-assisted Internet of Things (IoT) networks with a DRL technique. Radio map based 3D path planning for cellular-connected UAVs was proposed in [23] for mission completion time minimization with signal-to-interference-plus-noise ratio (SINR) constraints during the flight. In [42], dual-UAV-assisted secure communication was studied by jointly optimizing 3D trajectory and time switching allocation under energy constraints. However, the characteristics of video streaming have not been taken into account. Unlike the aforementioned studies, we study the QoE maximization problem of aerial video streaming in complex urban environments with 3D trajectory design and environment-aware communications.

C. Radio Map

Radio map, which contains rich information of the radio propagation environment for every location in a geographic area, has been employed for emerging UAV communications, spectrum management, interference control, and resource allocation [43], [44], [45]. In [43], a multi-UAV target searching scheme was proposed based on radio maps, which contain abundant information about received signal strength (RSS). A low complexity inter-cell interference coordination scheme was proposed in [44] for BSs equipped with massive-MIMO, where the coordination decision is made based on radio environment maps (REMs). The work in [45] proposed an A2G energy and information delivery framework based on binary channel feature map (CFM) by joint optimizing aerial platforms' positions and transmit parameters, where the binary CFM contains knowledge of the propagation environment in the flying area.

Although there are some research focusing on radio map construction from spatially distributed measurements, such measurements are performed at predetermined locations in general. A spectrum surveying scheme was proposed in [46], where UAVs were employed to collect radio measurements to construct a power map. The authors in [47] studied radio mapping with an efficient trajectory design for radio-aware UAVs, where performance enhancement were illustrated in terms of localization, connectivity, and sensing. The work in [48] proposed an emitter radio map disaggregation-based approach, where the emitters' radio maps are constructed by deep neural networks (DNNs). In [49], radio occupancy maps were estimated using DNNs, where the spatial structure of the propagation environment is learned from a dataset with measurements in other environments. Different from the above works, we employ an ECKM for environment-aware communications for aerial video streaming over complex urban environments, where both environment and channel information are utilized to facilitate 3D trajectory design, and the ECKM related information can also be constructed for simulated experience generation to facilitate the learning process.

Notations: In this paper, scalar variables and vectors are denoted by italic letters and boldface italic letters, respectively. In addition, $\|\cdot\|$ denotes the Euclidean norm and $[\cdot]^{\mathrm{T}}$ denotes transpose. $\mathbb{E}[\cdot]$ denotes the expectation. $\mathbb{R}^{m\times n}$ denotes the space of $m \times n$ real matrices, and $|X|$ denotes the cardinality of a set X.

III. System Model and Problem Statement

In this section, an aerial video streaming scenario is first introduced. Then, we present a 3D UAV flight model over urban areas, the A2G 3D communication model, and the video streaming model, based on which the optimization problem is formulated. The main notations used in this paper are summarized in Table I.

A. 3D UAV Flight Model Over Urban Areas

We consider an aerial video streaming scenario, where a rotary-wing UAV flies from an initial location to a destination

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

TABLE I
SUMMARY OF MAIN NOTATIONS AND DEFINITIONS

Notation	Definition
M	The number of terrestrial BSs
s_m	The m-th terrestrial BS
$\mathbf{g}_m$	The 3D coordinate of BS s_m
T	The total UAV operation time (s)
$\mathbf{q}(t)$	The 3D location of the UAV at time instant t
$\mathbf{v}(t)$	The UAV velocity at time t
$V_{\max}$	The maximum UAV speed (m/s)
$\mathbf{q}_s$, $\mathbf{q}_f$	The UAV's 3D location of the starting and ending positions
$H_{\min}$, $H_{\max}$	The minimum and maximum allowable altitude for the UAV
$\alpha_m(t)$	The indicator of transmission scheduling and association for BS s_m at time t
$\tilde{R}_m(t)$	The approximate achievable rate between the UAV and BS s_m at time t
$\tilde{r}$	The minimum required rate of video user
$r(t)$	The video playback rate for the video user at time t
$U(t)$	The QoE for the video user at time t
$\mathcal{F}$	The feasible region for UAV trajectory
$\mathcal{M}_m$	The ECKM for BS s_m
$\mathbf{M}_m$	The discretized ECKM for BS s_m
$\mathbf{R}$	The maximum rate map for the truncated region
$\beta_m(\mathbf{q})$	The large-scale channel gain between the UAV and BS s_m at location $\mathbf{q}$
$G_m(\mathbf{q})$	The antenna gain between the UAV and BS s_m at location $\mathbf{q}$
$\bar{R}^{\max}(\mathbf{q})$	The maximum achievable rate at location $\mathbf{q}$

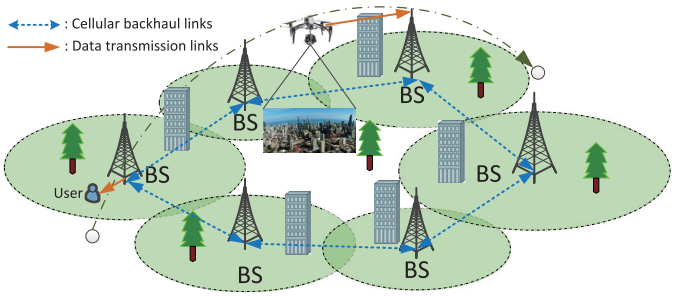


Fig. 1. Video transmission in a UAV-assisted surveillance application.

with an on-board camera capturing videos along the route. Denote $\mathbf{q}_s = [x_s, y_s, z_s]^{\mathrm{T}} \in \mathbb{R}^{3\times 1}$ and $\mathbf{q}_f = [x_f, y_f, z_f]^{\mathrm{T}} \in \mathbb{R}^{3\times 1}$ as the 3D location of the starting and ending positions, respectively. The video captured by the UAV is delivered to the ground BSs, which is then streamed to ground users through cellular backhaul links with high capacity. We assume that the capacity for cellular links from the BSs to ground video user are sufficient and the communication bottleneck for the video streaming only lies on the A2G links [10], [50]. For example, a ground VR user who takes VR helmets near the BS can share the views of the UAV along the flight, and would like to enjoy a journey with high QoE, as shown in Fig. 1. The cellular network consists of M terrestrial BSs, denoted by $\mathcal{M} = \{s_1, s_2, \ldots, s_M\}$. We consider an urban area $\mathcal{C} \subseteq \mathbb{R}^{3\times 1}$ with a large number of buildings, which may block not only the LoS communication links with the BSs, but also the UAV's moving path. Such physical environment is assumed to be sensed by the UAV (e.g., with Radar or camera) with a maximum sensing distance $d_{\max}$. The 3D coordinate of BS s_m is represented by $\mathbf{g}_m \in \mathbb{R}^{3\times 1}$. Let T denote the total UAV operation time. Then, the 3D location of the UAV at time instant t is denoted as $\mathbf{q}(t) = [x(t), y(t), z(t)]^{\mathrm{T}} \in \mathbb{R}^{3\times 1}, \forall t \in [0, T]$, where $\mathbf{q}(0) = \mathbf{q}_s$ and $\mathbf{q}(T) = \mathbf{q}_f$. Following the aerial regulations, the UAV should fly at an altitude ranging from $H_{\min}$ to $H_{\max}$, i.e., $H_{\min} \le z(t) \le H_{\max}, \forall t$. Furthermore, we define $\mathbf{v}(t) \triangleq \dot{\mathbf{q}}(t)$ as the UAV velocity at time t, and then we have $\|\mathbf{v}(t)\| \le V_{\max}, \forall t$, with $V_{\max}$ denoting the maximum UAV speed. We model the building height as a function $b(x, y)$ with respect to any horizontal coordinate $[x, y]^{\mathrm{T}} \in \mathbb{R}^{2\times 1}$ in the considered area $\mathcal{C}$. Thus, we obtain the feasible region for UAV trajectory $\mathbf{q}(t)$ as $\mathcal{F} = \{[x, y, z]^{\mathrm{T}} \in \mathcal{C} | \max\{H_{\min}, b(x, y)\} \le z \le H_{\max}\}$ such that building avoidance can be ensured.

B. 3D Communication Model

Denote $h_m(t)$ as the channel coefficient of A2G channel between the UAV and ground BS s_m at time t. Specifically, $h_m(t) = \sqrt{G_m(t)\beta_m(t)}\tilde{h}_m(t)$, where $G_m(t)$, $\beta_m(t)$, $\tilde{h}_m(t)$ account for the BS antenna gain, the large-scale channel gain, and the small-scale fading, respectively. The UAV is assumed to be equipped with a unit gain isotropic antenna. According to the 3GPP [51], each BS is equipped with directional antennas. Specifically, a uniform linear array (ULA) of 8 elements is placed vertically, while the sectorization technique is applied horizontally such that each BS consists of three sectors. Each array element is characterized by its half-power beamwidths ϕ_{3dB} and ψ_{3dB} along the horizontal and vertical dimensions, respectively, where $\psi_{3dB} = \phi_{3dB} = 65°$. Thus, the 3D antenna element gain A_e of BS s_m can be obtained with given vertical and horizontal angles $(\psi_m(t), \phi_m(t))$, where such angles depend on the UAV location $\mathbf{q}(t)$, i.e.,

$$\begin{aligned} A_e(\psi_m(t), \phi_m(t)) = G_{\max} \\ - \min\{-[A_{e,v}(\psi_m(t)) \\ + A_{e,h}(\phi_m(t))], A_v\}, \end{aligned} \tag{1}$$

where A_e consists of both vertical and horizontal radiation patterns and $G_{\max} = 8$ dBi is the maximum directional gain. $A_{e,v} = -\min\{12\left(\frac{\psi_m(t)-90}{\psi_{3dB}}\right)^2, SLA_v\}$, where $SLA_v = 30$ dB, denotes the side-lobe level limit. $A_{e,h} = -\min\{12\left(\frac{\phi_m(t)}{\phi_{3dB}}\right)^2, A_v\}$, where $A_v = 30$ dB, denotes the front-back ratio. Thus, the total antenna gain along the direction between the UAV and BS s_m at time t is $G_m(t) = A_e + AF$, where AF is the array factor given in [52].

According to the 3GPP model for urban Macro (UMa) [53], the A2G channel between the UAV and ground BS s_m at time t is an LoS channel when no obstacle exists between the UAV and BS s_m at time t, where the path loss measured in dB is

$$PL_m^L(t) = 28.0 + 22\log_{10}(d_m(t)) + 20\log_{10}(f_c), \tag{2}$$

with f_c denoting the carrier frequency and $d_m(t) \triangleq \|\mathbf{q}(t) - \mathbf{g}_m\|$ denoting the 3D distance between the UAV and BS s_m at time t. Otherwise, it is classified as an NLoS channel and the path loss is

$$\begin{aligned} PL_m^N(t) = -17.5 + (46 - 7\log_{10} z(t))\log_{10}(d_m(t)) \\ + 20\log_{10}(\frac{40\pi f_c}{3}), \end{aligned} \tag{3}$$

where $z(t)$ is the UAV's altitude at time t. Denote $c_m(t)$ as the channel blocking state at time t, with $c_m(t) = 1$ denoting

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

the LoS state while $c_m(t) = 0$ denoting the NLoS state. In practice, $c_m(t)$ can be exactly determined at location $\mathbf{q}(t)$ via checking whether the LoS path between the UAV and BS s_m is blocked or not. Therefore, we have

$$\beta_m(t) = c_m(t)10^{\frac{PL_m^L(t)}{10}} + (1 - c_m(t))10^{\frac{PL_m^N(t)}{10}}. \quad (4)$$

We assume that all BSs keeps active during the UAV's flight, since all BSs may be potentially associated with the UAV during its flight. Define $\alpha_m(t) \in \{0, 1\}$ as the indicator of transmission scheduling and association for BS s_m at time t. If the UAV decides to transmit video data to BS s_m at time t, then we have $\alpha_m(t) = 1$; otherwise, $\alpha_m(t) = 0$. We assume that at most one BS is scheduled for transmission at each time t, then we have $\sum_{m=1}^{M} \alpha_m(t) \leq 1, t \in [0, T]$. Denote $R_m(t)$ as the achievable rate between the UAV and BS s_m at time t, given by $R_m(t) = B \log_2 \left(1 + \frac{P|h_m(t)|^2}{\sigma^2}\right)$, where P is the UAV's maximum transmit power; B and σ^2 denote the channel bandwidth and noise power, respectively. Similar to [54], we assume that a sufficiently long channel code is employed such that the effect of small-scale fading has been averaged out, and then $R_m(t) \approx B \log_2 \left(1 + \frac{P\beta_m(t)G_m(t)}{\Gamma\sigma^2}\right) \triangleq \tilde{R}_m(t)$, where $\Gamma > 1$ reflects the gap between the theoretical Gaussian signaling and practical modulation-and-coding scheme.

C. Video Streaming Model

Suppose that dynamic adaptive streaming over HTTP (DASH) is adopted for video streaming and the video rate can be dynamically adapted to the channel condition [16], [55]. We assume that cellular users have sufficiently large playback buffers to store the received video data. Let $r(t)$ be the video playback rate for the video user at time t, who shares the UAV's view based on the videos transmitted from the UAV. Then we have $r(t) \geq \tilde{r}, \forall t$, where $\tilde{r}$ denote the minimum required rate of video user. QoE is adopted as the performance metric, which is mainly influenced by video quality, video rate variation, and rebuffering [55]. To avoid video rebuffering, the information-causality constraints have been introduced; i.e., at any time t, only the video data which has already been received can be played at user, i.e., $\int_0^t \sum_{m=1}^{M} \alpha_m(\tau)\tilde{R}_m(\tau)d\tau \geq \int_0^t r(\tau)d\tau, \forall t \in [0, T]$.

The video quality for video user at time t is evaluated through a logarithmic function with the property of diminishing returns [31], i.e., $\varpi \log \frac{\upsilon r(t)}{\tilde{r}}$, where ϖ and υ are positive constant parameters. We evaluate the video rate variation at time t as the squared difference between the current and time averaged playback rate, i.e., $|r(t) - \frac{1}{T}\int_0^T r(t)dt|^2$. As a result, we define the QoE at time t as

$$U(t) = \varpi \log \left(\frac{\upsilon r(t)}{\tilde{r}}\right) - \lambda |r(t) - \frac{1}{T}\int_0^T r(t)dt|^2, \quad (5)$$

where λ is a positive factor to balance the tradeoff between video quality and video rate variation. Note that video rate variation is negative since the QoE will be impaired by frequent changes of playback rate.

D. Problem Formulation

In this work, our objective is to maximize the total QoE over the 3D aerial cellular network by jointly optimizing the 3D UAV trajectory $\{\mathbf{q}(t)\}$, transmission scheduling $\{\alpha_m(t)\}$, as well as playback rate allocation $\{r(t)\}$, subject to the UAV's mechanical constraints as well as information-causality constraints. The optimization problem can be mathematically formulated as

$$\text{(P1)}: \max_{\{\mathbf{q}(t)\},\{\alpha_m(t)\},\{r(t)\}} \int_{t=0}^{T} U(t)dt$$

$$\text{s.t.} \quad \alpha_m(t) \in \{0, 1\}, \forall m, \forall t \in [0, T], \quad (6)$$

$$\sum_{m=1}^{M} \alpha_m(t) \leq 1, \forall t \in [0, T], \quad (7)$$

$$\int_0^t \sum_{m=1}^{M} \alpha_m(\tau)\tilde{R}_m(\tau)d\tau \geq \int_0^t r(\tau)d\tau, \forall t \in [0, T], \quad (8)$$

$$r(t) \geq \tilde{r}, \forall t \in [0, T], \quad (9)$$

$$\mathbf{q}(t) \in \mathcal{F}, \forall t \in [0, T], \quad (10)$$

$$\|\mathbf{v}(t)\| \leq V_{\max}, \forall t \in [0, T], \quad (11)$$

$$\mathbf{q}(0) = \mathbf{q}_s, \mathbf{q}(T) = \mathbf{q}_f. \quad (12)$$

In general, the optimal solution to (P1) is difficult to obtain due to the existence of binary constraints in (6) and the non-convex feasible region $\mathcal{F}$, as well as the complicated expression of $\tilde{R}_m(t)$. Since the generic path loss depends on the frequent variations of LoS/NLoS links and the practical BS antenna pattern consists of many non-continuous operations, $\tilde{R}_m(t)$ is not a continuous function generally, which is difficult to optimize.

To tackle the above challenges, we first consider the scenario with a perfect ECKM, where both location-specific environment and channel information are available prior to the UAV's flight, including the large-scale channel gain and antenna gain as well as the building height function $b(x, y)$. With ECKM, two graph algorithms with different complexities are proposed to obtain the UAV path, while an iterative PBCD algorithm is proposed to determine the playback rate and time duration allocation. Next, we consider the scenario without ECKM. We propose a dueling DQL solution for UAV path design, where the ECKM related information is also constructed for simulated experience generation to facilitate the learning process.

IV. Proposed Solution With Perfect ECKM

In this section, we first assume a scenario with a perfect ECKM, and then reformulate the problem into a more tractable form by analyzing the structure of the optimal solution, based on which graph algorithms are proposed to obtain the UAV path while an iterative PBCD algorithm is proposed to obtain the playback rate and time allocation along the UAV path.

With the ECKM, we can easily obtain the optimal transmission scheduling strategy of (P1) as $\alpha_{m^*}(t) = 1$ and $\alpha_{m'}(t) = 0, \forall m' \neq m^*$, where $m^* = \arg\max_{m \in \{1,2,\ldots,M\}} \tilde{R}_m(t)$. Let

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

$\tilde{R}^{\max}(t) \triangleq \max_{1\le m\le M} \tilde{R}_m(t)$. Then (P1) is simplified as

$$\begin{aligned}(\text{P2}): \max_{\{\mathbf{q}(t)\},\{r(t)\}} & \int_{t=0}^{T} U(t)dt \\ \text{s.t.} \quad & (9)-(12), \\ & \int_0^t \tilde{R}^{\max}(\tau)d\tau \ge \int_0^t r(\tau)d\tau, \forall t\in[0,T]. \end{aligned} \tag{13}$$

Note that $\tilde{R}^{\max}(t)$ depends on the location-specific channel knowledge $h_m(t)$, which is relevant to the UAV's 3D location $\mathbf{q}(t)$. As such, the location specific channel knowledge can be acquired from the CKM, such as the large-scale channel gain and antenna gain [22], [23], [24].

A. ECKM Based Reformulation

Different from CKM, in this paper, we construct an ECKM which provides not only channel knowledge but also environment information.

Definition 1: (ECKM): The ECKM for each BS s_m over the considered region $\mathcal{C}$ can be expressed as $\mathcal{M}_m = \{\mathfrak{M}_m(\mathbf{q})\}$. For any reachable location $\mathbf{q}\in\mathcal{F}$, $\mathfrak{M}_m(\mathbf{q}) = G_m(\mathbf{q})\beta_m(\mathbf{q})$. Otherwise, $\mathfrak{M}_m(\mathbf{q}) = -\infty$. $\beta_m(\mathbf{q})$ and $G_m(\mathbf{q})$ denote the large-scale channel gain and antenna gain between the UAV and BS s_m at location $\mathbf{q}$, respectively.

Due to definition 1, if $\mathfrak{M}_m(\mathbf{q}) = -\infty$, then we know that the location $\mathbf{q}$ is not a reachable location. Thus, the environment information about building distribution is also implied in the ECKM. In practice, the data used to construct an ECKM could be acquired via offline numerical simulation or online/offline measurements [22]. However, since the storage of each BS is limited, storing the entire data of ECKM for all locations is infeasible. As such, we consider a truncated 3D rectangle region, denoted by $\hat{\mathcal{C}}\subseteq\mathcal{C}$. In particular, the UAV's altitude ranges from $H_{\min}$ to $H_{\max}$, and the UAV's horizontal location can be assumed to be within a square region with a sufficiently large edge length L, e.g., $L \ge \sqrt{(x_f-x_s)^2+(y_f-y_s)^2}$, such that all possible horizontal locations during the flight can be covered. The region $\hat{\mathcal{C}}$ is discretized into a $D\times D\times Z$ grid with granularity Δ, where $D = \lceil \frac{L}{\Delta} \rceil$ and $Z = \lceil \frac{H_{\max}-H_{\min}}{\Delta} \rceil$, and there are ZD^2 grid points in total. Denote the set of all grid points as $\hat{\mathcal{C}}_G$. Let $\mathbf{q}_{i,j,k}$ be the (i,j,k)-th location in $\hat{\mathcal{C}}_G$, $i,j\in\{1,2,\ldots,D\}, k\in\{1,2,\ldots,Z\}$, then $\mathbf{q}_{i,j,k} = [i-\frac{1}{2}, j-\frac{1}{2}, k-\frac{1}{2}]^{\mathrm{T}}\Delta$. The grid points corresponding to the UAV's initial and final locations $\mathbf{q}_s$ and $\mathbf{q}_f$ are denoted as $\mathbf{g}_s$ and $\mathbf{g}_f$, respectively.

Definition 2: (Discretized ECKM): Based on the above 3D grid model, the discretized ECKM for each BS s_m is denoted by matrix $\mathbf{M}_m$. Let $[\mathbf{M}_m]_{i,j,k}$ be the (i,j,k)-th element in $\mathbf{M}_m$. Then $[\mathbf{M}_m]_{i,j,k} = G_m(\mathbf{q}_{i,j,k})\beta_m(\mathbf{q}_{i,j,k})$ if $\mathbf{q}_{i,j,k}\in\hat{\mathcal{C}}_G\bigcap\mathcal{F}\triangleq\mathcal{F}_G$; and $[\mathbf{M}_m]_{i,j,k} = -\infty$, otherwise.

The granularity Δ is chosen to be sufficiently small such that $G_m(\tilde{\mathbf{q}})\approx G_m(\mathbf{q}_{i,j,k})$, $\beta_m(\tilde{\mathbf{q}})\approx\beta_m(\mathbf{q}_{i,j,k}), \forall m$, for any 3D location $\tilde{\mathbf{q}}$ in the (i,j,k)-th grid cell. In this section, we assume that the perfect discretized ECKM $\{\mathbf{M}_m\}$ for all M BSs are available prior to the UAV's flight. In practice, such information can be downloaded off-line from the cellular network. As such, we can construct a so-called *maximum rate map (MRM)* that characterizes the maximum achievable rate $\tilde{R}^{\max}(t)$ at each grid point. Denote the achievable rate between the UAV and BS s_m when it is at location $\mathbf{q}_{i,j,k}$ as $\tilde{R}_m(\mathbf{q}_{i,j,k}) \triangleq B\log_2\left(1+\frac{PG_m(\mathbf{q}_{i,j,k})\beta_m(\mathbf{q}_{i,j,k})}{\Gamma\sigma^2}\right)$. Then the maximum achievable rate at $\mathbf{q}_{i,j,k}$ can be written as $\tilde{R}^{\max}(\mathbf{q}_{i,j,k}) = \max_m \tilde{R}_m(\mathbf{q}_{i,j,k})$.

Definition 3: (MRM): The MRM for the truncated region $\hat{\mathcal{C}}$ is denoted by a 3D matrix $\mathbf{R}\in\mathbb{R}^{D\times D\times Z}$, while the (i,j,k)-th element in $\mathbf{R}$ is $[\mathbf{R}]_{i,j,k} = \tilde{R}^{\max}(\mathbf{q}_{i,j,k})$ if $\mathbf{q}_{i,j,k}\in\mathcal{F}_G$; and $[\mathbf{R}]_{i,j,k} = 0$, otherwise.

In summary, the overall complexity of constructing the discretized ECKM $\{\mathbf{M}_m\}$ and MRM $\mathbf{R}$ over all locations in $\hat{\mathcal{C}}$ is given by $O(D^2ZM)$. Similarly, we can construct a *BS association map (BAM)* as $\mathbf{A}\in\mathbb{R}^{D\times D\times Z}$ that characterizes the optimal BS association at each grid point, where the (i,j,k)-th element in $\mathbf{A}$ can be calculated as $[\mathbf{A}]_{i,j,k} = \arg\max_m \tilde{R}_m(\mathbf{q}_{i,j,k})$ if $\mathbf{q}_{i,j,k}\in\mathcal{F}_G$, and $[\mathbf{A}]_{i,j,k} = 0$ otherwise.

With the given MRM $\mathbf{R}$, we adopt the path disretization technique [56] to make (P2) more tractable. Specifically, we divide the UAV path into N line segments with $N+1$ adjacent grid points, and approximate the UAV trajectory as the grid points $\{\mathbf{q}_n\}$ and time duration $\{\tau_n\}$, where τ_n represents the time duration along line segment between $\mathbf{q}_n$ and $\mathbf{q}_{n+1}$, $\frac{\Delta}{V_{\max}} \le \tau_n, 1\le n\le N$. With the 3D grids, the moving direction along the n-th line segment $\vec{\mathbf{v}}_n$ can be selected from the finite set $\mathcal{V} = \{[0,0,1]^{\mathrm{T}}, [0,0,-1]^{\mathrm{T}}, [0,1,0]^{\mathrm{T}}, [0,-1,0]^{\mathrm{T}}, [1,0,0]^{\mathrm{T}}, [-1,0,0]^{\mathrm{T}}\}$. The UAV can take turns along two consecutive line segments if their directions are different with sufficiently large UAV acceleration. The disretized form of $\tilde{R}_m(t)$, $\tilde{R}^{\max}(t)$, $r(t)$, $U(t)$ are denoted as $\tilde{R}_{mn}$, $\tilde{R}_n^{\max}$, r_n, U_n, respectively. In particular, $\tilde{R}_{mn}$ is approximated by the achievable rate for BS s_m at location $\mathbf{q}_n$, $\tilde{R}_n^{\max} = \max_m \tilde{R}_{mn}$, and $U_n = \varpi\log\frac{vr_n}{\tilde{r}} - \lambda|r_n - \frac{1}{T}\sum_{n=1}^{N} r_n\tau_n|^2$. The optimization problem (P2) can be discretized as:

$$\begin{aligned}(\text{P3}): \max_{N,\{\mathbf{q}_n\},\{\tau_n\},\{r_n\},\{\vec{\mathbf{v}}_n\}} & \sum_{n=1}^{N} U_n\tau_n \\ \text{s.t.} \quad & \mathbf{q}_n\in\mathcal{F}_G, \vec{\mathbf{v}}_n\in\mathcal{V}, \forall n, && (14)\\ & \mathbf{q}_{n+1} = \mathbf{q}_n + \Delta\vec{\mathbf{v}}_n, \forall n, && (15)\\ & \mathbf{q}_1 = \mathbf{g}_s, \mathbf{q}_{N+1} = \mathbf{g}_f, && (16)\\ & \sum_{n'=1}^{n}\tilde{R}_{n'}^{\max}\tau_{n'} \ge \sum_{n'=1}^{n} r_{n'}\tau_{n'}, \forall n, && (17)\\ & r_n\ge\tilde{r}, \forall n, && (18)\\ & \Delta/V_{\max}\le\tau_n, \forall n, && (19)\\ & \sum_{n=1}^{N}\tau_n = T. && (20)\end{aligned}$$

Although MRM $\mathbf{R}$ is given, (P3) is still a mixed-integer non-convex optimization problem.

B. Proposed Solution

With the feasible grid set $\mathcal{F}_G$, we can construct a graph $G(V,E)$ as follows:

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

Definition 4: Graph $G(V,E)$ is defined as: $V \triangleq \{\mathbf{q}|\mathbf{q} \in \mathcal{F}_G\}$, and $E \triangleq \{(\mathbf{q},\mathbf{q}')|\,\|\mathbf{q}-\mathbf{q}'\| = \Delta, \mathbf{q} \in V, \mathbf{q}' \in V\}$.

Thus, each feasible UAV path to (P3) corresponds to a path from $\mathbf{g}_s$ to $\mathbf{g}_f$ in graph G. Due to (19) and (20), we have $N \leq \frac{TV_{\max}}{\Delta}$. Note that all the paths from $\mathbf{g}_s$ to $\mathbf{g}_f$ in G with $N \leq \frac{TV_{\max}}{\Delta}$ can be obtained by using existing graph algorithms, e.g., the depth-first search (DFS) method with complexity $O(|V|!)$ [57]. If no such path is found, then (P3) will be infeasible. Note that for each feasible path from $\mathbf{g}_s$ to $\mathbf{g}_f$ in G, the UAV path $\{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}\}$ is exactly determined, and then the maximum rate $\{\tilde{R}_n^{\max}\}$ along such path is also determined. As a result, the optimal solution to (P3) can be obtained by searching all such paths and selecting the one with the maximum objective value. Thus, in the following, we only consider QoE maximization with given $\{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}, \{\tilde{R}_n^{\max}\}\}$, which is written as

$$\begin{aligned}(\text{P4}): &\max_{\{\tau_n\},\{r_n\}} \sum_{n=1}^{N} U_n \tau_n \\ \text{s.t.} \quad &(17)-(20).\end{aligned}$$

By introducing slack variables $\{y_n\}$, we obtain

$$\begin{aligned}(\text{P5}): &\max_{\{\tau_n\},\{r_n\},\{y_n\}} \sum_{n=1}^{N} (\varpi \log \frac{\upsilon r_n}{\tilde{r}} - y_n)\tau_n \\ \text{s.t.} \quad &(17)-(20), \\ &\lambda |r_n - \frac{1}{T}\sum_{n=1}^{N} r_n \tau_n|^2 \leq y_n, \forall n, \end{aligned} \tag{21}$$

which is equivalent to (P4) since y_n can always be decreased to achieve a larger objective value if the strict inequality holds in (21). By partitioning the optimization variables into two sets, i.e., $\{\{r_n\},\{y_n\}\}$ and $\{\tau_n\}$, we propose an efficient PBCD algorithm for solving problem (P5). Specifically, with any given $\{\tau_n\}$, (P5) can be reduced to

$$\begin{aligned}(\text{P6}): &\max_{\{r_n\},\{y_n\}} \sum_{n=1}^{N} (\varpi \log \frac{\upsilon r_n}{\tilde{r}} - y_n)\tau_n \\ \text{s.t.} \quad &(17), (18), (21),\end{aligned}$$

which is a convex optimization problem and thus can be effectively solved by existing solvers such as CVX [58]. With any given $\{\{r_n\},\{y_n\}\}$, (P5) can be reduced to

$$\begin{aligned}(\text{P7}): &\max_{\{\tau_n\}} \sum_{n=1}^{N} (\varpi \log \frac{\upsilon r_n}{\tilde{r}} - y_n)\tau_n \\ \text{s.t.} &(17), (19)-(21),\end{aligned}$$

which is also a convex optimization problem. The conventional block coordinate descent (BCD) method [54] alternately optimizes (P6) and (P7) in each iteration until convergence is achieved, which, however, may fail to update the time duration $\{\tau_n\}$ effectively. Specifically, with any given $\{\tau_n\}$ which satisfy (20), since (P6) aims to increase QoE by decreasing y_n, then equality holds in (21) after solving (P6) in each iteration. As a result, for optimizing the time duration in (P7), τ_n needs to be increased to improve the total QoE, and the optimization freedom for τ_n is severely limited due to the equality constraints (20) and (21), which may cause ineffective update of time duration in each iteration.

Algorithm 1 PBCD Algorithm for Solving Problem (P5)

1: Initialize $\rho_{ini} = \frac{N\Delta}{V_{\max}T}$ and $I_{\max}$. Let $l = 0$, $\rho^l = \rho_{ini}$, and $\rho_{step} = \frac{1-\rho_{ini}}{I_{\max}}$.
2: **repeat**
3: Solve the convex problem (P6) with given $\{\tau_n\}^l$, and denote the optimal solution as $\{\{r_n\}^{l+1}, \{y_n\}^{l+1}\}$;
4: $\rho^{l+1} = \min\{1, \rho^l + \rho_{step}\}$;
5: Solve the convex problem (P8) with given $\{\{r_n\}^{l+1}, \{y_n\}^{l+1}, \rho^{l+1}\}$, and denote the optimal solution as $\{\tau_n\}^{l+1}$;
6: Update $l = l + 1$;
7: **until** The fractional increase of the objective value is below a threshold κ;

To tackle such an issue, we propose a new PBCD method. It can be proved by contradiction that the optimal solution to (P5) is the same as that with a relaxed constraint $\sum_{n=1}^{N} \tau_n \leq T$. A temporary parameter $\rho \leq 1$ is employed on the right-hand-side of the above relaxed constraint and we obtain

$$\begin{aligned}(\text{P8}): &\max_{\{\tau_n\}} \sum_{n=1}^{N} (\varpi \log \frac{\upsilon r_n}{\tilde{r}} - y_n)\tau_n \\ \text{s.t.} \quad &(17), (19), (21), \\ &\sum_{n=1}^{N} \tau_n \leq \rho T.\end{aligned} \tag{22}$$

The main idea of PBCD is to solve (P8) with the newly introduced temporary parameter $\rho \leq 1$ instead of directly solving (P7). Specifically, ρ is gradually increased before solving problem (P8) in each iteration, until $\rho = 1$ is achieved. Then, the constraint (22) is relaxed after each iteration due to the increase of ρ, which thus permits a more effective time duration update compared to the conventional BCD. We summarize the details of the PBCD method in Algorithm 1. Denote $I_{\max}$ as the number of iterations required from the initial parameter ρ_{ini} to 1. Due to (19) and (22), we have $\rho \geq \frac{N\Delta}{V_{\max}T}$. Thus, the initial parameter is set as $\rho_{ini} = \frac{N\Delta}{V_{\max}T}$.

Theorem 1: With Algorithm 1, the sequence of objective values is monotonically non-decreasing with iteration number l, and thus Algorithm 1 is convergent.

Proof: Let $\{\{r_n\}^l, \{y_n\}^l, \{\tau_n\}^l\}$ be the solutions obtained in iteration l and the resulting objective value is η^l. In iteration $l+1$, the optimal solution to (P6) is first obtained with given $\{\tau_n\}^l$, and we have $\eta_{(P6)}^{l+1} \geq \eta^l$. Then, the relaxed problem (P8) is optimized with the given $\{\{r_n\}^{l+1}, \{y_n\}^{l+1}, \rho^{l+1}\}$ and constraint (22) is further relaxed after Step 4. The optimal solution to (P6) is always a feasible solution to (P8), and (P8) yields an objective value that improves upon that obtained in (P6), i.e., $\eta^{l+1} = \eta_{(P8)}^{l+1} \geq \eta_{(P6)}^{l+1}$. By combining the above arguments, we have $\eta^{l+1} \geq \eta^l$. Moreover, since the objective function is upper bounded, Algorithm 1 is convergent. ■

In Algorithm 1, (P5) is solved by solving a sequence of convex optimization problems (P6) and (P8) using the convex solver

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2 DFS Based Algorithm for Solving Problem (P3)

1: Construct a grid graph $G(V, E)$ based on Definition 4 and initialize $\eta^* = 0$;
2: Find all paths $\mathcal{P}$ from $\mathbf{g}_s$ to $\mathbf{g}_f$ over graph $G(V, E)$ with the DFS method;
3: **for** each path $\vec{p} \in \mathcal{P}$ **do**
4: Obtain $\{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}, \{R_n^{\max}\}\}$ based on $\vec{p}$ and MRM $\mathbf{R}$;
5: Solve problem (P5) with Algorithm 1 to obtain objective value η and solution $\{\{\tau_n\}, \{r_n\}\}$;
6: **if** $\eta > \eta^*$ **then**
7: $\eta^* = \eta$; $\{N^*, \{\mathbf{q}_n\}^*, \{\vec{\mathbf{v}}_n\}^*, \{\tau_n\}^*, \{r_n\}^*\} = \{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}, \{\tau_n\}, \{r_n\}\}$;
8: **end if**
9: **end for**

based on the interior-point method, and hence the overall complexity of Algorithm 1 is $O(N^{3.5}\log^2\frac{1}{\kappa})$ [54], where $\kappa > 0$ is the required accuracy. Thus, the algorithm for solving problem (P3) can be given as Algorithm 2.

Note that finding all paths from $\mathbf{g}_s$ to $\mathbf{g}_f$ could be time consuming. In the following, a low-complexity heuristic algorithm is proposed for (P3) based on a weighted graph. Intuitively, we would like to find a path with a larger QoE along that path. Considering the special case of (P3) with $\tilde{R}_n^{\max} = r_n, \forall n$, then the constraint (17) is always satisfied and we have $\tilde{R}_n^{\max} \geq \tilde{r}, \forall n$, due to (18). Then, U_n can be approximated as $U_n \approx \varpi \log \frac{v\tilde{R}_n^{\max}}{\tilde{r}} - \lambda|\tilde{R}_n^{\max} - \tilde{R}_{n-1}^{\max}|^2 \triangleq \tilde{U}_n$, where quality variation is approximated as $|\tilde{R}_n^{\max} - \tilde{R}_{n-1}^{\max}|^2$. Denote $\tilde{R}^{\max}$ as the maximum rate among MRM $\mathbf{R}$, i.e., $\tilde{R}^{\max} \triangleq \max\{\tilde{R}^{\max}(\mathbf{q}_{i,j,k}) | \mathbf{q}_{i,j,k} \in \mathcal{F}_G\}$. A weighted graph $\hat{G}(\hat{V}, \hat{E})$ is constructed as follows:

Definition 5: Graph $\hat{G}(\hat{V}, \hat{E})$ is defined as: $\hat{V} \triangleq \{\mathbf{q} | \mathbf{q} \in \mathcal{F}_G, \tilde{R}^{\max}(\mathbf{q}) \geq \tilde{r}\}$, $\hat{E} \triangleq \{(\mathbf{q}, \mathbf{q}') | \ \|\mathbf{q} - \mathbf{q}'\| = \Delta\}$. $w : E \to \mathbb{R}$, *where w is a weight function. In particular,* $w(\mathbf{q}, \mathbf{q}') = \varpi \log \frac{v\tilde{R}^{\max}}{\tilde{r}} - \varpi \log \frac{v\tilde{R}^{\max}(\mathbf{q})}{\tilde{r}} + \lambda|\tilde{R}^{\max}(\mathbf{q}) - \tilde{R}^{\max}(\mathbf{q}')|^2$. Note that $|\hat{V}| \leq |V|$ and each edge weight is non-negative, where a smaller weight corresponds to a larger QoE along that edge. Inspired by such observation, we find the shortest weighted path over $\hat{G}$ by utilizing Dijistra Algorithm [57] with complexity $O(|\hat{V}|^2)$, and apply Algorithm 1 to optimize time duration and playback rate allocation. The details are summarized in Algorithm 3. Note that MRM $\mathbf{R}$ is given in advance with the perfect ECKM, which is fully utilized in the graph based algorithms. When the perfect ECKM is unavailable, then MRM $\mathbf{R}$ is unknown in advance, which makes (P3) more difficult to solve. To tackle such issue, a DQL based solution with MRM mapping is proposed in the following section, where no perfect ECKM is required. Thus, the performances of the proposed algorithms with perfect ECKM serve as upper bounds for that of the DQL based solution without such information. The flow chart structure of the proposed solutions is shown in Fig. 2.

V. Proposed Solution Without Perfect ECKM

In this section, we consider the more practical scenario when the perfect ECKM is unavailable. By reformulating the problem as an MDP, a model-free DRL algorithm is proposed, where the UAV intelligently makes the most advantageous decisions by only local observations of environment and sampled rate measurements. In addition, the ECKM related information is also constructed for simulated experience generation to facilitate the learning process.

Algorithm 3 Weighted Graph Algorithm for Solving (P3)

1: Construct a weighted graph $\hat{G}(\hat{V}, \hat{E})$ as in Definition 5;
2: Find the shortest weighted path $\vec{p}$ from $\mathbf{g}_s$ to $\mathbf{g}_f$ over graph $\hat{G}(\hat{V}, \hat{E})$ with Dijistra Algorithm;
3: Obtain $\{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}, \{\tilde{R}_n^{\max}\}\}$ based on $\vec{p}$ and MRM;
4: Solve problem (P5) with Algorithm 1 to obtain solution $\{\{\tau_n\}, \{r_n\}\}$;

A. Problem Reformulation

By discretizing the airspace into 3D grids, the original problem (P2) is discretized as in (P3). Different from Section IV, in this section, $\{\tilde{R}_n^{\max}\}$ is unavailable due to the lack of ECKM prior the flight. However, the UAV can sense the surrounding physical environment within distance $d_{\max}$ and perform rate measurements during the flight, which can be used to learn the optimized solution. Inspired by Algorithm 3, we decompose problem (P3) into two subproblems, i.e., UAV path optimization, as well as time duration and playback rate allocation. With a given UAV path $\{N, \{\mathbf{q}_n\}, \{\vec{\mathbf{v}}_n\}\}$ as well as measured $\{R_n^{\max}\}$ along such path, the time duration and playback rate allocation can be obtained by solving (P5) with Algorithm 1. Thus, in the following, we will focus on the UAV path optimization problem, given by

$$\text{(P9)}: \max_{N, \{\mathbf{q}_n\}\{\vec{\mathbf{v}}_n\}} \sum_{n=1}^{N} \tilde{U}_n$$
$$\text{s.t.} \quad (14) - (16),$$
$$\tilde{R}_n^{\max} \geq \tilde{r}, \forall n, \tag{23}$$
$$N \leq TV_{\max}/\Delta, \tag{24}$$

where we maximize $\sum_{n=1}^{N} \tilde{U}_n$ by omitting the time duration and leave it in the time duration optimization subproblem (P5). Constraint (23) is imposed to ensure (17) and (18), while (24) is imposed to ensure (19) and (20). In practice, $\tilde{R}_n^{\max}$ can be measured at $\mathbf{q}_n$ by adopting an existing handover mechanism with continuous reference signal received power (RSRP) measurements [40]. Recall that the physical environment can be sensed by the UAV with a maximum sensing distance d_{max}. Thus, the UAV can determine whether a 3D location within its sensing range is feasible location or not during its flight. As a result, problem (P9) can be modeled as an MDP $(\mathcal{S}, \mathcal{A}, \mathcal{R})$, where $\mathcal{S}$ is the *state space* which consists of all possible states of the UAV at each time slot; $\mathcal{A}$ is the *action space* which consists of the UAV's available actions at each time slot; $\mathcal{R}$ is the *reward function* which maps the UAV's state spaces as well as action spaces at the current time slot to its expected reward. In particular, the UAV is treated as an agent that interacts with the environment. At each time step n, the UAV observes state $\mathbf{q}_n$; based on which it chooses the action $\vec{\mathbf{v}}_n$. Then the state

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

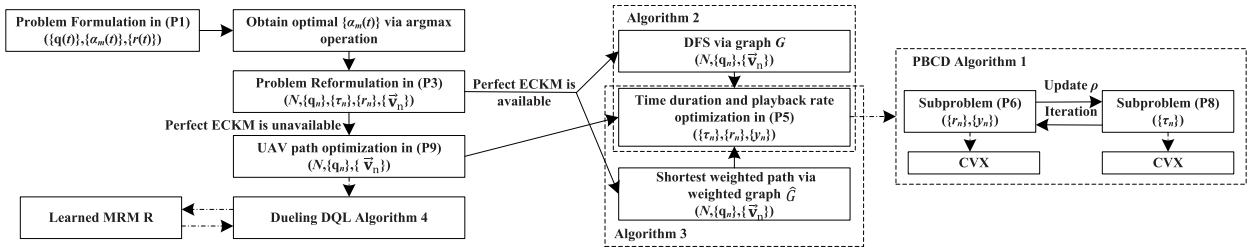


Fig. 2. A flow chart for illustration of the proposed solutions.

turns into $\mathbf{q}_{n+1}$ and the UAV receives reward $\hat{R}_n$ based on the calculation of $\tilde{U}_n$ and the rate measurement rate $\tilde{R}^{\max}_{n+1}$. In the following, we define the above elements.

- *State*: We define the state space by the feasible 3D location of UAV as $\mathcal{S} = \mathcal{F} \subseteq \mathbb{R}^{3\times 1}$, where the state of the UAV at time step n is defined as $\mathbf{q}_n \in \mathcal{S}$. $\mathbf{q}_s$ is the initial state while $\mathbf{q}_f$ is the final state.
- *Action*: The UAV at state $\mathbf{q}_n$ can choose an action $\vec{\mathbf{v}}_n$ from the action space at time step n. By dividing the airspace into 3D grids, the action space can be defined as $\mathcal{A} = \mathcal{V}$, and the states of the UAV are finite district grids.
- *Reward function*: We define a reward function $\mathcal{R} = \{\hat{R}_n\}$, which awards the UAV for reaching its destination with a larger $\sum_{n=1}^{N} \tilde{U}_n$, and penalizes the UAV for moving or violating constraints (14) and (23), i.e.,

$$\hat{R}_n = \begin{cases} \hat{R}_{des}, & \mathbf{q}_{n+1} = \mathbf{q}_f, \\ -\hat{R}_{out}, & \mathbf{q}_{n+1} \notin \mathcal{F}, \\ \mu_1 \tilde{U}_n - \mu_2 I_{n+1} - \mu_3, & \text{otherwise}. \end{cases} \tag{25}$$

where $I_{n+1} = 1$ when $\tilde{R}^{\max}_{n+1} < \tilde{r}$, and $I_{n+1} = 0$ otherwise. μ_1, μ_2, and μ_3 are positive weighting factors.

In the following, we propose a dueling DQL solution with environment information mapping to construct MRM, which can be used for simulated experience generation in DQL to reduce the required actual agent-environment interactions. Our proposed DQL consists of an offline training process and an online execution process. During the training process, we collect training data and train the DQN models offline. After the training process, we execute the well-trained DQN models to learn the optimized flying strategy for the UAV according to its current state.

B. Proposed Solution

Denote $Q_\zeta(s,a)$ as the Q-value of the UAV, which is defined as the accumulated reward when the UAV takes action a in state s and follows its policy ζ afterwards, i.e., $Q_\zeta(s,a) = \mathbb{E}_\zeta[\sum_{k=0}^{\infty} \gamma^k \hat{R}_{n+k} | \mathbf{q}_n = s, \vec{\mathbf{v}}_n = a]$, where γ is the discount factor reflecting the importance of immediate and future rewards. In (P9), the objective function corresponds to $\gamma = 1$, which means that all rewards are equally important. At each state s, the optimal policy of the UAV is to select the action that can maximize its Q-value, i.e., $\zeta^*(s) = \arg\max_{a\in\mathcal{A}} Q_\zeta(s,a)$. As a result, we should specify the Q-function $Q_\zeta(s,a)$ to find the optimal policy $\zeta^*(s)$. Due to the infinite state-action space, it is impossible to obtain exact Q-functions for the UAV. Instead, in the following we adopt a DQN with the dueling network architecture [59], also named dueling DQN, to approximate the state-action value $Q(s,a)$ with weights $\boldsymbol{\theta}$.

The proposed algorithm employs the following techniques for stabilizing and improving the performance of the neural network. *Experience replay buffer*: Replay buffer $\mathcal{B}$ is utilized to store the transitions for supporting the neural network to overcome potential instability. When $\mathcal{B}$ is full, a mini-batch of samples is randomly selected for training the networks. Due to the finite buffer size of $\mathcal{B}$, the content in buffer $\mathcal{B}$ is always up-to-date and then the neural networks can learn from the new samples. *Dueling neural network*: To improve the convergence rate and stability, the dueling neural network is employed [59], which combines two streams of the advantage function and value function. In particular, the Q-function can be obtained by combining the two streams' outputs as

$$Q(\mathbf{q},\vec{\mathbf{v}};\boldsymbol{\theta},\boldsymbol{\theta}_A,\boldsymbol{\theta}_V) = \mathcal{V}(\mathbf{q};\boldsymbol{\theta}_V) + \left(\mathcal{G}(\mathbf{q},\vec{\mathbf{v}};\boldsymbol{\theta}_A) - \frac{1}{\mathcal{A}}\sum_{\vec{\mathbf{v}}'} \mathcal{G}(\mathbf{q},\vec{\mathbf{v}}';\boldsymbol{\theta}_A)\right), \tag{26}$$

where $\mathcal{V}(\mathbf{q};\boldsymbol{\theta}_V)$ is the value function estimated by one of the streams with a fully-connected layer. Advantage function $\mathcal{G}(\mathbf{q},\vec{\mathbf{v}};\boldsymbol{\theta}_A)$ depicts the importance of each state related action, and $\mathcal{G}(\mathbf{q},\vec{\mathbf{v}}';\boldsymbol{\theta}_A)$ is output by the other stream, with $\boldsymbol{\theta}_A$ and $\boldsymbol{\theta}_V$ denoting the parameters of the corresponding networks. *Target networks*: To stabilize the training, a target network Q' with parameters $\boldsymbol{\theta}'$ is employed for target value estimation [60]. $\boldsymbol{\theta}'$ will be updated by the parameters of original neural network Q, i.e., $\boldsymbol{\theta}$, after a number of iterations. The update of neural network parameters are performed by minimizing the loss function defined as

$$\mathbb{L}(\theta) = \mathbb{E}_{\mathbf{q},\vec{\mathbf{v}},\hat{R},\mathbf{q}'}\left[\left(y^{DQL} - Q(\mathbf{q},\vec{\mathbf{v}};\boldsymbol{\theta},\boldsymbol{\theta}_A,\boldsymbol{\theta}_V)\right)^2\right], \tag{27}$$

where $y^{DQL} = \hat{R} + \max_{\vec{\mathbf{v}}'\in\mathcal{A}} Q'(\mathbf{q}',\vec{\mathbf{v}}';\boldsymbol{\theta}',\boldsymbol{\theta}_A,\boldsymbol{\theta}_V)$.

Environment Information Mapping: To reduce the number of agent-environment interactions such as location determination, obstacle sensing, and rate measurements, environment information mapping is employed. In particular, the UAV can only measure the rate of the location where it actually visits. Note that it is a supervised learning problem to predict the rate $\tilde{R}^{\max}(\mathbf{q}), \forall \mathbf{q} \in \mathcal{C}$ with any finite measurements $\{\mathbf{q}_n, \tilde{R}^{\max}(\mathbf{q}_n)\}$. A feedforward fully-connected artificial neural network (ANN) with parameters $\boldsymbol{\xi}$ is adopted for MRM estimation. The obtained $\tilde{R}^{\max}(\mathbf{q})$ based on rate measurement can not only be used as new input data to improve the MRM

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

estimation, but also be used for determining the next UAV flying direction $\vec{\mathbf{v}}$. Although the estimated MRM $\mathbf{R}$ might be initially inaccurate, its accuracy will be improved with more accumulated real experiences, based on which the UAV can predict the estimated return for each trajectory it would take, without actual agent-environment interactions. Therefore, more simulated experiences can be generated based on $\mathbf{R}$, where the UAV can update its policy by using both real experience and simulated experience.

We summarize the details of the learning algorithm in Algorithm 4. We would like to clarify that Algorithm 4 is a learning algorithm, which can be performed at the ground station with the measured rate acquired from the UAV. Once the DQN is well trained and properly loaded to the mission UAV, the mission UAV only needs to select the best action as shown at Step 5. To further reduce the actual UAV flight requirement, the system may first pre-train the policy offline with the offline collected rate measurements, or with simulation-generated rate samples based on the available knowledge on the radio propagation environment, and then further refine the policy by actual UAV flight with online learning. This thus alleviates the requirement for actual UAV flight, and renders the proposed algorithm more practical for training. On the other hand, our paper focus on designing a learning algorithm specifically for the given scenario, where the generalization ability for different cities will be left as future work. In Algorithm 4, $\Psi^{\max}$ denotes the maximum number of episodes, and $N_{\max} = \lfloor \frac{TV_{\max}}{\Delta} \rfloor$. At step 5, the UAV selects the optimal action with probability $1-\epsilon$ and chooses a random action with probability ϵ. Algorithm 4 adopts the MRM learning operations and the estimated MRM is updated in step 10 by using the actual rate measurement data, where MRM is also used in step 12 for simulated experience generation. For each actual UAV experience, $\tilde{N}$ steps will be taken in the simulated trajectory. Regardless of real or simulated experience, the same action-value update based on dueling DQL is performed, which is guaranteed to converge [59]. The complexity of Algorithm 4 is given by $O(\Psi^{\max} N_{\max}^2 \tilde{N})$.

Algorithm 4 The Dueling DQL Algorithm for Problem (P9) With MRM Mapping

1: Initialize the neural network Q and the target network Q' with the parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, respectively; Initialize the replay memory pool $\mathcal{B}$;
2: **for** episode $= 1$ to $\Psi^{\max}$ **do**
3: Randomly initialize the state $\mathbf{q}_1$. Set $n = 1$;
4: **while** $\mathbf{q}_n \neq \mathbf{q}_f$ and $n \leq N_{\max}$ **do**
5: Obtain the action $\vec{\mathbf{v}}_n$ of the UAV based on the ϵ-greedy mechanism;
6: Execute the action $\vec{\mathbf{v}}_n$ and observe the next state $\mathbf{q}_{n+1}$, estimate the reward $\hat{R}_n$ according to (25);
7: Store the transition $(\mathbf{q}_n, \vec{\mathbf{v}}_n, \hat{R}_n, \mathbf{q}_{n+1})$ in $\mathcal{B}$; Randomly select a mini-batch of transitions $(\mathbf{q}_j, \vec{\mathbf{v}}_j, \hat{R}_j, \mathbf{q}_{j+1})$ from $\mathcal{B}$;
8: Obtain the Q-value function in (26); Update the network parameters by using gradient descent for loss function minimization defined in (27);
9: $n = n + 1$;
10: Add the measured rate $(\mathbf{q}_{n+1}, \tilde{R}^{\max}(\mathbf{q}_{n+1}))$ to database $\mathcal{B}'$; Sample random minibatch from $\mathcal{B}'$ and update the network parameter $\boldsymbol{\xi}$ for estimated MRM $\mathbf{R}$;
11: **for** $\tilde{n} = 1$ to $\tilde{N}$ **do**
12: Perform similar operations as steps 3-9 for the simulated experience with reward calculating based on estimated MRM $\mathbf{R}$;
13: **end for**
14: Update the target network parameters as $\boldsymbol{\theta}' = \boldsymbol{\theta}$ after a number of iterations;
15: **end while**
16: **end for**

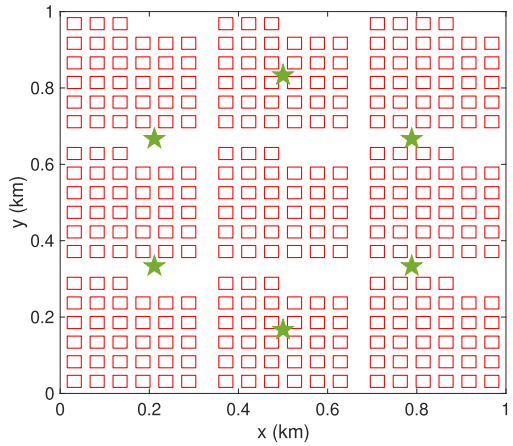
VI. Performance Evaluation

In this section, we provide the system settings and evaluate the performance of the graph based PBCD algorithm with a perfect ECKM and the dueling DQL algorithm without perfect ECKM through simulations. We also study the impact of different parameters on the performance of the proposed algorithms.

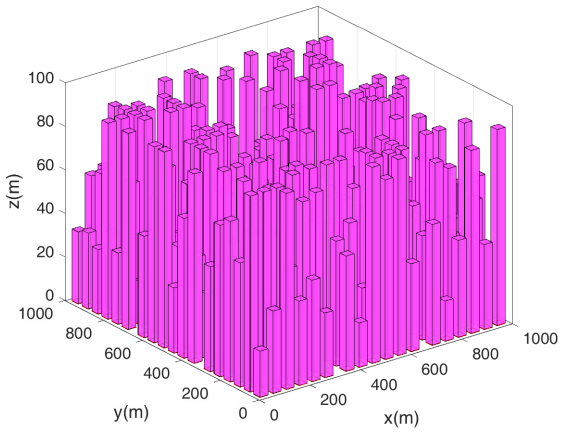
A. Simulation Setting

As shown in Fig. 3(a), we consider an aerial video streaming scenario with a cellular-connected UAV and $M = 6$ ground BSs, which are uniformly distributed in a square urban area with width 1.0 km. A ground video user located near an BS operates the UAV and shares the UAV's view based on the videos transmitted from the UAV, as shown in Fig. 1. All BSs are assumed to be at the same altitude of $H_G = 25$ m. The locations and heights of the buildings in the urban area are generated based on the model given by the International Telecommunication Union (ITU) [61], as shown in Fig. 3(b). For the DQL algorithm, we set $\Psi^{\max} = 2000$, $\hat{R}_{des} = 200$, $\hat{R}_{out} = 10^4$, $\mu_1 = 0.1$, $\mu_2 = 40$, $\mu_3 = 1$, $\epsilon = 0.5$, and $\tilde{N} = 10$. There are 5 hidden layers in the dueling DQN in Algorithm 4, where the numbers of neurons of the first 4 hidden layers are 512, 256, 128, and 128, respectively. Unless otherwise stated, the other parameters related to the simulations are set as follows: $\varpi = 0.8$, $\upsilon = 400$, $\beta_0 = -60$ dB, $\Gamma\sigma^2 = -110$ dBm, $V_{\max} = 20$ m/s, $H_{\min} = 50$ m, $H_{\max} = 200$ m, $f_c = 2$ GHz, $P = 0.1$ W, $B = 1$ MHz, $\Delta = 10$ m, $I_{\max} = 5$, $d_{\max} = 50$ m, $\kappa = 10^{-5}$, $\tilde{r} = 5$ Mbps, $T = 100$ s. With the specific locations and heights of the buildings, the ECKM is generated based on 3GPP channel models as specified in Section III-A for any location in the service area. To evaluate the performance of the proposed algorithms, we consider the following benchmarks for comparisons: 1) *Time minimization benchmark*, which minimizes the mission completion time with SINR requirement based on a given radio map as in [23]; 2) *DRL benchmark without mapping*, which is the state-of-the-art DRL benchmark without MRM mapping as in [11] and [41]; 3) *2D benchmark with mapping*, where DRL with radio mapping is utilized and the UAV flies at a sufficiently high altitude to avoid buildings as in [40]. Note that a radio map is given in the time minimization

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

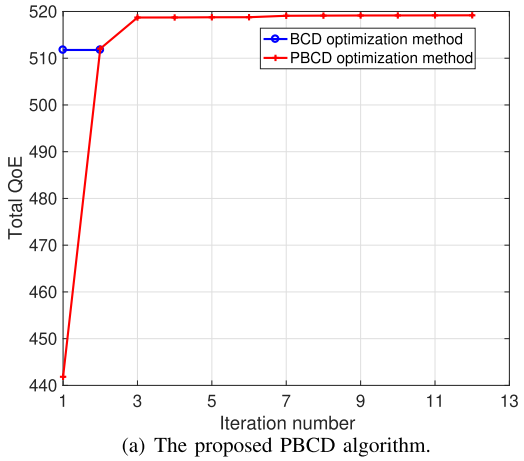


(a) The horizonal locations of ground BSs and buildings.

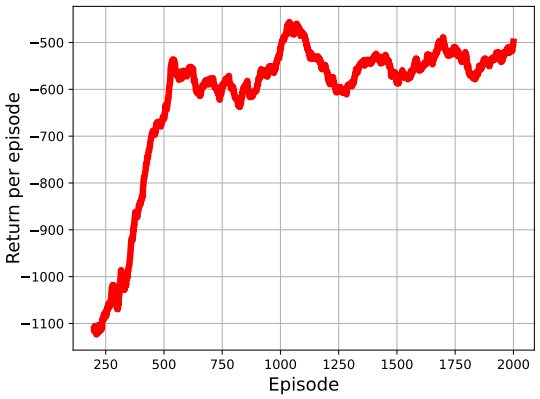


(b) The building distribution.

Fig. 3. The urban scenario considered in our evaluation.



(a) The proposed PBCD algorithm.



(b) Algorithm 4.

Fig. 4. Convergence behaviour of the proposed Algorithms.

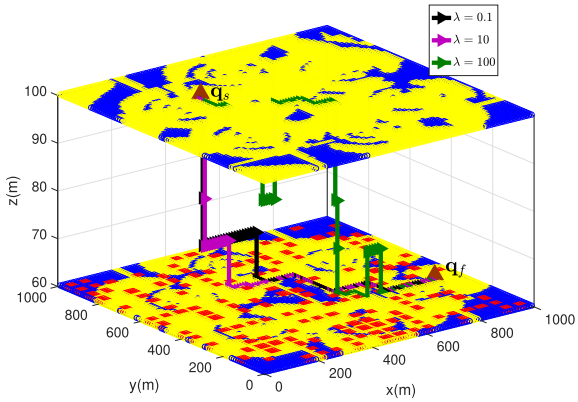


Fig. 5. Optimized trajectories with perfect ECKM.

benchmark, while such information is unavailable in the other two benchmarks.

B. Convergence of the Proposed Algorithms

Fig. 4(a) depicts the convergence behaviour of Algorithm 1 and the conventional BCD algorithm. It can be seen that the conventional BCD algorithm converges within 2 iterations with its objective value almost unchanged. For the proposed PBCD Algorithm, although it converges after 12 iterations with a low starting point, it can achieve a higher objective value than that of the conventional BCD. The reason is that although the starting point is low, since the initial parameter ρ_{ini} is less than 1, a larger search space is provided for each iteration with the increase of parameter ρ, resulting in more flexible optimization compared with BCD.

Fig. 4(b) depicts the moving average return per episode for the proposed DRL-based Algorithm, where the length of the moving window is 200 episodes. From Fig. 4(b), we can see that the average return increases with the training episode and saturates after sufficiently large episodes, which demonstrates the convergence of the proposed Algorithm 4. The UAV can learn from its experience for improving the long-term reward, which is upper-bounded by the accumulated QoE for moving from $\mathbf{q}_s$ to $\mathbf{q}_f$.

We first consider the performance with a given perfect ECKM. The optimized trajectories with different values of the balancing factor λ are shown in Fig. 5. Fig. 5 also describes the coverage map of the considered area, where the yellow areas denote the areas in which the UAV can achieve a maximum rate no less than $\tilde{r}$, while the blue areas denote that with maximum rate less than $\tilde{r}$. In addition, the red areas denote the obstacle areas with buildings where the UAV can not reach. It is observed that the coverage map varies with different heights, which is as expected due to the complex antenna pattern and building distribution. From Fig. 5, we can see that the UAV tries to move to the destination within the yellow areas for larger rates while avoiding the red and blue areas. As λ increases, the UAV tends to fly with less video rate variation. Intuitively, decreasing video rate variation will

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

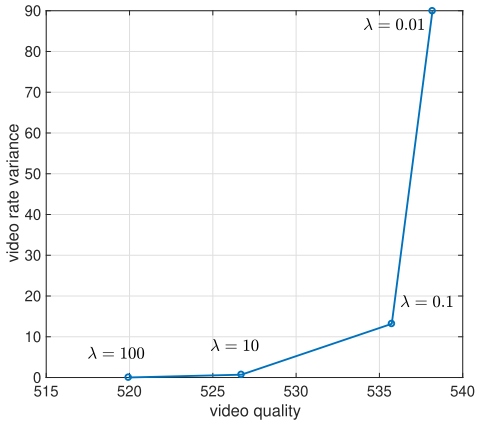
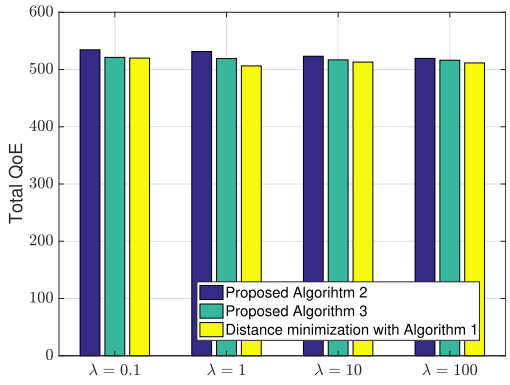
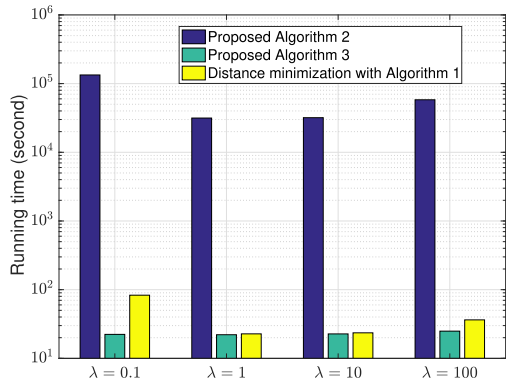


Fig. 6. The tradeoff curve.



(a) The total QoE versus λ.



(b) The running time versus λ.

Fig. 7. Performance comparisons with perfect ECKM.

restrict the increase of video quality, which is verified in Fig. 6. Fig. 6 shows the total video rate variation versus the total video quality with different values of λ. Thus, the QoE maximization should balance the tradeoff between decreasing video rate variation and increasing the video quality, which has already been captured by our proposed algorithms.

We compare the performance of the proposed algorithms in Fig. 7(a) and Fig. 7(b) with the perfect ECKM. The results for running time are obtained by solving the formulated problems efficiently through Matlab R2021b tools and CVX with solver SDPT3 over a computer with Quad core CPU 1.60 GHz. For fair comparison, we also consider a distance minimization approach, which minimizes the flying distance

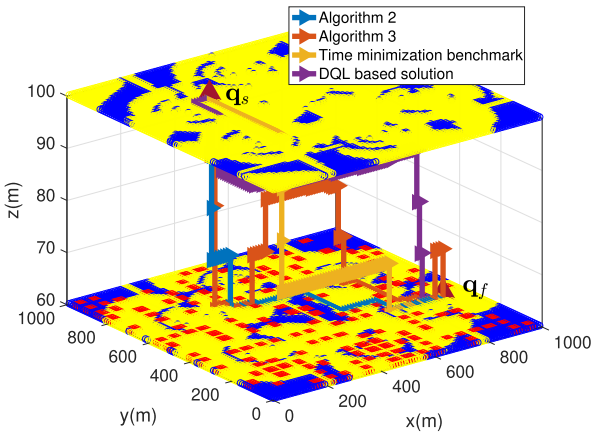


Fig. 8. UAV trajectories with different schemes without perfect ECKM.

by finding the shortest path over unweighted graph G but optimizes time duration and playback rate allocation with our proposed Algorithm 1. It is observed that Algorithm 2 achieves the largest QoE with the highest running time. This is expected since it selects the maximum QoE value from all paths in G, which is time consuming. Algorithm 3 achieves a larger QoE than the distance minimization approach since more heuristic information are utilized for path selecting. It is also observed that Algorithm 3 can achieve a comparable QoE as Algorithm 2 but with much less execution time. Thus, Algorithm 3 is more preferred than the other schemes with the perfect ECKM.

C. Performance When the Perfect ECKM Is Unavailable

Next, we consider the practical scenario and evaluate the performance of the proposed DQL-based solution when the perfect ECKM is unavailable and $\lambda = 0.1$. From Fig. 8, we can see that with the DQL-based solution, even without the perfect ECKM, the UAV is able to fly to the destination with larger achievable rates by successfully avoiding the red and blue areas, such that the buildings can be avoided and higher QoE can be achieved. This is because the UAV can intelligently make the most advantageous decisions of moving with the dueling DQN neural network by leveraging surrounding environment sensing and rate measurements. As more learning experience is accumulated, better performance can be achieved by exploiting of the learned MRM for path planning. Furthermore, with the simulated experience generation in Algorithm 4, fewer agent-environment interactions are required. To validate the quality of MRM estimated by Algorithm 4, Fig. 9 shows the mean error ratio of the learned MRM, which is calculated by comparing the predicted rates using the learned MRM versus their actual values given a set of randomly selected locations. Although the mean error ratio is initially large, it decreases rather quickly with the increase of episode number as more rate measurements are accumulated, which validates the effectiveness of Algorithm 4 for learning with MRM estimation.

Note that Algorithm 2 and Algorithm 3 are based on the perfect ECKM, which can serve as upper bounds for the

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

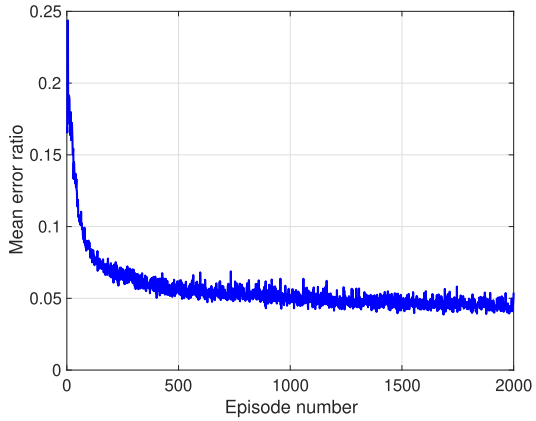


Fig. 9. The mean error ratio of MRM estimation.

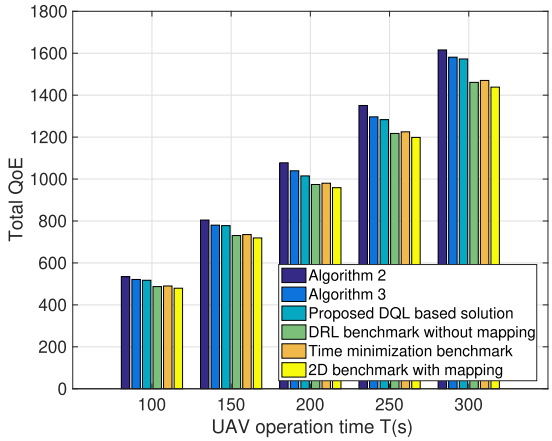


Fig. 10. Performance comparisons with the DQL-based solution.

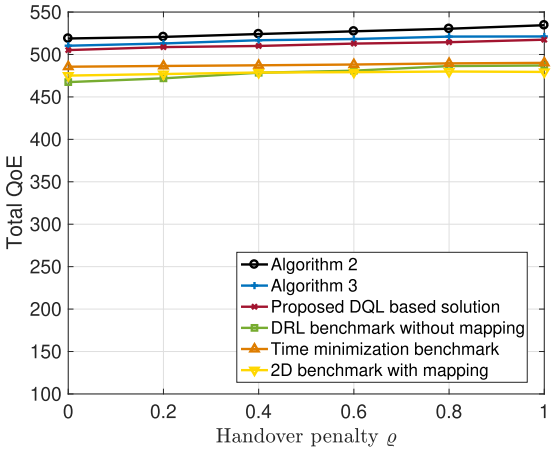


Fig. 11. The impact of handover penalty.

DQL-based solution without such information. It is observed in Fig. 10 that the performance gaps between the proposed DQL-based solution and the upper bounds are small, and thus the effectiveness of the DQL-based solution is demonstrated. Furthermore, the proposed DQL-based solution outperforms the other benchmarks, where the performance gains are more pronounced with a larger UAV operation time T. The additional gain is brought by the joint design of 3D UAV path as well as time duration and playback rate allocation. Although the time minimization benchmark is also based on a given radio map, it leads to a lower QoE since only time is minimized and the QoE requirements for aerial video streaming are ignored. Although radio mapping is also employed in 2D benchmark with mapping, its achieved QoE is lower than that of our proposed solution, since the flexibility of 3D flight is fully utilized in our proposed solution.

To investigate the impact of handovers, we introduce handover penalty factor $\varrho \in [0, 1]$ as in [62], where smaller value of ϱ corresponds to a bigger performance impact. If handover event occurs at location $\mathbf{q}_n$, then the normalized maximum achievable rate for the UAV at location $\mathbf{q}_n$ can be expressed as $\varrho \tilde{R}_n^{max}$. Fig. 11 shows the total QoE for different schemes with different values of handover penalty factor ϱ. As expected, the total QoE decreases with the decrease of ϱ, since smaller value of ϱ corresponds to a bigger performance impact over achievable rate. However, the performance impact is limited. From Fig. 11, we can see that the performance degradation ratio with handover effect is below 3% even with $\varrho = 0$. This is because in the considered scenario, the total handover duration is very small compared to the whole time duration T. As such, it is reasonable to assume that the overall performance degradation is negligible by employing appropriate handover procedure [8].

VII. Conclusion

In this paper, a data driven framework of video streaming over aerial 3D cellular networks was proposed. The total QoE of video streaming was maximized in a complex urban scenario by jointly optimizing 3D UAV trajectory, transmission scheduling, and playback rate adaption with building avoidance. To overcome the difficulties of environment-aware communications, we proposed a novel ECKM based approach. Specifically, efficient algorithms were proposed based on a perfect ECKM to obtain performance upper bounds by utilizing graph models and the iterative PBCD method. For the more practical scenario without an accurate ECKM, we proposed a dueling DQL solution with MRM construction to facilitate the learning based path design. The design framework is also applicable to the scenario with 3D beamforming with beamforming design for interference mitigation. In addition, the corresponding problem with accurate handover effect model will be left as future work.

References

[1] F. Metzger et al., "An introduction to online video game QoS and QoE influencing factors," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1894–1925, 3rd Quart., 2022.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May 2020.

[3] 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 Secur.*, vol. 16, pp. 1959–1972, 2021.

[4] M. Mozaffari, W. Saad, M. Bennis, Y. Nam, and M. Debbah, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2334–2360, 3rd Quart., 2019.

[5] S. Hu, Q. Wu, and X. Wang, "Energy management and trajectory optimization for UAV-enabled legitimate monitoring systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 142–155, Jan. 2021.

[6] Q. Wu et al., "A comprehensive overview on 5G-and-beyond networks with UAVs: From communications to sensing and intelligence," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 2912–2945, Oct. 2021.

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

[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] S. Hu, X. Yuan, W. Ni, and X. Wang, "Trajectory planning of cellular-connected UAV for communication-assisted radar sensing," *IEEE Trans. Commun.*, vol. 70, no. 9, pp. 6385–6396, Sep. 2022.

[10] 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.

[11] C. Zhan, H. Hu, S. Mao, and J. Wang, "Energy-efficient trajectory optimization for aerial video surveillance under QoS constraints," in *Proc. IEEE Conf. Comput. Commun.*. London, U.K., May 2022, pp. 1559–1568.

[12] 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.

[13] H. Zhou, F. Hu, M. Juras, A. B. Mehta, and Y. Deng, "Real-time video streaming and control of cellular-connected UAV system: Prototype and performance evaluation," *IEEE Wireless Commun. Lett.*, vol. 10, no. 8, pp. 1657–1661, Aug. 2021.

[14] 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.

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

[16] 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.

[17] J. Ji, K. Zhu, and L. Cai, "Trajectory and communication design for cache-enabled UAVs in cellular networks: A deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, early access, Jun. 13, 2022, doi: 10.1109/TMC.2022.3181308.

[18] B. Chang, W. Tang, X. Yan, X. Tong, and Z. Chen, "Integrated scheduling of sensing, communication, and control for mmWave/THz communications in cellular connected UAV networks," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2103–2113, Jul. 2022.

[19] C. Luo, M. N. Satpute, D. Li, Y. Wang, W. Chen, and W. Wu, "Fine-grained trajectory optimization of multiple UAVs for efficient data gathering from WSNs," *IEEE/ACM Trans. Netw.*, vol. 29, no. 1, pp. 162–175, Feb. 2021.

[20] W. Xu et al., "Minimizing the deployment cost of UAVs for delay-sensitive data collection in IoT networks," *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, pp. 812–825, Apr. 2022.

[21] Y. Hsu and R. Gau, "Reinforcement learning-based collision avoidance and optimal trajectory planning in UAV communication networks," *IEEE Trans. Mobile Comput.*, vol. 21, no. 1, pp. 306–320, Jan. 2022.

[22] Y. Zeng and X. Xu, "Toward environment-aware 6G communications via channel knowledge map," *IEEE Wireless Commun.*, vol. 28, no. 3, pp. 84–91, Jun. 2021.

[23] 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.

[24] S. Bi, J. Lyu, Z. Ding, and R. Zhang, "Engineering radio maps for wireless resource management," *IEEE Wireless Commun.*, vol. 26, no. 2, pp. 133–141, Apr. 2019.

[25] 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.

[26] L. Zhang and J. Chakareski, "UAV-assisted edge computing and streaming for wireless virtual reality: Analysis, algorithm design, and performance guarantees," *IEEE Trans. Veh. Technol.*, vol. 71, no. 3, pp. 3267–3275, Mar. 2022.

[27] 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.

[28] I. Comsa, G. Muntean, and R. Trestian, "An innovative machine-learning-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.

[29] F. Hu, Y. Deng, and A. H. Aghvami, "Cooperative multigroup broadcast 360° video delivery network: A hierarchical federated deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4009–4024, Jun. 2022.

[30] Z. Liu and Y. Jiang, "Cross-layer design for UAV-based streaming media transmission," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 7, pp. 4710–4723, Jul. 2022.

[31] 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.

[32] X. Zhang, J. Wang, and H. V. Poor, "Statistical delay and error-rate bounded QoS provisioning for 6G mURLLC over AoI-driven and UAV-enabled wireless networks," in *Proc. IEEE INFOCOM*, Vancouver, BC, Canada, May 2021, pp. 1–10.

[33] A. Rahmati, X. He, I. Guvenc, and H. Dai, "Dynamic mobility-aware interference avoidance for aerial base stations in cognitive radio networks," in *Proc. IEEE Conf. Comput. Commun.*, Paris, France, Apr. 2019, pp. 595–603.

[34] W. Xu et al., "Throughput maximization of UAV networks," *IEEE/ACM Trans. Netw.*, vol. 30, no. 2, pp. 881–895, Apr. 2022.

[35] 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. Technol.*, vol. 69, no. 11, pp. 12843–12856, Nov. 2020.

[36] M. Banagar and H. S. Dhillon, "3D two-hop cellular networks with wireless backhauled UAVs: Modeling and fundamentals," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6417–6433, Aug. 2022.

[37] L. Bertizzolo et al., "Streaming from the air: Enabling drone-sourced video streaming applications on 5G open-RAN architectures," *IEEE Trans. Mobile Comput.*, vol. 22, no. 5, pp. 3004–3016, May 2023.

[38] A. Trotta, F. D. Andreagiovanni, M. Di Felice, E. Natalizio, and K. R. Chowdhury, "When UAVs ride a bus: Towards energy-efficient city-scale video surveillance," in *Proc. IEEE Conf. Comput. Commun.*, Honolulu, HI, USA, Apr. 2018, pp. 1043–1051.

[39] H. Huang, A. V. Savkin, and W. Ni, "Online UAV trajectory planning for covert video surveillance of mobile targets," *IEEE Trans. Autom. Sci. Eng.*, vol. 19, no. 2, pp. 735–746, Apr. 2022.

[40] 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.

[41] K. K. Nguyen, T. Q. Duong, T. Do-Duy, H. Claussen, and L. Hanzo, "3D UAV trajectory and data collection optimisation via deep reinforcement learning," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2358–2371, Apr. 2022.

[42] W. Wang et al., "Robust 3D-trajectory and time switching optimization for dual-UAV-enabled secure communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 11, pp. 3334–3347, Nov. 2021.

[43] C. He, Y. Dong, and Z. Jane Wang, "Radio map assisted multi-UAV target searching," *IEEE Trans. Wireless Commun.*, early access, Dec. 20, 2022, doi: 10.1109/TWC.2022.3227933.

[44] W. B. Chikha, M. Masson, Z. Altman, and S. B. Jemaa, "Radio environment map based inter-cell interference coordination for massive-MIMO systems," *IEEE Trans. Mobile Comput.*, early access, Nov. 17, 2022, doi: 10.1109/TMC.2022.3222763.

[45] X. Xia, Y. Wang, K. Xu, and Y. Xu, "Toward digitalizing the wireless environment: A unified A2G information and energy delivery framework based on binary channel feature map," *IEEE Trans. Wireless Commun.*, vol. 21, no. 8, pp. 6448–6463, Aug. 2022.

[46] R. Shrestha, D. Romero, and S. P. Chepuri, "Spectrum surveying: Active radio map estimation with autonomous UAVs," *IEEE Trans. Wireless Commun.*, vol. 22, no. 1, pp. 627–641, Jan. 2023.

[47] D. Gesbert, O. Esrafilian, J. Chen, R. Gangula, and U. Mitra, "UAV-aided RF mapping for sensing and connectivity in wireless networks," *IEEE Wireless Commun.*, pp. 1–7, 2022.

[48] S. Shrestha, X. Fu, and M. Hong, "Deep spectrum cartography: Completing radio map tensors using learned neural models," *IEEE Trans. Signal Process.*, vol. 70, pp. 1170–1184, 2022.

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.

[49] Y. Teganya and D. Romero, "Deep completion autoencoders for radio map estimation," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1710–1724, Mar. 2022.
[50] S. Zhang, Y. Zeng, and R. Zhang, "Cellular-enabled UAV communication: A connectivity-constrained trajectory optimization perspective," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2580–2604, Mar. 2019.
[51] *Study on 3D Channel Model for LTE*, 3GPP, document TR 36.873, Dec. 2017.
[52] M. Rebato, L. Resteghini, C. Mazzucco, and M. Zorzi, "Study of realistic antenna patterns in 5G mmWave cellular scenarios," in *Proc. IEEE Int. Conf. Commun., Rel., Modeling Symp.*, Kansas City, MO, USA, May 2018, pp. 1–6.
[53] *Technical Specification Group Radio Access Network: Study on Enhanced LTE Support for Aerial Vehicles*, 3GPP, document TR 36.777, Dec. 2017.
[54] 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.
[55] Z. Jiang, C. Xu, J. Guan, Y. Liu, and G. Muntean, "Stochastic analysis of DASH-based video service in high-speed railway networks," *IEEE Trans. Multimedia*, vol. 21, no. 6, pp. 1577–1592, Jun. 2019.
[56] 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.
[57] D. B. West, *Introduction to Graph Theory*. 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.
[58] M. Grant and S. Boyd. (2016). *CVX: MATLAB Software for Disciplined Convex Programming*. [Online]. Available: http://cvxr.com/cvx
[59] Z. Wang, T. Schaul, M. Hessel, H. Van Hasselt, M. Lanctot, and N. De Freitas, "Dueling network architectures for deep reinforcement learning," in *Proc. ICML*, Jun. 2016, pp. 1995–2003.
[60] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning," in *Proc. 30th Conf. Artif. Intell. (AAAI)*, Mar. 2016, pp. 2094–2100.
[61] *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 1410-5, Feb. 2012.
[62] B. Galkin, E. Fonseca, R. Amer, L. A. DaSilva, and I. Dusparic, "REQIBA: Regression and deep Q-learning for intelligent UAV cellular user to base station association," *IEEE Trans. Veh. Technol.*, vol. 71, no. 1, pp. 5–20, Jan. 2022.

Cheng Zhan (Member, IEEE) received the B.Eng. and Ph.D. degrees in computer science from the School of Computer Science, University of Science and Technology of China, Anhui, China, in 2006 and 2011, respectively. From 2009 to 2010, he was a Research Assistant with the Department of Computer Science, City University of Hong Kong. From 2016 to 2017, he was a Visiting Scholar with the Department of Electrical and Computer Engineering, National University of Singapore. He is currently an Associate Professor with the School of Computer and Information Science, Southwest University, China. His research interests include unmanned aerial vehicle communications, multimedia communications, wireless sensor networks, and network coding. He served as a TPC Member for the IEEE ICC, GLOBECOM, WCNC, and UIC.

Han Hu (Member, IEEE) received the B.E. and Ph.D. degrees from the University of Science and Technology of China, China, in 2007 and 2012, respectively. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, China. His research interests include multimedia networking, edge intelligence, and space-air-ground integrated networks. He served as a TPC Member for Infocom, ACM MM, AAAI, and IJCAI. He received several academic awards, including the Best Paper Award of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY in 2019, the Best Paper Award of *IEEE Multimedia Magazine* in 2015, and the Best Paper Award of IEEE GLOBECOM 2013. He served as an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and *Ad Hoc Networks*.

Zhi Liu (Senior Member, IEEE) received the B.E. degree from the University of Science and Technology of China, China, and the Ph.D. degree in informatics from the National Institute of Informatics. He is currently an Associate Professor with The University of Electro-Communications, Japan. His research interests include video network transmission, vehicular networks, and mobile edge computing. He was a recipient of the IEEE StreamComm 2011 Best Student Paper Award, the 2015 IEICE Young Researcher Award, and the ICOIN 2018 Best Paper Award. He is also an Editorial Board Member of *Wireless Networks* (Springer) and IEEE OPEN JOURNAL OF THE COMPUTER SOCIETY.

Jing Wang (Member, IEEE) received the Ph.D. degree from Peking University in 2011. She is currently an Associate Professor with the School of Information, Renmin University of China. She has published papers on top computer conferences, such as MICRO, ISCA, HPCA, and IEEE/ACM TRANSACTIONS journals. Her research interests include edge intelligence and data analytics, computer system for artificial intelligence, and energy-efficient computing. She served as a TPC Member for ISCA, NAS, and ASDB. She received the Beijing Nova Award. She received the Best Paper Award of ICCD and the Featured Paper of IEEE TRANSACTIONS ON COMPUTERS.

Nan Cheng (Senior Member, IEEE) received the B.E. and M.S. degrees from the Department of Electronics and Information Engineering, Tongji University, Shanghai, China, in 2009 and 2012, respectively, and the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, in 2016. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, University of Toronto, from 2017 to 2019. He is currently a Professor with the State Key Laboratory of ISN and the School of Telecommunications Engineering, Xidian University, Shaanxi, China. He has published over 90 journal articles in IEEE TRANSACTIONS and other top journals. His current research interests include B5G/6G, AI-driven future networks, and space-air-ground integrated networks. He serves as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, and *Peer-to-Peer Networking and Applications*, and serves/served as the guest editor for several journals.

Shiwen Mao (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is currently a Professor and an Earle C. Williams Eminent Scholar and the Director of the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grids. He was a co-recipient of the 2022 Best Journal Paper Award of the IEEE ComSoc eHealth Technical Committee (TC), the 2021 Best Paper Award of *Digital Communications and Networks* (Elsevier/KeAi) Journal, the 2021 IEEE INTERNET OF THINGS JOURNAL Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the 2018 Best Journal Paper Award and the 2017 Best Conference Paper Award from IEEE ComSoc Multimedia TC, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He was a co-recipient of the Best Paper Award from IEEE ICC 2022 and 2013, IEEE GLOBECOM 2019, 2016, and 2015, and IEEE WCNC 2015, and the Best Demo Award from IEEE INFOCOM 2022 and IEEE SECON 2017. He was the General Chair of IEEE INFOCOM 2022, the TPC Chair of IEEE INFOCOM 2018, and the TPC Vice-Chair of IEEE GLOBECOM 2022. He is the Editor-in-Chief of IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING.

Authorized licensed use limited to: Auburn University. Downloaded on February 10,2025 at 17:20:44 UTC from IEEE Xplore. Restrictions apply.