IEEE Open Journal of
Vehicular Technology

Received 29 May 2025; accepted 28 June 2025. Date of publication 7 July 2025;
date of current version 4 August 2025. The review of this article was coordinated by Editor Sudharman K. Jayaweera.

Digital Object Identifier 10.1109/OJVT.2025.3586774

Multi-UAV Reinforcement Learning With Realistic Communication Models: Recent Advances and Challenges

TIZIANA CATTAI [1] (Member, IEEE), FRANCESCO FRATTOLILLO [2], ANDREA LACAVA [1,3] (Graduate Student Member, IEEE), PRASANNA RAUT [4], JENNIFER SIMONJAN [4], SALVATORE D'ORO [3] (Member, IEEE), TOMMASO MELODIA [3], EVGENII VINOGRADOV [4], ENRICO NATALIZIO [4] (Senior Member, IEEE), STEFANIA COLONNESE [1] (Senior Member, IEEE), FRANCESCA CUOMO [1] (Senior Member, IEEE), AND LUCA IOCCHI [2]

[1]Department of Information Engineering, Electronics and Telecommunication, Sapienza University of Rome, 00185 Rome, Italy
[2]Department of Computer, Control and Management Engineering Sapienza University of Rome, 00185 Rome, Italy
[3]Institute for the Wireless Internet of Things, Northeastern University, Boston MA 02120 USA
[4]Technology Innovation Institute Abu Dhabi, Abu Dhabi 9639, UAE

CORRESPONDING AUTHORS: TIZIANA CATTAI; FRANCESCO FRATTOLILLO (e-mail: tiziana.cattai@uniroma1.it; frattolillo@diag.uniroma1.it).

(Tiziana Cattai and Francesco Frattolillo contributed equally to this work.)

This work was supported in part by Cooperative Networked UAVs Project funded by Technology Innovation Institute and in part by the National Recovery and Resilience Plan (PNRR) Ministry of Education and Research (MUR) Project under Grant PE0000013-FAIR.

ABSTRACT The interest in applications related to Multi-Unmanned Aerial Vehicle (UAV) systems has been growing exponentially inthe last few years. Reinforcement Learning (RL) presents one of the most popular alternatives for solving Multi-UAV tasks, thanks to its flexible requirements for modelingthe problem. However, it is often applied to abstractions of the original problem, thus leaving to next development phases the integration of RL solutions to the actual systems. This choice may not guarantee the overall optimal performance of the implemented system. In this survey, we analyze the literature on Multi-UAV applications that utilize reinforcement learning, with particular attention to works that consider realistic communication channels. We focus on identifying the key variables that influence communication and whether these variables are integrated within the RL framework or considered externally. Additionally, we identify key trends, challenges, and future directions in the field, providing a comprehensive overview for researchers and practitioners interested in the practical deployment of RL-based Multi-UAV systems.

INDEX TERMS Multi-UAV, reinforcement learning, 5G, communications.

I. INTRODUCTION

Multi-Unmanned Aerial Vehicle (UAV) systems have been largely studied for their potential applications in several domains, including search and rescue [1], [2], delivery [3], [4], surveillance [5], [6], and communication service providers. The complexity of generating behaviors for multiple UAVs in dynamic and uncertain environments requires the development of effective control and decision-making strategies. To this end, RL is becoming a common technique to provide effective solutions due to its ability to deal with general environments and problems, with minimal and flexible requirements for modeling such problems.

It is also worth noting that some of the works that mathematically formulate the problem demonstrate that its solution is NP-hard [7], [8] or MINP [9]. Consequently, solving the problem using direct optimization techniques is highly complex, paving the way for approaches based on RL that do not require formal modeling of the problem since they learn through trial and error.

On the other hand, in general, RL methods do not scale well with the complexity of the problems, especially in multi-agent scenarios [10], [11] and also applying RL to real robotic systems may lead to dangerous and unpredictable behaviors [12], [13]. Therefore, the training environments used for training

RL agents are usually designed as abstractions of the real target environment. Such an abstraction aims to provide simpler models of the complex dynamics of UAV systems, their interaction with the environment, and their communication aspects, considering both downstream/upstream communications and UAV-to-UAV communications, which are typical of many applications.

While the abstract environments are necessary for effectively and efficiently training RL agents, the computed solutions may not be directly applicable to the real application domains, and they may lose the optimality guarantees achieved in the abstract model [14]. More generally, solutions that are optimal in the abstract environment used for RL training may be sub-optimal, may perform very poorly, or may even be fully impractical when deployed in the real application domain.

In this article, we are interested in studying the robustness of RL solutions computed in abstract environments when deployed in actual application scenarios, which is known as the Sim-to-Real transfer problem. Among the many features of Multi-UAV systems, our main interest in this paper is in communication aspects. More specifically, we investigate whether the current literature explores the impact of realistic communication aspects on the performance of RL solutions proposed for Multi-UAV applications. The paper will show that many of these points, though not all, are covered by the current literature, and it will identify areas of open challenges left for future research.

A. RESEARCH QUESTIONS

The general goal is to provide a structured approach to understand how realistic communication aspects are integrated and their implications in Multi-UAV RL frameworks. To this end, we analyzed the most relevant and most recent works that address this problem.

More specifically, our aim is to address the following Research Questions (RQs):

RQ1 *How are realistic communication aspects integrated in the UAVs' and environments' models?*

RQ2 *How are realistic communication aspects integrated into RL problem definition?*

RQ3 *How does performance degrade when realistic communication aspects are used for multi-UAV?*

The first question **RQ1** aims to identify which realistic communication aspects or specific features have been considered to model the multi-UAV framework. The answer to this research question is useful for evaluating the significance and practical applicability of the proposed solutions for actual deployment on real systems.

The second question **RQ2** refers to investigate which communication features are explicitly considered in the definition of the RL problem. Addressing this question should help to understand which are the most useful components to take into consideration when developing work on Multi-UAV systems in the context of RL. Finally, the third question **RQ3** focuses on experimental analysis and aims to assess the impact of realistic communication on the expected performance of the Multi-UAV system, with the final goal of understanding the strengths and limitations of the proposed solutions.

Taken together, the answers to the three RQs enable a clearer understanding of the literature on RL for Multi-UAV applications from the perspective of actual deployments in environments in which the communication aspects are essential factors.

The rest of this article is structured as follows. In Section II, we introduce some fundamental notions related to UAV systems, communication, and RL; Section III describes the methodology used to survey the literature; in Section IV we compare other review and survey papers on the topic, pointing out the main contribution of our work; Section V is focused on answering the research questions presented in the current section; following the analysis of **RQ2** related to the integration of communication-related aspects inside RL problems, we also thought that it would be interesting to describe more in detail some of the RL solutions used and the main applications considered in the analyzed literature which is the focus of Section VI; finally, in Section VII we summarize the findings and propose suggestions for researchers interested in working on this topic.

II. BACKGROUND

A. UAVS: AN INTRODUCTION TO THE TECHNOLOGY

UAVs, better known as drones, are flying platforms functioning with no human pilot on board. While they were originally made for military purposes, now they have become useful in many civil, commercial, and scientific applications [15]. The widespread use of civil UAVs is driven by the ability of these platforms to promptly access hazardous or difficult-to-reach locations.

The UAV-enabled use-cases span over emergency response [16], [17], package delivery [18], [19], infrastructure inspection [20], entertainment [21], and environmental [22] and traffic monitoring [23], [24] among others. For disaster response, UAVs provide real-time information on what is happening in the affected areas, which assists in locating the citizens that need rescue and allocating resources efficiently. Delivery systems have also witnessed a revolution with the creation of UAVs, which facilitate fast and contact-free delivery of items like goods, medicine, or even other services in areas that are severely underdeveloped or just hard to reach. For monitoring, UAVs are used to track and observe important national structures, monitor the borders, and even track illegal operations. In addition, UAVs are key players in environmental conservation, where they gather data to produce reports for things like deforestation, climate change, monitoring tracking for wildlife, and animal temperature studies. Beyond the impressive commercial promises, UAVs expose a potential to contribute to achieving many of the United Nations (UN)'s Sustainable Development Goals [25]. Fig. 1 repors a Schematic representation of a multi-UAV communication system.

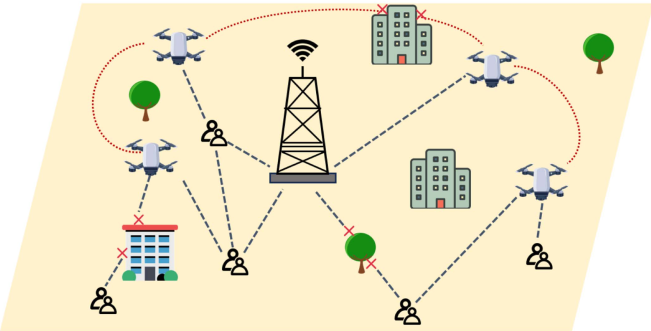

FIGURE 1. Schematic representation of a multi-UAV communication system. UAVs serve as communication relays to connect ground users and a central base station. Dashed lines represent active communication links, and red crosses indicate obstacles such as buildings and trees that block the signal.

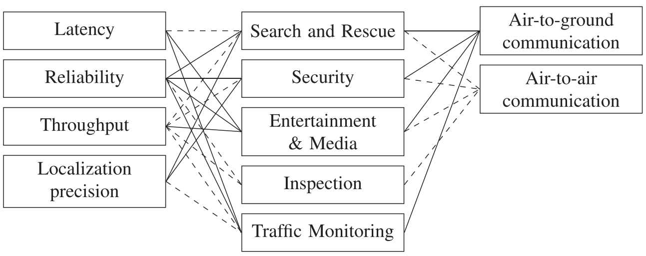

FIGURE 2. Relationships between communication metrics, use cases, and communication channels: solid and dashed lines indicate mandatory and use-case-dependent requirements, respectively [26].

B. UAV AND COMMUNICATION

The operations described above require continuous communication links for control, mission updates, and data retrieval [26], [27]. While flight controllers manage navigation, additional information must be exchanged with ground stations (i.e., Air-to-Ground communication is required) or other UAVs (Air-to-Air) to support real-time tasks, such as high-definition video streaming or sensor data collection. These demands introduce a set of constraints on link quality, latency, and bandwidth that may be satisfied by a variety of communication technologies (e.g., 5G, Wi-Fi, ad-hoc networks, and others).

Mission goals often impose stringent latency and throughput requirements along with the need for a reliable connection. Search and rescue teams, for example, rely on stable links to execute real-time maneuvers and to timely deliver information to the ground coordination center, while inspection crews may require sufficiently high data rates to transmit high-resolution imagery. Moreover, the communication requirements may vary during the mission: for example, higher video definition may be required to enhance object detection or vision-based navigation performance [28]. Fig. 2 highlights these relationships by showing how specific communication metrics align with UAV use cases and communication channels.

Apart from the communication requirements, missions define the desired UAV dynamics. For instance, a routine security patrol requires planning a trajectory and ensuring that all communication requirements are satisfied [29]. On the other hand, a drone relay needs to find an optimal position ensuring maximum coverage extension [30]. Moreover, when multiple UAVs are required to perform a mission, we face the problem of jointly optimizing communication, drone dynamics, data processing, and decision-making.

C. CHANNEL AND COMMUNICATION MODELING

Methods for analyzing and modeling the behavior of radio signals between UAVs and ground nodes have evolved from terrestrial modeling efforts [26], [31]. UAVs often fly at altitudes beyond typical terrestrial link assumptions, causing new signal propagation characteristics. The most important feature of Air-to-Ground (A2G) and Air-to-Air (A2A) channels is the higher probability of having Line-of-Sight (LOS) connection [31], [32]. In general, such links are more stable and experience less signal attenuation, however, UAV mobility can cause sudden obstructions of LOS resulting in sharp and fast signal level drops. Additionally, reflections from buildings and terrain irregularities can lead to highly variable channel responses [33] often exhibiting non-stationary behavior [34]. Accurately predicting channel attenuation, fading, and interference helps in designing more efficient communication-enabled UAV solutions.

Beyond propagation, UAV communication systems experience challenges related to resource allocation [35], [36], interference management [37], dynamic network topology [38]. For example, in the case of communicating with the ground infrastructure such as cellular networks, seamless handover across communication nodes or base stations represents another critical aspect of UAV connectivity [39], [40]. As drones travel over diverse terrains or move out of range of a given ground station, they must transition to a new station with minimal interruption. Failing to manage these handovers adequately can cause significant latency spikes, packet losses, or Radio Link Failures (RLFs). In addition, collaborative swarm operations have elevated the need for scalable and robust A2A connectivity and multi-UAV coordination and cooperation.

D. RL IN UAV MISSIONS

RL has emerged as a powerful approach to address decision-making and control challenges in multi-UAV operations, particularly within complex missions with multiple metrics to account for. A Multi-agent Reinforcement Learning (MARL) problem is usually described through a framework called Markov Game (MG), or stochastic games [41], which is defined by a tuple $\langle N, \mathcal{S}, \mathcal{A}, T, R \rangle$, where: N is the number of players (agents); $\mathcal{S}$ is the set of environment states shared by all agents; $\mathcal{A}$ is the set of *joint* actions $\mathcal{A} = \mathcal{A}^1 \times \cdots \times \mathcal{A}^N$, where $\mathcal{A}^i$ is the set of actions available to agent i; $T : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the *joint* transition function; $R : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \rightarrow [0, 1]$ is the *joint* reward function. The goal in MARL is to find a set of N functions called policies π^i, with $i \in N$, that maximize the expected joint sum of discounted rewards: $\sum_{k=t}^{T} \gamma^k R(s_k, a_k)$, where $0 \leq \gamma < 1$ is called the discount

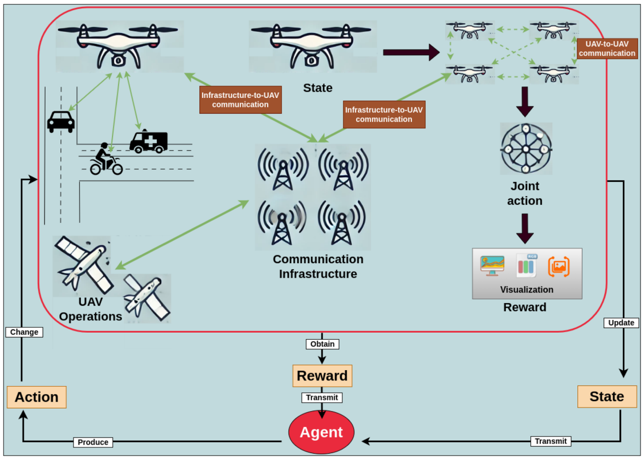

FIGURE 3. **Background: multi-UAV RL with communication.**

factor. RL frameworks enable UAVs to learn optimal behaviors through interaction with their environment by leveraging the core components of states, actions, and rewards. In this context:

- *States* represent the UAV's current conditions, including its position, velocity, battery level, and communication metrics.
- *Actions* encompass the set of possible maneuvers or communication strategies the UAV can employ, such as adjusting flight paths, altering transmission power, or switching communication channels.
- *Rewards* are designed to reflect mission objectives, such as maintaining robust connectivity, minimizing latency, maximizing throughput, or conserving energy.

By iteratively selecting actions that maximize cumulative rewards, UAVs can autonomously optimize their communication performance in dynamic environments.

Early RL applications in UAVs focused on single-drone tasks like trajectory optimization for maintaining LoS links and basic interference management. As UAV operations have become more complex, RL has been extended to address multi-objective and multi-agent scenarios crucial for robust communication systems (see Fig. 3). For instance, RL algorithms have been employed to dynamically allocate communication resources, manage handovers between ground stations, and coordinate multiple UAVs to form relay networks that enhance coverage and reliability.

E. SIM-TO-REAL IN MULTI-UAV RL

Applying RL directly in real-world applications is a very challenging problem. There are multiple factors that contribute to the complexity of this process. Firstly, RL suffers from the problem of sample inefficiency, which means that algorithms typically require large amount of data to learn effective policies. However, collecting such data in the physical world can be a very expensive and time-consuming operation. Moreover, due to the trial-and-error nature of reinforcement learning, certain exploratory actions may lead to undesirable and dangerous behaviors [12]. To address these challenges, a common solution is to train RL agents in a simulated environment that approximates the real-world dynamics and then transfer the learned behavior to the real world. This process is commonly referred to as Sim-to-Real [14].

Sim-to-Real also has its own set of challenges. The most important is denoted as *reality gap*, which is the mismatch between the simulated and the real control-law caused by small modeling inaccuracies in the simulator [42]. Despite some solutions to the problem, such as domain randomization [43], which allows learning robust policies by randomizing some aspects of the simulated environment, the reality gap is still an open challenge.

The number of works that apply sim-to-real to UAVs are limited, and they are mainly related to single UAV collision-free navigation tasks [44], [45]. Recently [46] successfully applied Deep Reinforcement Learning (DRL) to a UAV for competing in a top-level drone racing competition. It is even more complex to apply sim-to-real to Multi-UAV applications. [47] proposed a digital-twin based solution for swarm cooperative target searching based on MARL policy. The field of sim-to-real in the context of reinforcement learning applied to UAV systems is both a highly fascinating and complex area. There is still much to explore, and it is hoped that future work will focus more effort in this direction.

III. SURVEY METHODOLOGY

In this section, we describe the article selection methodology used to identify the papers to be included in this survey, focusing specifically on literature that integrates realistic communication aspects with Multi-UAV RL systems.

Numerous conferences have explored topics related to multi-UAV and RL. However, for reasons of feasibility and to ensure a comprehensive review of articles presenting novel methods in detail, we decided to exclude conference proceedings from our search. Instead, we focused on IEEE Xplore, one of the most important scientific databases, because of its significance in the field and its highly practical search and filtering tools.

We started from the survey published in 2023 [48], which only contains papers until the end of 2022, with the aim to extend that study, which was mainly focused on reviewing deep RL algorithms for multi-UAV. We initially focused our search on journal papers from 2015 by identifying keywords related to communication, such as Path Loss, Signal-to-Noise ratio, Age of Information, Quality of Service, Channel Allocation, and co-channel Interference. This enabled us to obtain a reproducible query with logical operators.

The proposed query specifically required the inclusion of some keywords aiming to properly identify the topic of the survey. Indeed, we identified the necessary words as *Multi-UAV*, *reinforcement learning*, *Communication*, and *5G*, because they ensure that only articles addressing the core topics of interest are selected. Additionally, we incorporated

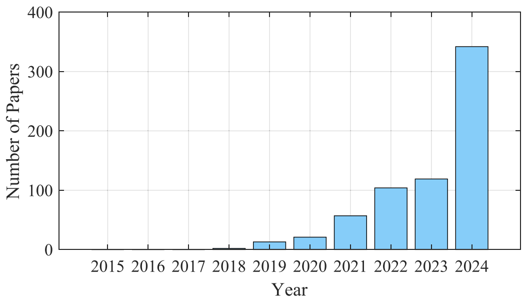

FIGURE 4. Bar plot of the number of papers identified with the proposed query.

another set of optional keywords, linked by the OR operator, to capture specific aspects of communication realism and refer to our prior knowledge on the topic. Those terms relate to signal propagation and quality, such as *Path Loss*, *Signal-to-Noise Ratio*, and *Signal to Interference Noise Ratio*. We also included communication metrics and models, with keywords like *Quality of Service*, *NodeB*, *Channel modeling*, *Co-channel interference*, and *Channel allocation*. Furthermore, the query included references to application contexts, such as *Streaming*, *Video*, and *Measurements*, as well as simulation techniques like *Ray tracing*, to ensure the inclusion of journal papers that present realistic communication aspects.

Using the defined query, we obtained a total number of 658 papers from 2015 to 2024. In Fig. 4, we report the bar plot of the number of papers found by the query and the years in order to analyze the distribution of papers related to the survey's topic over time. We observed that there are no papers published before 2018 and only two in 2018. Interestingly, there has been a significant increase in the number of journal publications in recent years, with a notable peak in 2024. This trend proves the growing interest in the integration of realistic communication aspects within Multi-UAV RL systems.

The articles examined in the subsequent sections incorporate at least some of these variables within their models. These variables span multiple levels of the protocol architecture, ranging from the physical and Medium Access Control layers to the network layer, each capturing distinct aspects of communication dynamics, quality, or performance. In the following, we briefly describe the key quantities that correspond to the optional keywords used in the query.

The analyzed studies employ various models for path loss, which characterizes the attenuation of signal power as it propagates through space, distinguishing between scenarios where a line of sight exists and those where it does not. Channel modeling entails the mathematical representation of the communication environment, incorporating factors such as multipath propagation, fading, and interference. Typically, these models assume a non-quadratic power decay with distance to account for real-world environmental factors, such as obstacles. The Signal-to-Noise-Ratio (SNR), which quantifies the power of the received signal relative to background noise, serves as a key determinant of channel quality and available throughput. Frequently, this metric appears in the more general form of the Signal to Interference plus Noise Ratio (SINR), particularly in scenarios where multiple UAVs share a common access medium. In most cases, interference is mitigated through the use of orthogonal subchannels, achieved by assigning individual communication links either non-overlapping frequency bands, distinct time slots, or orthogonal modulation codes.

Channel allocation plays a crucial role in the optimal assignment of UAV communication channels, enhancing network performance while minimizing interference. However, co-channel interference may still arise when multiple communication links utilize adjacent physical channels, a phenomenon accurately accounted for by the SINR. In many studies, SINR calculations assume that transmitting UAV base stations do not employ power control mechanisms to compensate for variations in receiver distance; instead, transmission power is typically kept constant, with coverage constraints defined by a minimum SINR threshold.

For UAVs engaged in data collection from Internet of Things (IoT) devices monitoring environmental conditions, a critical communication parameter in the design of reward functions is the Age of Information (AoI), which measures the freshness of the information exchanged or collected. Additionally, UAV communication performance is assessed through established Quality of Service (QoS) parameters used in communication networks, including latency, jitter, and packet delivery rate. Furthermore, next generation networks enable the adoption of advanced capabilities such as Reconfigurable Intelligent Surface (RIS) and Multi-access Edge Computing (MEC) functionalities, further enhancing UAV communication efficiency. The specific UAV communication architecture ultimately relates to the RL objective under consideration, with model choices tailored to optimize various aspects of UAV network performance in dynamic environments.

IV. RELATED SURVEYS

There are multiple surveys and reviews on multi-UAV applications focusing on different aspects, but we will focus on the surveys identified through the previously described methodology.

Several communication aspects have been analyzed in the different works, for instance, [33], [49] analyzed the literature on channel modeling of UAVs as Low altitude Platform (LAP), while [50] explored the most recent specifications by 3rd Generation Partnership Project (3GPP) as well as recent quantitative experiments on wireless communications and path planning in UAV networks. Similarly, [51] reviewed the main solutions and challenges classified according to the application domain, and papers [26], [52] reviewed the use-cases, main problems, and architectures involving UAV communications. In addition, [53] tackled issues related to network layer-related communications. Key aspects associated with non-terrestrial networks are addressed in [54], where authors reviewed channels in mmwave and beamforming and integration in IoT networks, including MARL applications for mobile edge computing. The general problem of aerial networks is presented in [55], [56] and the survey [57] discusses the localization challenges and solutions in aerial and space networks.

Systems of UAVs as cyber-physical systems are examined in [78], which reviews U2G and U2U (i.e., A2A) channel models typically assuming LOS channels with degradation dependent on distance—and explicitly identifies the backhaul network as an open issue due to potential link blockage or interference with the access link. Aerial ad hoc networks and related routing challenges are reviewed in [73] and [74], [75], respectively, and the standardization of aerial networks is discussed in [76].

Other surveys concentrate on specific aspects of networking and communications, addressing security and routing issues such as UAV detection mechanisms [70], the application of blockchain technologies for secure UAV communications [71], and energy-efficient communication methods crucial for UAV operation [72]. The application of game theory to UAV communications is reviewed in [59], where it is applied to optimize coverage, trajectories, task allocation, energy consumption, security, load balancing, and offloading, despite its high computational complexity making it less suitable for multi-agent systems. Moreover, [66] addresses UAV support for the Internet of Everything (IoE) and the associated scalability challenges, while [67], [68] focus on the role of UAVs in MEC.

Several recent works examine the use of RL techniques for UAV communications [58], although they vary substantially in terms of their communication models, the realism of their evaluations, and the specific Artificial Intelligence (AI)/RL algorithms they employ. Surveys such as [60] and [61] provide broad overviews of AI-driven methodologies in wireless networks, including UAV scenarios, but they do not focus on the unique challenges of multi-UAV coordination or realistic, detailed channel models. Similarly, [62] offers a tutorial-style discussion of AI solutions for wireless systems without addressing the specifics of UAV mobility and communication links. Another survey, [69], is oriented toward IoT use cases and thus has limited applicability to cellular-based UAV environments, while [65] reviews UAV-assisted data collection for IoT without considering multi-UAV scenario. Meanwhile, [63] presents a more architectural perspective, highlighting how AI tasks can be distributed between edge nodes and the cloud in infrastructure-less networks, though it assumes perfect or static Channel State Information (CSI) and does not explicitly target UAV mobility issues. A more hands-on approach appears in [64], which surveys practical UAV implementations, including hardware and onboard AI modules, as well as potential communication technologies like Long Range (LoRa), Bluetooth, Wi-Fi, and Long Term Evolution (LTE). However, it, too, does not deeply address the nuances of multi-UAV collaboration or the use of RL algorithms for managing these networks. In contrast, [77] focuses on cellular networks and standard handover procedures, dedicating only a small section to UAV-related challenges.

To provide a clear categorization of the existing surveys, we have organized them into five macro-categories, as shown in Table 1. These include studies focusing on computational methodologies such as reinforcement learning, deep learning, and game theory for UAV and multi-UAV optimization; surveys analyzing UAV-assisted services, such as sensing, computing offloading, and IoT integration; works focused on networking and communication protocols, including security and blockchain, routing, energy efficiency, and standardization; surveys on aerial and ground communication, covering aspects related to channel modeling, beamforming, non-terrestrial networks, localization, and U2G/ U2U communication; and, finally, survey papers oriented to applications, with a focus on architectures, network-layer challenges, and practical implementations leveraging technologies like LoRa, Bluetooth, Wi-Fi, and LTE.

Together, those surveys provide an overview of the challenges and advancements in multi-UAV communications, but they tend to focus on isolated aspects of multi-UAV systems without providing an integrated view of how communication models affect the modeling and performance of multi-agent RL approaches.

V. ANALYSIS

In this section, we explore the different methods adopted by the papers identified with the proposed methodology focusing on our research questions. During this analysis, we have considered only articles that include multi-UAV applications in which agents are the UAV devices and that implement some realistic communication channel modeling. To be more specific, we are going to exclude from this analysis all the works that came out from our query and that do not include RL solutions [79], [80], or that are single agent [81], [82], or in which the agents are not the UAVs but other entities such as Ground Users (GUs) [83], [84].

A. INTEGRATION OF REALISTIC COMMUNICATION ASPECTS IN UAV MODELS AND ENVIRONMENTS (RQ1)

This subsection describes how the papers that we have previously identified have incorporated realistic communication aspects into UAV models and the environments used for training.

First of all, it is important to identify a taxonomy with the key communication features that can be considered in multi-UAV architecture. To this aim, we mention:

- at the physical/MAC layer: latency, bandwidth, energy, delay, loss, errors;
- at the networking layer: routing, hierarchy, delays, quality of service, backhaul;
- at the application/session/transport layer: adoption of reliable, unreliable transport protocol, session control protocol, content delivery servers, and mobile computing.

At the physical/Medium Access Control (MAC) layer, multi-UAV networks have to address key aspects such as latency, bandwidth limitations, energy consumption, delay,

TABLE 1. Categorization of Related Surveys Organized by Macro-Categories

Survey Reference(s)	Macro-Category	Focus
[58], [59], [60], [61], [62], [63], [64]	Methodological approaches	computational methods such as RL, deep learning, and game theory for UAV/Multi-UAV optimization.
[65], [66], [67], [68], [69]	Services	sensing, computing offloading, and IoT integration.
[70], [71], [72], [73], [74], [75], [76], [77], [53]	Networking and Communication Protocols	security, routing, blockchain for UAV security, energy-efficient UAV communication, and standardization of aerial networks.
[49], [33], [50], [54], [55], [56], [57], [78]	Aerial and Ground Communication	channel modeling, mmWave, beamforming, non-terrestrial networks, localization, and UAV-to-Ground (U2G) / UAV-to-UAV (U2U) communication.
[51], [26], [52], [64]	Applications	architectures, civil application domains, and practical implementations including LoRa, Bluetooth, Wi-Fi, and LTE.
our survey	All Communication elements + RL	RL approaches for Multi-UAV applications with realistic communication models

packet loss, and error rates. Recent research has proposed different ways to tackle those issues, such as adaptive modulation, robust error estimation, and energy-aware MAC protocols to dynamically optimize physical layer parameters and improve overall network functioning [85].

In this context, the network layer has a crucial role in enabling efficient communications in multi-UAV systems. Routing protocols generally adopt hybrid strategies, combining deterministic and stochastic techniques to select optimal paths based on metrics such as link stability, available bandwidth, latency, and energy consumption. For instance, in [86], authors have presented a trajectory and packet routing method to optimize quality metrics like SINR, queue delay, and energy. However, there are many routing protocols proposed in UAVs, such as topology-based, position-based, deterministic, stochastic, and hierarchical, able to divide the network into clusters or levels, reducing routing overhead and improving the scalability and including QoS metrics [87].

The idea of including appropriate transport protocols in multi-UAV networks is fundamental to allow key requirements, such as latency and data quality, particularly in video transmission and mobile computing applications. RL approaches have the potential to improve video delivery with dynamic modifications of compression and transmission to guarantee Quality of Experience (QoE) under different and dynamic network conditions [88]. Furthermore, UAV-assisted edge computing frameworks enable efficient session control and content delivery by optimizing trajectory planning and resource allocation through multi-agent deep RL, decreasing the total system cost and improving service differentiation [89]. In addition, hierarchical federated learning methods enables cooperative content distribution among UAV clusters, improving scalability and adaptability for high-demand applications such as 360-degree video streaming [90].

The reviewed works generally focus on the integration of LOS and Non-Line-of-Sight (NLOS) conditions, which are critical for modeling realistic U2G and U2U communication. In those papers, the models used to define LOS/NLOS differ in their formulation based on environmental parameters, elevation angles, and probabilistic modeling. For instance, in some of the papers, such as [79], [81], [91], [92], [93], an exponential-based loss probability model is adopted, where an exponential model is used to derive the probability of LOS, which is a function of the elevation, the carrier frequency and other environmental factors. Another possible approach used in [7], [94] is to define the LOS probability by the estimation of how UAV height and transmission distance impact on the LOS probability using a power law. Another possibility consists in modelling the LOS/NLOS connection between UAVs and devices using functions of the LOS/NLOS probability and the transmitting distance in logarithmic expressions [8]. Differently from basic logarithmic expressions, alternative approaches use a log-normal model for shadowing that explicitly takes into consideration random variations due to obstacles and reflections [95]. More specific frameworks can be formulated according to the applications, such as in [96], authors use a pure LOS model.

Another key aspect often considered in the papers is the presence of noise or interference in proposed methodologies. Indeed, the noise model can be applied in different points depending on the system design, but the different approaches differ for the network scenario and how interference is considered in each case. For instance, in [7], [95], [97], noise is simply applied as the background white Gaussian noise over a specific bandwidth, and the interference is assumed negligible. Differently, in paper [8], [83], both uplink and downlink scenarios are considered with additive noise and also interference coming from other cells (inter-cell interference (ICI)). In [79], authors take into consideration the

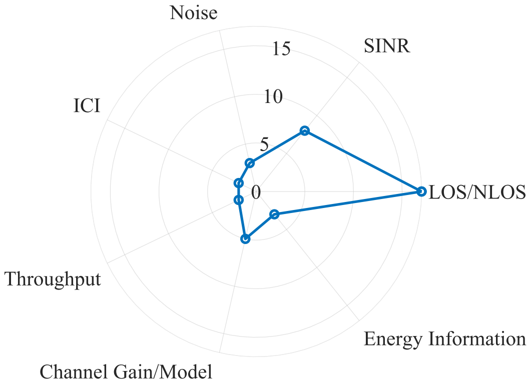

FIGURE 5. Radar plot representing the number of papers including the different communication-related features in the physical/MAC layer.

SINR for each user as a measure of the quality by considering additive Gaussian noise and the presence of interference, but alternatively considering the additional complexity of Non-Orthogonal Multiple Access (NOMA) systems can be explored [80], [98], [99]. Some of the reviewed papers also consider energy consumption with different strategies. For instance, in [100], authors quantifies the energy cost during task offloading by considering the consumption for local computations, transmission (using an optional cost related to UAV flying), and hovering, focusing on the trade-offs between local executions and offloading. Differently, a non-linear energy harvesting model is applied in [81] to measure circuit saturation effects, and [94] intuitively integrates energy transfer within a UAV-enabled wireless communication network scenario, using a Time Division Multiple Access (TDMA) protocol and studying the probabilistic nature of communication channels to optimize energy delivery and uplink communication. Channel behavior has been explored via the channel gain [80], [83], [101], which can be defined as function of the distance between GU and UAVs [83], or its functioning model [102], [103], which can be a parametric random model [102]. Another aspect related to the physical/MAC layer taken into consideration is the latency in [8], where the total round time for scheduling a device is the sum of the local model update latency, uplink local model upload latency, global model aggregation latency, and downlink global model broadcast latency.

It is important to note that most papers incorporate realistic communication models at the physical layer, especially through the LOS and NLOS. This is evident in Fig. 5, where we have a number of papers covering the different communications aspects with a specific focus on the physical/MAC layer.

Very few articles concentrate on the networking and application/session/transport layers. For example, the QoE is explored in [7], where it is measured as the Mean Opinion Score (MOS) that is a function of delay and rate. The execution time delay, which refers to the processing time of an offloaded task on computational nodes, is considered in [100]. Moreover, the study in [80] focuses on the challenges related to limited energy storage, and the total energy consumption is separated into task execution at UAVs and cloud servers, offloading costs from devices to UAVs and from UAVs to cloud, and the energy required for UAV hovering, and it also decomposes the user delay in stages relates to offloading processing and result transfer stages. Caching mechanisms and user associations are investigated in [95]. In addition, some cross-layer elements, such as those related to privacy and security, are investigated in [8], [83].

Taken together, there is no unique approach to incorporate realistic communication aspects into multi-UAV models. Different studies focus on distinct aspects of communication, depending on their objectives and application scenarios.

B. INCORPORATING REALISTIC COMMUNICATION INTO RL PROBLEM DEFINITIONS (RQ2)

This section examines studies that incorporate realistic communication aspects into RL problems, specifically those addressing RQ2. It highlights what are the main communication-related variables included in the different elements that make up a RL problem. The goal is to identify useful components and strategies for advancing State-of-the-Art (SOTA) methods in multi-UAV contexts.

We include in our analysis all the variables that directly or indirectly are related to communication, except for the position of the elements of the problems, such as the UAVs, the Ground Devices (GDs) Cartesian or polar coordinates. Moreover, we exclude variables related to some computational task completion, such as the computational capacity [9], the number of Central Processing Unit (CPU) cycles, the time maximum time delay of input task [9], or the computational resources to allocate to a device [91].

Starting from the formalization of a RL problem provided in Section II-D, we note that in order to influence the behavior of the agents, communication parameters should be included either in the action, in the state space, or in the reward function. We analyze each component to identify key features across different applications and optimization goals while grouping similar concepts to avoid excessive detail.

State: The state of the agent allows access to direct information when making decisions. In some of the works, the UAVs need to communicate with GDs, and information about their subchannel allocation may be included in their state space as a binary variable [8]. In some applications, the UAVs are required to upload or download some payload. The percentage of completion or the remaining part of the payload can be included in the state [8]. The state can also include variables related to the channel condition between devices [93], [104] such as the SINR [9], [100] or the channel gain [98], [99]. In [105], the authors consider a specific application where the Multi-UAV network also includes a RIS. In such a case, they include the reflected channel gains since the RIS reflects the signal from UAVs to other devices. In [101] a similar scenario is presented, which included the sum rate of mobile devices associated with UAV at the previous time step.

In [95], the authors include the transmission delay in the state, while in [106] they include the available transmission power. In task-offloading-related applications, the state may also include the parameters related to bandwidth [92], [107], [108] or whether the channel is already busy because it is connected to a device [96]. In IoT applications, some works also include the average throughput of IoT devices in the state space [94]. Other works included energy information, which depends on wireless communication [109], [110] or additional information to identify whether user equipment obtains wireless communication service or not [109].

Actions: In RL, actions allow the agent to transition from one state to another. Usually, an RL agent starts by picking an action randomly, and during the training, it learns which actions return better long-term outcomes, and it starts to exploit them. Being able to correctly decide whether to explore or stick with good actions is called the exploration-exploitation dilemma, and it is still an open challenge. Despite the fact that usually, the action space of the UAVs is limited to their movement [7], [93], [94], [96], [109], there are different works that include actions related to communication. Such actions includes the choice of a specific transmission power [8], [9], [95], [98], [99], [100], [105], [106], [110]. Deciding which device to communicate with [8] or even which channel to allocate to a specific device [8]. Some applications also includes offloading actions [91], [92], [104], [106], [107], [110], [111]. In [101], the authors consider UAV-mounted RIS, and their action includes the phase shift of the RIS.

Rewards: The reward is a function that describes the objective that a RL wants to maximize. In the context of UAVs, some elements are related to the minimization of the time needed/delay to complete some communication-related action [8], [9], [95], [97], [100], [106], [107], [110] or in MEC applications, it may also be referred to as AoI [92]. In IoT applications, is usually a reward based on the average throughput of IoT devices [94], [96] or sum rate [101] or sum bit rate [93]. Throughput is also used under the constraint of guaranteeing user fairness [98], [99]. Another objective is usually the minimization of energy consumption; such reward function usually includes elements related to transmission power [105], [106] or offloading decisions [104]. In [91], the authors consider a MEC scenario, and their proposed reward includes a penalty for offloading the task instead of computing it locally. The reward may also depend on the bandwidth condition [108] or on the quality of service offered [109]. [7] is the only work that consider a communication-related parameter only in the reward function. In this work, the UAVs are used as Aerial Base Station (ABS) in 3D space, and the reward is a function of the MOS of the ground users.

Fig. 6 visualizes the distribution of these RL components across the selected papers, emphasizing the integration of realistic communication metrics into RL-based multi-UAV models, highlighting the increasing focus on adapting RL problem definitions to reflect the complex and dynamic nature of multi-UAV communication networks.

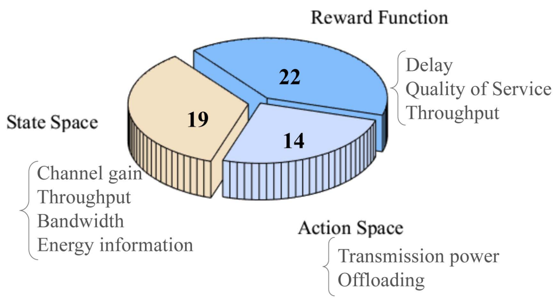

FIGURE 6. Pie chart representing the number of papers that include communication inside a specific element of the RL.

C. PERFORMANCE EVALUATION IN REALISTIC COMMUNICATION SCENARIOS (RQ3)

This subsection reviews papers assessing the impact of realistic communication on multi-UAV system performance, taking into consideration works that address RQ3. It discusses experimental results measuring model robustness, sensitivity to communication variables, and performance degradation under adverse conditions. The aim is to identify the limitations of current solutions and outline future research directions. Of the papers surveyed in this work, we note that only a few actually implement realistic communications or use elaborate channel models, leaving this RQ3 mostly unfulfilled by the current SOTA. Indeed, wireless channel models are crucial in multi-UAV problems, affecting the training and testing conditions of AI agents and underscoring the need to consider their impact in system design [112].

Works such as [79] demonstrate that effective three-dimensional (3D) placement of UAV-mounted base stations combined with careful spectrum allocation can mitigate the interference inherent in realistic environments. Likewise, energy-efficient 3D deployments in partial offloading scenarios have been proposed in [80], highlighting strategies that jointly optimize transmit power and resource allocation to balance coverage, throughput, and battery constraints. By incorporating more realistic channel models and communication assumptions, these solutions begin to reveal how dynamic link conditions can hinder performance and necessitate robust algorithms for interference management and UAV coordination.

Another line of research addresses mission-critical tasks like offloading or data processing, often focusing on dynamic, time-varying environments. For instance, multi-UAV networks employing NOMA and MEC have been studied in [98], where trajectory planning must be continuously adapted in concert with scheduling and power control. Similarly, [83] explores MARL methods for controlling UAV positions, transmit power, and decoding orders. Although these learning-based frameworks typically lead to improved spectral efficiency, recent findings highlight their sensitivity to abrupt changes in channel quality or interference levels—particularly under adverse conditions. Performance can degrade unless the agents are retrained or augmented with safety mechanisms that anticipate disturbances.

In parallel, some works examine the aspects of security and privacy. A UAV-assisted MEC system adopting RL to thwart eavesdropping attacks is discussed in [83], illustrating how constraints on UAV trajectory design and resource allocation can reduce flexibility in the presence of malicious adversaries. Further, [8] investigates privacy-preserving approaches via federated learning, ensuring user data remains on local devices and mitigating privacy leakage. While these methods highlight the importance of robust communication—especially under uncertain or contested settings—they also underscore the inherent trade-offs between reliability, resource usage, and security. Once more, the omission of rigorous channel modeling in many existing approaches points to unresolved vulnerabilities and performance gaps.

From an algorithmic standpoint, advanced techniques leveraging quantum-inspired clustering and Spiking Neural Networks (SNNs) have been explored to optimize UAV placement and spectrum usage [99]. While initial results are promising—offering reduced complexity and better resilience in noisy conditions—empirical tests still tend to simplify channel or mobility assumptions. Indeed, [99] provides evidence that these new paradigms can handle dynamic factors to a degree, but more realistic modeling of fading, interference, and user distributions would be necessary to confirm robustness in the real world.

Experimental results reported in the literature often differ significantly due to the lack of a standardized scenario, which is sometimes caused by the specific application targeted in each work. The datasets used for training and testing RL algorithms are usually custom-made and personalized, making comparisons difficult. Concerning some key parameter used in the papers, we mention that most works simulate scenarios with fewer than five UAVs, typically within an area of 500 m $\times$ 500 m, although coverage areas vary widely across studies, ranging from as small as 50 m $\times$ 50 m to as large as 2000 m $\times$ 2000 m. The current lack of standardization in UAV research can be partially mitigated through the use of advanced and well-established testbeds for channel emulation and UAV experimentation [113]. In particular, wireless platforms such as Colosseum [114] and the Aerial Experimentation and Research Platform for Advanced Wireless (AERPAW) [115] offer standardized and reproducible environments for the development and evaluation of reinforcement learning algorithms for UAVs. Colosseum enables real-time emulation of realistic RF channels, full-stack 5G/Open RAN deployments, and dynamic network topologies, supporting the validation of control and communication strategies in complex wireless scenarios. AERPAW is the first wireless testbed designed to explore the integration of 5G technologies with autonomous drone systems, which can indeed help the validation of multi-RL approaches over a real-world testbed.

To validate their proposed methods, authors commonly compare different RL algorithms for the same task or vary the parameters of their scenario. Typical performance metrics include convergence behavior, computation offloading efficiency, resource allocation, coverage ratio, task execution delay, and energy consumption [9], [91], [100]. Some studies also report AoI and throughput [92]. Only a few works, such as [104], provide qualitative but interesting comparisons of both algorithms and scenarios. Most contributions focus on benchmarking RL algorithms, such as Deep Deterministic Policy Gradient (DDPG), Soft Actor-Critic (SAC), and Deep Q-Learning (DQL) [9], [91], [97], [100], [104]. Alternatively, some studies fix the learning algorithm and investigate the impact of varying scenario parameters, such as the number of UAVs [95], transmit power [111], or environmental scenario [97]. An overview of the experimental evaluation settings and performance metric for quantitative analysis is synthetically reported in Table 2.

TABLE 2. Commonly Adopted Experimental Evaluation Settings and Performance Metrics

Experimental settings	Performance metrics
Number of UAV	• Convergence performance
2-7	• Energy consumption
Coverage [mxm]	• Time delay
50x20 - 5000x5000	• Task offloading rate
Coverage [devices]	• Coverage ratio
$\sim$ 20 IoT	• Latency
$\sim$ 10 mobile users	• Age of Information

VI. ALGORITHMS AND APPLICATIONS

In the following section, we take a closer look at some interesting aspects that were not covered by the previous research questions. Specifically, we analyze the main algorithms used in this context to identify the most commonly adopted solutions. Additionally, we examine the primary applications in which UAVs with realistic communication models are employed.

A. ALGORITHMS ANALYSIS

In [116], the authors investigate a RIS-assisted scenario with Double-Deep Q-Network (DQN) to control a UAV's trajectory, ensuring connectivity with terrestrial User Equipment (UE). Although they capture some mobility aspects, the communication framework remains somewhat high-level in terms of per-user metrics. Another study, [117], adopts Q-learning for beamforming in a cellular network with multiple UAVs, using a more rigorous channel model that highlights the degradation in signal quality when UAVs move outside main beam coverage. Similarly, [118] integrates both DQN and DDPG to jointly optimize UAV path planning and beamforming, again leveraging detailed propagation assumptions with RIS.

Several other papers highlight the interplay between UAV mobility and resource optimization. In [119], Actor-Critic (AC) RL techniques guide the positioning of UAVs that function as mobile edge servers, jointly optimizing user association and UAV trajectory in a 5G environment. A similarly elaborate approach is found in [120], where the authors use a multi-step dueling Double-DQN to plan three-dimensional

TABLE 3. DRL Algorithms Used in multi-UAV Applications With Realistic Communication Channels and Their Corresponding References

Algorithm	References
Q-learning	[7, 118, 120]
Asynchronous Advantage AC (AC) algorithm	[8]
DDPG	[9, 97, 95, 93] [118, 121]
DQL	[97, 94]
Proximal Policy Optimization (PPO)	[105, 101]
Twin Delay DDPG (DDPG)	[91]
Multi-Agent DDPG (DDPG)	[100]
H-MAAC (Heterogeneous Multi-agent AC)	[107]
H-A2C (Heterogeneous Advantage AC (AC))	[92]
Mutual DQL	[98]
SAC	[104]
CommNet	[109]
CA2C (Compound-action AC)	[110]

UAV trajectories under a realistic channel model (including LOS/NLOS effects and antenna tilt) to maintain robust connectivity. Though not primarily focused on communication throughput, [121] also includes a detailed channel model while exploring wireless power transfer among UAVs, applying DDPG-based RL to manage battery recharging tasks.

A summary of the main RL algorithms used across the surveyed papers is shown in Table 3. Collectively, these studies underscore the diversity of RL frameworks such as Q-learning, DQN variants, DDPG, AC applied to UAV communications. They also highlight the varying degrees of real-world fidelity in the network assumptions, from simplified channel modeling to more sophisticated treatments incorporating path loss, LOS/NLOS probabilities, beam misalignment, and mobility constraints. Nevertheless, there remains a need for comprehensive methods that combine the rigor of detailed channel modeling with robust multi-UAV coordination and real data or field tests. This gap motivates further research into RL-based UAV communications under realistic conditions, as few of the surveyed papers fully integrate all aspects of real-time multi-UAV operation, end-to-end performance objectives, and physically validated channel dynamics.

B. MAIN APPLICATIONS

UAVs are used to provide connection to GUs by maximizing their MOS [7] or in multi-objective cases, they can optimize data rate while also minimizing latency [99]. In [8], the authors consider a privacy-preserving application for machine learning model training using data from GDs. UAVs are also used as MEC to help IoT devices [9], cellular devices [98], smart mobile devices [92] and in scenarios such as maritime communication [97], or intelligent agriculture [110], either by collecting data, by offloading tasks, by performing some information processing or even for caching contents [95]. In the context of IoT applications, UAVs can also be used to broadcast energy signals to charge IoT devices [94]. In [96], the authors consider a multi-UAV application that includes both the previous scenarios where UAVs harvest fresh data, and they also broadcast energy signals to recharge IoT devices. A multi-UAV-aided MEC network, which also includes charging stations that allow UAVs to recharge in order to sustain longer missions, is considered in [91]. A MEC scenario where there are also eavesdroppers, and ground jammers need to generate jamming signals in order to disturb them is considered in [111], but it includes a single UAV. UAVs are also employed in Flying Ad-Hoc Networks (FANETs), where they are required to work together in groups. An example is [108] which uses Network Virtualization (NV) technologies and Service Function Chain (SFC) to facilitate terrestrial networks. The same study also uses UAVs in combination with satellitesin space-air-ground integrated network (SAGIN).Multi-UAV networks are also used in combination with Reconfigurable Intelligent Surfaces in order to improve the network efficiency [105] or in more complex scenarios, including interference from unknown jammers [101]. Finally, UAVs are also used in the context of federated learning, where agents are trained on local and private data, and they can only share their network parameters [107].

VII. CONCLUSIONS AND FUTURE DIRECTIONS

In this review, we began by providing an overview of UAV systems, communication frameworks, and RL techniques. We then examined the existing survey papers in the literature, identifying a notable gap in the application of realistic communication systems within RL solutions for multi-UAV applications. Specifically, we outlined the methodology employed in our research and presented a set of research questions aimed at understanding the implementation of realistic communication in simulated multi-UAV systems. These questions were designed to investigate how realistic communication is integrated into multi-UAV RL problems, how such elements influence problem formulation, and how the presence of realistic communication, as opposed to abstract communication models, impacts results and solution transferability to real-world contexts.

Despite significant progress, there remain persistent limitations. First, many solutions rely on partially observable or even idealized network states, whereas unpredictable channel fluctuations or node failures are the norm in actual deployments. Second, energy constraints for both UAVs and mobile users often force difficult trade-offs between mission longevity and throughput gains. Third, although various methods tackle security or privacy, end-to-end encryption, and trust verification for UAV-aided edge learning are still evolving. Looking ahead, future research should focus on:

1) *Adaptive Learning and Control:* RL paradigms capable of rapidly adapting to channel degradation or security breaches without extensive retraining.
2) *Integrated UAV–Terrestrial Infrastructures:* Coordinated scheduling between UAV swarms and terrestrial

cells, ensuring reliable and high-throughput coverage during demand spikes.

3) *Robust Security and Privacy Mechanisms:* Lightweight cryptographic methods or blockchain-based solutions that protect sensitive data with minimal overhead.
4) *Use of Common Testbeds:* Development and adoption of frameworks for multi-UAV applications that can be shared by the research community. It would be useful to introduce tables with the parameters used, in order to enable a quantitative comparison between different studies
5) *Extensive Field Trials:* Validating theoretical frameworks with large-scale experimental data to capture propagation anomalies and varied user behaviors, ultimately closing the gap between simulation assumptions and practical deployment.

Altogether, while multi-UAV communication systems hold promise for the next generation of wireless networks, realistic channel modeling and robust communication assumptions remain underexplored. Incorporating these aspects more thoroughly will not only help realize consistent performance and quality of service under adverse conditions but will also guide researchers in developing integrated solutions that satisfy both technical and operational demands. The majority of studies are still conducted in abstract and simulated environments. Furthermore, while these studies provide valuable insights, there remains a lack of comprehensive validation to assess the efficacy of these solutions in real-world settings. Given the rapid growth of research in this domain, we recommend a stronger focus on the integration of realistic communication models in future work. This would not only enhance the applicability of RL solutions but also ensure their robustness and reliability in real-world multi-UAV applications.

REFERENCES

[1] S. Waharte and N. Trigoni, “Supporting search and rescue operations with UAVs,” in *Proc. Int. Conf. Emerg. Secur. Technol.*, 2010, pp. 142–147.

[2] E. T. Alotaibi, S. S. Alqefari, and A. Koubaa, “LSAR: Multi-UAV collaboration for search and rescue missions,” *IEEE Access*, vol. 7, pp. 55817–55832, 2019.

[3] Y. Han et al., “Two-stage heuristic algorithm for vehicle-drone collaborative delivery and pickup based on medical supplies resource allocation,” *J. King Saud Univ.- Comput. Inf. Sci.*, vol. 35, no. 10, 2023,Art. no. 101811.

[4] K. Kuru, D. Ansell, W. Khan, and H. Yetgin, “Analysis and optimization of unmanned aerial vehicle swarms in logistics: An intelligent delivery platform,” *IEEE Access*, vol. 7, pp. 15804–15831, 2019.

[5] N. H. Motlagh, M. Bagaa, and T. Taleb, “UAV-based IoT platform: A crowd surveillance use case,” *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 128–134, Feb. 2017.

[6] H. Huang and A. V. Savkin, “Aerial surveillance in cities: When UAVs take public transportation vehicles,” *IEEE Trans. Automat. Sci. Eng.*, vol. 20, no. 2, pp. 1069–1080, Apr. 2023.

[7] X. Liu, Y. Liu, and Y. Chen, “Reinforcement learning in multiple-UAV networks: Deployment and movement design,” *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 8036–8049, Aug. 2019.

[8] H. Yang, J. Zhao, Z. Xiong, K. Y. Lam, S. Sun, and L. Xiao, “Privacy-preserving federated learning for UAV-enabled networks: Learning-based joint scheduling and resource management,” *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3144–3159, Oct. 2021.

[9] A. M. Seid, G. O. Boateng, S. Anokye, T. Kwantwi, G. Sun, and G. Liu, “Collaborative computation offloading and resource allocation in multi-UAV-assisted IoT networks: A deep reinforcement learning approach,” *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12203–12218, Aug. 2021.

[10] L. Canese et al., “Multi-agent reinforcement learning: A review of challenges and applications,” *Appl. Sci.*, vol. 11, no. 11, 2021, Art. no. 4948. [Online]. Available: https://www.mdpi.com/2076-3417/11/11/4948

[11] K. Zhang, Z. Yang, and T. Başar, *Multi-Agent Reinforcement Learning: A Selective Overview of Theories and Algorithms*. Cham, Switzerland: Springer International Publishing, 2021, pp. 321–384. [Online]. Available: https://doi.org/10.1007/978-3-030-60990-0_12

[12] J. Kober and J. Peters, “Reinforcement Learning in Robotics: A Survey,” in *Learn. Motor Skills*. Cham, Switzerland: Springer International Publishing, 2014, pp. 9–67. [Online]. Available: https://doi.org/10.1007/978-3-319-03194-1_2

[13] J. García and F. Fernández, “A comprehensive survey on safe reinforcement learning,” *J. Mach. Learn. Res.*, vol. 16, no. 42, pp. 1437–1480, 2015. [Online]. Available: http://jmlr.org/papers/v16/garcia15a.html

[14] W. Zhao, J. P. Queralta, and T. Westerlund, “Sim-to-real transfer in deep reinforcement learning for robotics: A survey,” in *Proc. 2020 IEEE Symp. Ser. Comput. Intell.*, 2020, pp. 737–744.

[15] S. A. H. Mohsan, M. A. Khan, F. Noor, I. Ullah, and M. H. Alsharif, “Towards the unmanned aerial vehicles (UAVs): A comprehensive review,” *Drones*, vol. 6, no. 6, 2022, Art. no. 147.

[16] M. Erdelj, E. Natalizio, K. R. Chowdhury, and I. F. Akyildiz, “Help from the sky: Leveraging UAVs for disaster management,” *IEEE Pervasive Comput.*, vol. 16, no. 1, pp. 24–32, Jan.–Mar. 2017.

[17] A. Trotta, L. Montecchiari, M. D. Felice, and L. Bononi, “A GPS-free flocking model for aerial mesh deployments in disaster-recovery scenarios,” *IEEE Access*, vol. 8, pp. 91558–91573, 2020.

[18] N. Cherif, W. Jaafar, H. Yanikomeroglu, and A. Yongacoglu, “3D aerial highway: The key enabler of the retail industry transformation,” *IEEE Commun. Mag.*, vol. 59, no. 9, pp. 65–71, Sep. 2021.

[19] A. Colpaert, M. Raes, E. Vinogradov, and S. Pollin, “Drone delivery: Reliable cellular UAV communication using multi-operator diversity,” in *Proc. ICC 2022 - IEEE Int. Conf. Commun.*, 2022, pp. 1–6.

[20] M. Erdelj, O. Saif, E. Natalizio, and I. Fantoni, “UAVs that fly forever: Uninterrupted structural inspection through automatic UAV replacement,” *Ad Hoc Netw.*, vol. 94, Nov. 2019, Art. no. 101612. [Online]. Available: https://doi.org/10.1016/j.adhoc.2017.11.012

[21] S. Ortiz, C. T. Calafate, J.-C. Cano, P. Manzoni, and C. K. Toh, “A UAV-based content delivery architecture for rural areas and future smart cities,” *IEEE Internet Comput.*, vol. 23, no. 1, pp. 29–36, Jan./Feb. 2019.

[22] O. Alvear, N. R. Zema, E. Natalizio, and C. T. Calafate, “Using UAV-based systems to monitor air pollution in areas with poor accessibility,” *J. Adv. Transp.*, vol. 2017, no. 1, 2017, Art. no. 8204353. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2017/8204353

[23] H. Menouar, I. Guvenc, K. Akkaya, A. S. Uluagac, A. Kadri, and A. Tuncer, “UAV-enabled intelligent transportation systems for the smart city: Applications and challenges,” *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 22–28, Mar. 2017.

[24] A. Saboor, E. Vinogradov, Z. Cui, S. Coene, W. Joseph, and S. Pollin, “Elevating the future of mobility: UAV-enabled intelligent transportation systems,” in *Proc. 7th Int. Conf. Adv. Commun. Technol. Netw.*, 2024, pp. 1–7.

[25] I. Gryech, E. Vinogradov, A. Saboor, P. S. Bithas, P. T. Mathiopoulos, and S. Pollin, “A systematic literature review on the role of UAV-enabled communications in advancing the UN’s sustainable development goals,” *Front. Commun. Netw.*, vol. 5, 2024, Art. no. 1286073. [Online]. Available: https://www.frontiersin.org/journals/communications-and-networks/articles/10.3389/frcmn.2024.1286073

[26] E. Vinogradov, H. Sallouha, S. D. Bast, M. Azari, and S. Pollin, “Tutorial on UAVs: A blue sky view on wireless communication,” vol. 14, no. 4, pp. 395–395, 2018.

[27] Y. Zeng, R. Zhang, and T. J. Lim, “Wireless communications with unmanned aerial vehicles: Opportunities and challenges,” *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 36–42, May 2016.

[28] Q. Chen, H. Zhu, L. Yang, X. Chen, S. Pollin, and E. Vinogradov, "Edge computing assisted autonomous flight for UAV: Synergies between vision and communications," *IEEE Commun. Mag.*, vol. 59, no. 1, pp. 28–33, Jan. 2021.

[29] K. S. Kappel, T. M. Cabreira, J. L. Marins, L. B. D. Brisolara, and P. R. Ferreira, "Strategies for patrolling missions with multiple UAVs," *J. Intell. Robotic Syst.*, vol. 99, pp. 499–515, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:203138177

[30] L. Zhu, J. Zhang, Z. Xiao, X. Cao, X.-G. Xia, and R. Schober, "Millimeter-wave full-duplex UAV relay: Joint positioning, beamforming, and power control," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2057–2073, Sep. 2020.

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

[32] A. Saboor, E. Vinogradov, Z. Cui, A. Al-Hourani, and S. Pollin, "A geometry-based modelling approach for the line-of-sight probability in UAV communications," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 364–378, 2024.

[33] W. Khawaja, I. Guvenc, D. W. Matolak, U.-C. Fiebig, and N. Schneckenburger, "A survey of air-to-ground propagation channel modeling for unmanned aerial vehicles," *IEEE Commun. Surv. Tut.*, vol. 21, no. 3, pp. 2361–2391, thirdquarter 2019.

[34] A. Colpaert, Z. Cui, E. Vinogradov, and S. Pollin, "3D non-stationary channel measurement and analysis for MaMIMO-UAV communications," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 6061–6072, May 2024.

[35] P. Luong, F. Gagnon, L.-N. Tran, and F. Labeau, "Deep reinforcement learning-based resource allocation in cooperative UAV-assisted wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7610–7625, Nov. 2021.

[36] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, and W. Jiang, "Multi-agent DRL for task offloading and resource allocation in multi-UAV enabled IoT edge network," *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4531–4547, Dec. 2021.

[37] U. Challita, W. Saad, and C. Bettstetter, "Interference management for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2125–2140, Apr. 2019.

[38] M. Matracia, M. A. Kishk, and M.-S. Alouini, "On the topological aspects of UAV-assisted post-disaster wireless communication networks," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 59–64, Nov. 2021.

[39] A. Colpaert, E. Vinogradov, and S. Pollin, "3D beamforming and handover analysis for UAV networks," in *Proc. 2020 IEEE Globecom Workshops*, 2020, pp. 1–6.

[40] N. Cherif, W. Jaafar, H. Yanikomeroglu, and A. Yongacoglu, "RL-based cargo-UAV trajectory planning and cell association for minimum handoffs, disconnectivity, and energy consumption," *IEEE Trans. Veh. Technol.*, vol. 73, no. 5, pp. 7304–7309, May 2024.

[41] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *Mach. Learn. Proc.*, Morgan Kaufmann, 1994, pp. 157–163.

[42] E. Salvato, G. Fenu, E. Medvet, and F. A. Pellegrino, "Crossing the reality gap: A survey on sim-to-real transferability of robot controllers in reinforcement learning," *IEEE Access*, vol. 9, pp. 153171–153187, 2021.

[43] I. Akkaya et al., "Solving Rubik's cube with a robot hand," 2019, *arXiv:1910.07113*.

[44] F. Sadeghi and S. Levine, "CAD2RL: Real single-image flight without a single real image," 2016, *arXiv:1611.04201*.

[45] H. I. Ugurlu, X. H. Pham, and E. Kayacan, "Sim-to-real deep reinforcement learning for safe end-to-end planning of aerial robots," *Robotics*, vol. 11, no. 5, 2022, Art. no. 109. [Online]. Available: https://www.mdpi.com/2218-6581/11/5/109

[46] E. Kaufmann, L. Bauersfeld, A. Loquercio, M. Müller, V. Koltun, and D. Scaramuzza, "Champion-level drone racing using deep reinforcement learning," *Nature*, vol. 620, no. 7976, pp. 982–987, 2023.

[47] P. Cao, L. Lei, G. Shen, S. Cai, X. Liu, and X. Liu, "UAV swarm cooperative search based on scalable multiagent deep reinforcement learning with digital twin-enabled sim-to-real transfer," *IEEE Trans. Mobile Comput.*, vol. 24, no. 6, pp. 5173–5188, Jun. 2025.

[48] F. Frattolillo, D. Brunori, and L. Iocchi, "Scalable and cooperative deep reinforcement learning approaches for multi-UAV systems: A systematic review," *Drones*, vol. 7, no. 4, 2023, Art. no. 236.

[49] A. A. Khuwaja, Y. Chen, N. Zhao, M. S. Alouini, and P. Dobbins, "A survey of channel modeling for UAV communications," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 2804–2821, Fourthquarter 2018.

[50] G. Geraci et al., "What will the future of UAV cellular communications be? A flight from 5G to 6G," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 3, pp. 1304–1335, thirdquarter 2022.

[51] H. Shakhatreh et al., "Unmanned aerial vehicles (UAVs): A survey on civil applications and key research challenges," *IEEE Access*, vol. 7, pp. 48572–48634, 2019.

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

[53] B. Li, Z. Fei, and Y. Zhang, "UAV communications for 5G and beyond: Recent advances and future trends," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2241–2263, Apr. 2019.

[54] M. M. Azari et al., "Evolution of non-terrestrial networks from 5G to 6G: A survey," *IEEE Commun. Surv. Tut.*, vol. 24, no. 4, pp. 2633–2672, Fourthquarter 2022.

[55] N.-N. Dao et al., "Survey on aerial radio access networks: Toward a comprehensive 6G access infrastructure," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1193–1225, Secondquarter 2021.

[56] S. Zhang, D. Zhu, and Y. Wang, "A survey on space-aerial-terrestrial integrated 5G networks," *Comput. Netw.*, vol. 174, 2020, Art. no. 107212.

[57] H. Sallouha, S. Saleh, S. D. Bast, Z. Cui, S. Pollin, and H. Wymeersch, "On the ground and in the sky: A tutorial on radio localization in ground-air-space networks," *IEEE Commun, Surveys Tuts.*, vol. 27, no. 1, pp. 218–258, Feb. 2025.

[58] Y. Bai, H. Zhao, X. Zhang, Z. Chang, R. Jäntti, and K. Yang, "Toward autonomous multi-UAV wireless network: A survey of reinforcement learning-based approaches," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 4, pp. 3038–3067, Fourthquarter 2023.

[59] M. E. Mkiramweni, C. Yang, J. Li, and W. Zhang, "A survey of game theory in unmanned aerial vehicles communications," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3386–3416, Fourthquarter 2019.

[60] N. C. Luong et al., "Applications of deep reinforcement learning in communications and networking: A survey," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3133–3174, Fourthquarter 2019.

[61] T. Li et al., "Applications of multi-agent reinforcement learning in future internet: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 24, no. 2, pp. 1240–1279, Secondquarter 2022.

[62] A. Feriani and E. Hossain, "Single and multi-agent deep reinforcement learning for AI-enabled wireless networks: A tutorial," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1226–1252, Secondquarter 2021.

[63] W. Xu, Z. Yang, D. W. K. Ng, M. Levorato, Y. C. Eldar, and M. Debbah, "Edge learning for B5G networks with distributed signal processing: Semantic communication, edge computing, and wireless sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 9–39, Jan. 2023.

[64] A. N. Wilson, A. Kumar, A. Jha, and L. R. Cenkeramaddi, "Embedded sensors, communication technologies, computing platforms and machine learning for UAVs: A review," *IEEE Sensors J.*, vol. 22, no. 3, pp. 1807–1826, Feb. 2022.

[65] Z. Wei et al., "UAV-assisted data collection for Internet of Things: A survey," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 15460–15483, Sep. 2022.

[66] Y. Liu, H.-N. Dai, Q. Wang, M. K. Shukla, and M. Imran, "Unmanned aerial vehicle for internet of everything: Opportunities and challenges," *Comput. Commun.*, vol. 155, pp. 66–83, 2020.

[67] M. Abrar, U. Ajmal, Z. M. Almohaimeed, X. Gui, R. Akram, and R. Masroor, "Energy efficient UAV-enabled mobile edge computing for IoT devices: A review," *IEEE Access*, vol. 9, pp. 127779–127798, 2021.

[68] F. Zhou, R. Q. Hu, Z. Li, and Y. Wang, "Mobile edge computing in unmanned aerial vehicle networks," *IEEE Wireless Commun.*, vol. 27, no. 1, pp. 140–146, Feb. 2020.

[69] W. Chen, X. Qiu, T. Cai, H.-N. Dai, Z. Zheng, and Y. Zhang, "Deep reinforcement learning for Internet of Things: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1659–1692, thirdquarter 2021.

[70] G. E. M. Abro, S. A. B. Zulkifli, R. J. Masood, V. S. Asirvadam, and A. Laouiti, "Comprehensive review of UAV detection, security, and communication advancements to prevent threats," *Drones*, vol. 6, no. 10, 2022, Art. no. 284.

[71] S. Hafeez et al., “Blockchain-assisted UAV communication systems: A comprehensive survey,” *IEEE Open J. Veh. Technol.*, vol. 4, pp. 558–580, 2023.

[72] H. Jin et al., “A survey of energy efficient methods for UAV communication,” *Veh. Commun.*, vol. 41, 2023, Art. no. 100594.

[73] A. Baltaci, E. Dinc, M. Ozger, A. Alabbasi, C. Cavdar, and D. Schupke, “A survey of wireless networks for future aerial communications (FACOM),” *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2833–2884, Fourthquarter 2021.

[74] O. S. Oubbati, M. Atiquzzaman, P. Lorenz, M. H. Tareque, and M. S. Hossain, “Routing in flying ad hoc networks: Survey, constraints, and future challenge perspectives,” *IEEE Access*, vol. 7, pp. 81057–81105, 2019.

[75] D. S. Lakew, U. Sa’ad, N.-N. Dao, W. Na, and S. Cho, “Routing in flying ad hoc networks: A comprehensive survey,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1071–1120, Secondquarter 2020.

[76] A. Fotouhi et al., “Survey on UAV cellular communications: Practical aspects, standardization advancements, regulation, and security challenges,” *IEEE Commun. Surveys Tuts.*, vol. 21, no. 4, pp. 3417–3442, Fourthquarter 2019.

[77] M. S. Mollel et al., “A survey of machine learning applications to handover management in 5G and beyond,” *IEEE Access*, vol. 9, pp. 45770–45802, 2021.

[78] H. Wang, H. Zhao, J. Zhang, D. Ma, J. Li, and J. Wei, “Survey on unmanned aerial vehicle networks: A cyber physical system perspective,” *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 1027–1070, Secondquarter 2020.

[79] C. Zhang, L. Zhang, L. Zhu, T. Zhang, Z. Xiao, and X. G. Xia, “3D deployment of multiple UAV-mounted base stations for UAV communications,” *IEEE Trans. Commun.*, vol. 69, no. 4, pp. 2473–2488, Apr. 2021.

[80] N. Nouri, J. Abouei, A. R. Sepasian, M. Jaseemuddin, A. Anpalagan, and K. N. Plataniotis, “Three-dimensional multi-UAV placement and resource allocation for energy-efficient IoT communication,” *IEEE Internet Things J.*, vol. 9, no. 3, pp. 2134–2152, Feb. 2022.

[81] Y. Yu, J. Tang, J. Huang, X. Zhang, D. K. C. So, and K. K. Wong, “Multi-objective optimization for UAV-assisted wireless powered IoT networks based on extended DDPG algorithm,” *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 6361–6374, Sep. 2021.

[82] M. M. H. Qazzaz, S. A. R. Zaidi, D. C. McLernon, A. Salama, and A. A. Al-Hameed, “Optimizing search and rescue UAV connectivity in challenging terrain through multi Q-learning,” in *Proc. 11th Int. Conf. Wireless Netw. Mobile Commun.*, 2024, pp. 1–8.

[83] W. Lu et al., “Secure transmission for multi-UAV-assisted mobile edge computing based on reinforcement learning,” *IEEE Trans. Netw. Sci. Eng.*, vol. 10, no. 3, pp. 1270–1282, May/Jun. 2023.

[84] Y. Wang, C. Zhang, T. Ge, and M. Pan, “Computation offloading via multi-agent deep reinforcement learning in aerial hierarchical edge computing systems,” *IEEE Trans. Netw. Sci. Eng.*, vol. 11, no. 6, pp. 5253–5266, Nov./Dec. 2024.

[85] S. Poudel and S. Moh, “Energy-efficient and fast MAC protocol in UAV-aided wireless sensor networks for time-critical applications,” *Sensors*, vol. 20, no. 9, 2020, Art. no. 2635.

[86] M. M. Alam and S. Moh, “Joint trajectory control, frequency allocation, and routing for UAV swarm networks: A multi-agent deep reinforcement learning approach,” *IEEE Trans. Mobile Comput.*, vol. 23, no. 12, pp. 11989–12005, Dec. 2024.

[87] M. Y. Arafat and S. Moh, “Routing protocols for unmanned aerial vehicle networks: A survey,” *IEEE Access*, vol. 7, pp. 99694–99720, 2019.

[88] 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.

[89] Z. Ning, Y. Yang, X. Wang, Q. Song, L. Guo, and A. Jamalipour, “Multi-agent deep reinforcement learning based UAV trajectory optimization for differentiated services,” *IEEE Trans. Mobile Comput.*, vol. 23, no. 5, pp. 5818–5834, May 2024.

[90] 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.

[91] L. Zhao, Y. Yao, H. Zhou, H. Wang, and V. C. M. Leung, “TD3-based collaborative computation offloading and charging scheduling in multi-UAV-assisted MEC networks,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2024, pp. 1–6.

[92] H. Li, J. Zhang, H. Zhao, Y. Ni, J. Xiong, and J. Wei, “Joint optimization on trajectory and resource for freshness sensitive UAV-assisted MEC system,” in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2024, pp. 1–6.

[93] R. Tang, R. Zhang, Y. Xu, and C. Yuen, “Deep reinforcement learning-based resource allocation for multi-UAV-assisted full-duplex wireless-powered IoT networks,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 10, no. 6, pp. 2236–2251, Dec. 2024.

[94] J. Tang, J. Song, J. Ou, J. Luo, X. Zhang, and K. K. Wong, “Minimum throughput maximization for multi-UAV enabled WPCN: A deep reinforcement learning method,” *IEEE Access*, vol. 8, pp. 9124–9132, 2020.

[95] H. Muslih, S. M. A. Kazmi, M. Mazzara, and G. Baye, “Cache sharing in UAV-enabled cellular network: A deep reinforcement learning-based approach,” *IEEE Access*, vol. 12, pp. 43422–43435, 2024.

[96] O. S. Oubbati, M. Atiquzzaman, H. Lim, A. Rachedi, and A. Lakas, “Synchronizing UAV teams for timely data collection and energy transfer by deep reinforcement learning,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6682–6697, Jun. 2022.

[97] Y. Liu, J. Yan, and X. Zhao, “Deep reinforcement learning based latency minimization for mobile edge computing with virtualization in maritime UAV communication network,” *IEEE Trans. Veh. Technol.*, vol. 71, no. 4, pp. 4225–4236, Apr. 2022.

[98] R. Zhong, X. Liu, Y. Liu, and Y. Chen, “Multi-agent reinforcement learning in NOMA-aided UAV networks for cellular offloading,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1498–1512, Mar. 2022.

[99] M. Saravanan and V. Pathmanaban, “Optimizing UAV base station positioning through quantum-inspired solution workflow,” in *Proc. 20th Int. Conf. Distrib. Comput. Smart Syst. Internet Things*, 2024, pp. 347–352.

[100] A. M. Seid, G. O. Boateng, B. Mareri, G. Sun, and W. Jiang, “Multi-agent DRL for task offloading and resource allocation in Multi-UAV enabled IoT edge network,” *IEEE Trans. Netw. Service Manag.*, vol. 18, no. 4, pp. 4531–4547, Dec. 2021.

[101] Z. U. A. Tariq, E. Baccour, A. Erbad, M. Hamdi, and M. Guizani, “RL-based adaptive UAV swarm formation and clustering for secure 6G wireless communications in dynamic dense environments,” *IEEE Access*, vol. 12, pp. 125609–125628, 2024.

[102] M. Alnakhli, E. M. Mohamed, and M. M. Fouda, “Bandwidth allocation and power control optimization for multi-UAVs enabled 6G network,” *IEEE Access*, vol. 12, pp. 67405–67415, 2024.

[103] O. T. H. Alzubaidi, M. N. Hindia, K. Dimyati, K. A. Noordin, and F. Qamar, “Interference mitigation based on joint optimization of NTBS 3D positions and RIS reflection in downlink NOMA hetNets,” *IEEE Access*, vol. 12, pp. 98750–98767, 2024.

[104] C. Huang, G. Chen, P. Xiao, Y. Xiao, Z. Han, and J. A. Chambers, “Joint offloading and resource allocation for hybrid cloud and edge computing in SAGINs: A decision assisted hybrid action space deep reinforcement learning approach,” *IEEE J. Sel. Areas Commun.*, vol. 42, no. 5, pp. 1029–1043, May 2024.

[105] K. K. Nguyen, S. R. Khosravirad, D. B. D. Costa, L. D. Nguyen, and T. Q. Duong, “Reconfigurable intelligent surface-assisted multi-UAV networks: Efficient resource allocation with deep reinforcement learning,” *IEEE J. Sel. Topics Signal Process.*, vol. 16, no. 3, pp. 358–368, Apr. 2022.

[106] C. Li, K. Jiang, G. He, F. Bing, and Y. Luo, “A computation offloading method for multi-UAVs assisted MEC based on improved federated DDPG algorithm,” *IEEE Trans. Ind. Informat.*, vol. 20, no. 12, pp. 14062–14071, Dec. 2024.

[107] Z. Zhu, S. Wan, P. Fan, and K. B. Letaief, “Federated multiagent actor-critic learning for age sensitive mobile-edge computing,” *IEEE Internet Things J.*, vol. 9, no. 2, pp. 1053–1067, Jan. 2022.

[108] Y. Lu, C. Jiang, L. Tan, J. Zhang, P. Zhang, and C. Rong, “UAV dynamic service function chains deployment based on security considerations: A reinforcement learning method,” *IEEE Internet Things J.*, vol. 11, no. 24, pp. 39731–39743, Dec. 2024.

[109] J. Kim, S. Park, S. Jung, and C. Cordeiro, “Cooperative multi-UAV positioning for aerial internet service management: A multi-agent deep reinforcement learning approach,” *IEEE Trans. Netw. Service Manag.*, vol. 21, no. 4, pp. 3797–3812, Aug. 2024.

[110] F. Khoramnejad, A. Syed, W. S. Kennedy, and M. Erol-Kantarci, "Energy and delay aware general task dependent offloading in UAV-aided smart farms," *IEEE Trans. Netw. Service Manag.*, vol. 21, no. 5, pp. 5033–5048, Oct. 2024.

[111] Y. Ding et al., "Online edge learning offloading and resource management for UAV-assisted MEC secure communications," *IEEE J. Sel. Topics Signal Process.*, vol. 17, no. 1, pp. 54–65, Jan. 2023.

[112] Z. Cui, K. Guan, C. Briso-Rodríguez, B. Ai, Z. Zhong, and C. Oestges, "Channel modeling for UAV communications: State of the art, case studies, and future directions," 2020, *arXiv:2012.06707*.

[113] PAWR, "Platforms for advanced wireless research," 2020. Accessed: May 19, 2025. [Online]. Available: https://advancedwireless.org/

[114] M. Polese et al., "Colosseum: The open RAN digital twin," *IEEE Open J. Commun. Soc.*, vol. 5, pp. 5452–5466, 2024.

[115] V. Marojevic, I. Guvenc, R. Dutta, M. L. Sichitiu, and B. A. Floyd, "Advanced wireless for unmanned aerial systems: 5G standardization, research challenges, and AERPAW architecture," *IEEE Veh. Technol. Mag.*, vol. 15, no. 2, pp. 22–30, Jun. 2020.

[116] X. Liu, Y. Liu, and Y. Chen, "Machine learning empowered trajectory and passive beamforming design in UAV-RIS wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 7, pp. 2042–2055, Jul. 2021.

[117] L. Li et al., "Millimeter-wave networking in the sky: A machine learning and mean field game approach for joint beamforming and beam-steering," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6393–6408, Oct. 2020.

[118] L. Wang, K. Wang, C. Pan, and N. Aslam, "Joint trajectory and passive beamforming design for intelligent reflecting surface-aided UAV communications: A deep reinforcement learning approach," *IEEE Trans. Mobile Comput.*, vol. 22, no. 11, pp. 6543–6553, Nov. 2023.

[119] J. Chen et al., "Deep reinforcement learning based resource allocation in multi-UAV-aided MEC networks," *IEEE Trans. Commun.*, vol. 71, no. 1, pp. 296–309, Jan. 2023.

[120] H. Xie, D. Yang, L. Xiao, and J. Lyu, "Connectivity-aware 3D UAV path design with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 70, no. 12, pp. 13022–13034, Dec. 2021.

[121] O. S. Oubbati, A. Lakas, and M. Guizani, "Multiagent deep reinforcement learning for wireless-powered UAV networks," *IEEE Internet Things J.*, vol. 9, no. 17, pp. 16044–16059, Sep. 2022.