TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 29/31 pp418–432
DOI: 10.26599/TST.2024.9010071
Volume 30, Number 1, February 2025

Deep-EERA: DRL-Based Energy-Efficient Resource Allocation in UAV-Empowered Beyond 5G Networks

Shabeer Ahmad, Jinling Zhang*, Ali Nauman, Adil Khan, Khizar Abbas, and Babar Hayat

Abstract: The rise of innovative applications, like online gaming, smart healthcare, and Internet of Things (IoT) services, has increased demand for high data rates and seamless connectivity, posing challenges for Beyond 5G (B5G) networks. There is a need for cost-effective solutions to enhance spectral efficiency in densely populated areas, ensuring higher data rates and uninterrupted connectivity while minimizing costs. Unmanned Aerial Vehicles (UAVs) as Aerial Base Stations (ABSs) offer a promising and cost-effective solution to boost network capacity, especially during emergencies and high-data-rate demands. Nevertheless, integrating UAVs into the B5G networks presents new challenges, including resource scarcity, energy efficiency, resource allocation, optimal power transmission control, and maximizing overall throughput. This paper presents a UAV-assisted B5G communication system where UAVs act as ABSs, and introduces the Deep Reinforcement Learning (DRL) based Energy Efficient Resource Allocation (Deep-EERA) mechanism. An efficient DRL-based Deep Deterministic Policy Gradient (DDPG) mechanism is introduced for optimal resource allocation with the twin goals of energy efficiency and average throughput maximization. The proposed Deep-EERA method learns optimal policies to conserve energy and enhance throughput within the dynamic and complex UAV-empowered B5G environment. Through extensive simulations, we validate the performance of the proposed approach, demonstrating that it outperforms other baseline methods in energy efficiency and throughput maximization.

Key words: Deep Reinforcement Learning (DRL); Unmanned Aerial Vehicles (UAVs); resource allocation; energy efficiency; 5G and beyond network

1 Introduction

The emergence of novel and interactive applications, like online video streaming, Augmented Reality (AR) or Virtual Reality (VR), Internet of Things (IoT), digital twins, online gaming, smart health care, and various industrial verticals, require high data rate, uninterrupted connection, and low latency. At present, mobile networks face a significant challenge in accommodating a wide range of applications, each with distinct Quality of Service (QoS) needs[1–3]. Due to that, a cost-effective deployment solution is needed to

• Shabeer Ahmad, Jinling Zhang, and Babar Hayat are with School of Electronic Engineering, Beijing University of Posts and Telecommunication, Beijing 1000876, China. E-mail: shabeer@bupt.edu.cn; zhangjl@bupt.edu.cn; babar@bupt.edu.cn.
• Ali Nauman is with Department of Information and Communication Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea. E-mail: anauman@ynu.ac.kr.
• Adil Khan is with School of Information Engineering, Xi’an Eurasia University, Xi’an 710065, China. E-mail: adilkhan@eurasia.edu.
• Khizar Abbas is with Department of Computer Science, Hanyang University, Seoul 04763, Republic of Korea. E-mail: khizarabbas@hanyang.ac.kr.
* To whom correspondence should be addressed.
Manuscript received: 2023-12-05; revised: 2024-02-16; accepted: 2024-04-08

enhance the spectral efficiency of Fifth-Generation (5G) networks in highly dense areas by increasing high data rates with minimum cost. Conversely, Unmanned Aerial Vehicles (UAVs) have emerged as viable and mostly adopted in performing search and rescue operations, emergency scenarios, efficient real-time monitoring, agriculture field monitoring and control, goods delivery, military operations, and communication networks. More specifically, UAVs are the best and most cost-effective solution to increase the communication network capacity and handle various scenarios, such as network devices malfunctioning, network failure due to natural distastes, abrupt surges in data rate demands due to a growing user base, as well as deployments in densely populated regions to enhance QoS and expand coverage in marine networks[4]. UAVs are most effective as Aerial Base Stations (ABSs), that can be deployed rapidly at low cost and fly anywhere without human involvement. In the beginning, UAVs were launched in military services. After that, they have been widely adopted in various civilian fields, such as agriculture, cargo, and wireless communications. UAVs provide essential features, like low cost, rapid deployment, adaptive communication, high mobility, Line of Sight (LoS) communication, and many more[5–7]. Due to the innovative features of UAVs, they have become an important part of Beyond 5G (B5G) wireless systems to ensure seamless connectivity and high throughput in a heterogeneous network context. The Third-Generation Partnership Project (3GPP) presents the use of UAVs with existing Long-Term Evolution (LTE) networks for enhancing network coverage and data rates[8]. UAVs can be used in wireless communication systems in various use cases: flying relay nodes or intermediate nodes for providing connectivity among two nodes such as source and destination, flying User Equipment (UE) for data collection purposes and remote sensing, and ABS, which can be deployed in infrastructure malfunction cases, to provide connectivity in natural disaster areas or the areas where it is hard to deploy 5G infrastructure and highly dense area to maintain service QoS. Moreover, UAVs can be used in communication networks as caching and power transmitters[9].

Integrating UAVs as ABS in wireless networks is a promising and noteworthy advancement, garnering considerable attention in academic and industrial circles. UAVs are poised to enhance system energy efficiency and user experiences in various settings, such as schools, shopping malls, and other high-traffic areas. They can serve as ABSs to bolster network capacity, coverage, reliability, and energy efficiency, while operating as mobile terminals within cellular networks. The resource allocation and power management research for UAV-supported 5G emergency wireless communications offers solutions for resource scarcity in disaster-stricken regions, extending communication endurance and improving user satisfaction[10,11].

As a supporting component of B5G systems, UAVs demonstrate impressive gains in expanding network coverage and enhancing overall network performance. However, several complex challenges remain to be addressed. In the context of UAVs as communication equipment, existing research has concentrated mainly on interference management, energy limitations, and trajectory designing and optimization. While trajectory optimization primarily aims to maximize their impact, it is vital to consider UAV stability and energy consumption as communication devices. Moreover, most of these studies have yet to account for 5G application scenarios, making them less applicable to existing 5G networks[12]. An exception is the work by Gao et al.[13], which introduces a 5G network system for emergency scenarios by integrating UAVs with a 5G system, offering greater flexibility and quicker response time than traditional emergency networks. The fusion of UAV and B5G networks offers the potential to create a more flexible communication system (for emergency cases) with promising development prospects. However, this integration introduces various new challenges of resource scarcity, energy efficiency, resource allocation, Base Station (BS) selection, optimal power control, and maximizing overall throughput. More specifically, the extensive deployment of densely packed BSs poses an energy consumption challenge, necessitating innovative approaches to increase energy efficiency while maintaining better Quality of Experience (QoE) to customers[14–16].

Moreover, Deep Reinforcement Learning (DRL) demonstrates remarkable adaptability in managing dynamic and intricate settings, making it a valuable tool for intelligent UAV control to enhance communication network performance[17]. Recently, DRL has emerged as a prominent research trend within Artificial Intelligence (AI). DRL has proven valuable for addressing resource allocation challenges in heterogeneous networks[18–20]. In this work[21], a Deep Q-Network (DQN) based resource allocation approach is introduced. This framework handles complex real-time control issues and leverages energy harvesting within Ultra-Dense Networks (UDNs) without requiring prior knowledge of Channel State Information (CSI), energy arrivals, and data rates. This innovative methodology mirrors human learning through extensive data training, a feat beyond the reach of conventional methods. It utilizes a dynamic trial-and-error exploration approach to engage with the environment. This leads to unparalleled levels of resource allocation automation and optimization, particularly for handling resource optimization with multiple purposes problems, involving energy, incomplete CSI, and data rates, which are typically insurmountable challenges using traditional convex optimization techniques.

1.1 Research motivation and contributions

AI approaches, particularly DRL methods, have emerged as promising solutions to effectively address the challenges associated with the UAVs as an ABS, including spectrum sharing, scheduling, optimal power control, resource allocation, link selection, energy efficiency, trajectory planning and management, QoS assurance, and throughput maximization. However, conventional Reinforcement Learning (RL) algorithms, such as Q-Learning (QL) and DQN, encounter limitations in handling the large continuous state and action spaces characteristic of the dynamic environment. Consequently, there is a pressing need for a robust and efficient DRL mechanism capable of effectively handling the vast amounts of data inherent in UAV-assisted B5G network environments while providing optimal decision-making capabilities.

To resolve the problems associated with UAVs used as ABSs in the B5G networks for increasing network coverage and capacity, we have introduced a DRL-based Energy-Efficient Resource Allocation (Deep-EERA) mechanism for maximizing energy efficiency and throughput. We transform the network resource allocation problem, like power allocation and throughput maximization, into an optimization problem, and design a Deep Deterministic Policy Gradient (DDPG) mechanism for optimal decision-making. The DDPG algorithm is best suited for our UAVs-assisted wireless network environment, because it is best for a dynamic environment with higher dimensionality of data. This DDPG agent learns from the simulation environment, where UAVs and BSs are used to serve users with better network capacity and provide seamless connectivity. The DDPG agent controls the power transmission of BSs, and UAVs, and maximizes the data rate by minimizing energy consumption. We have performed various energy efficiency, throughput, and energy consumption experiments to validate our Deep-EERA mechanism. The simulation outcomes reveal our proposed mechanism’s superiority over traditional approaches and achieve higher throughput by maximizing energy efficiency.

1.2 Paper organization

The remaining manuscript is organized as follows. Section 2 explains the relevant literature on using UAVs in wireless communication networks. Section 3 presents the DRL DDPG resource allocation mechanism in UAVs-assisted wireless network systems for B5G networks. This section also provides the architecture of the proposed system with the help of formulation and modeling. Section 4 explains the simulations and experimental details of the implemented system. It also explains the comparative analysis and results achieved through our system. In the end, we have concluded the paper by summarizing our achievements in this paper in Section 5.

2 Related Literature

This section extensively reviews the existing literature on integrating UAVs in wireless communication networks. It delves into the multifaceted challenges in this context, primarily focusing on throughput optimization, energy efficiency, QoS assurance, and the pivotal role of AI, RL, and DRL techniques in addressing these challenges.

Several noteworthy and significant studies[22, 23] have assumed either statistical or complete knowledge of environmental factors, including energy arrival patterns and real channel states. However, pinpointing the precise corresponding distribution for these factors can be challenging. Researchers have increasingly turned to RL methods as a practical and effective solution to address the uncertainties related to energy harvesting processes and channel states. Moreover, RL can offer flexibility for handling topological changes while demanding relatively low computational resources and implementation efforts compared to alternative paradigms, such as swarm intelligence, neural networks, and software agents in the context of an AI-enabled future[24].

In addition to that, we delve into the diverse RL-based approaches that have been applied in recent research literature. For instance, Savaglio et al.[25] developed an intelligent QL-enabled MAC protocol (QL-MAC) for Wireless Sensor Networks (WSNs). The QL-MAC protocol empowers individual nodes to independently optimize wake-up schedules, conserving energy by minimizing radioactivity through trial-and-error learning. However, it should be noted that applying QL to large-scale networks can introduce dimensionality problems due to the impracticality of constructing a comprehensive Q table.

Fadlullah et al.[26] demonstrated the suitability of Deep Learning (DL) technology for crafting model-free network strategies capable of addressing the challenges posed by the changing network environment driven by significant growing traffic demands and network applications. Mnih et al.[27] introduced a DRL mechanism which combines RL with DL, offering potential solutions for tackling complex challenges. Subsequently, Li et al.[28] and Du et al.[29] employed DQN to address radio resource allocation challenges in future networks. Their implemented DQN-based method integrates Deep Neural Networks (DNN) as function approximation into the QL algorithm, enabling handling large state spaces. Chu et al.[30] developed a scheduling algorithm using Long Short-Term Memory (LSTM) and DQN methods. They aimed to develop strategies to maximize and optimize the uplink rate in IoT cellular systems. Mohammadi et al.[31] ventured into extending the semi-supervised RL approach using DQN to address indoor localization based on Bluetooth with low power consumption. Their model incorporates a deep autoencoder model to learn optimal agent actions. Omoniwa et al.[32] implemented a multi-agent Double DQN (DDQN) framework to optimize UAV trajectory design while enhancing energy efficiency, particularly in mitigating UAV cell interference. Their approach demonstrates superior energy-saving capabilities through comprehensive simulations compared to conventional baseline methods. Reference [33] focuses on system throughput maximization and energy efficiency using energy cooperation and harvesting technology in ultra-dense networks. An optimal DRL algorithm is designed with the goal of throughput maximization within a limited time frame. This approach is devised to address the challenge of not having prior knowledge of channel conditions and energy arrival patterns. Additionally, they have developed a multiagent DRL method to tackle the dimensionality issue arising from the large number of states and action sets. A comparative analysis is conducted with two conventional algorithms: conservative and greedy. The simulation outcomes show the effectiveness of the multi-agent DRL mechanism in achieving higher average throughput.

Silver et al.[34] introduced a DDPG method recognized for its effectiveness in managing state spaces with high dimensionality and continuous action spaces. Similarly, Li et al.[21] proposed a DDPG-based method to optimize the power control scheme within the context of UDN networks. The results of the DDPG method show superior performance than other RL mechanisms, such as DQN and QL mechanisms. Nevertheless, it is important to mention that these DRL-based studies may not account for the increase in system users, potentially leading to expanded state and action dimensions and resulting in dimensionality challenges.

Li et al.[35] explored a novel online flight resource allocation system, called the DDPG-based Flight Resource Allocation Scheme (DDPG-FRAS). This mechanism has been developed to enhance the real-time optimization of UAV flight control operations, ensuring efficient scheduling of data collection tasks throughout its trajectory. The primary objective is the minimization of packet loss in the sensor network. Through empirical findings, it becomes evident that enlarging the buffer size can significantly reduce packet loss, yielding an impressive enhancement of up to 47.9%.

Peng and Shen[36] explored the combined challenges of resource management and vehicle association within the MEC and UAVs-supported vehicular network. A multi-objective resource optimization has been formulated to efficiently manage the computational capabilities, caching resources, and spectrum allocation for MEC-enabled UAVs serving as ABS. To address the issue of stringent delay requirements of vehicular networks, they developed a DRL-based DDPG for efficient resource management, which gives an optimal decision on resource allocation and vehicle association. The experimental outcomes reveal that the proposed mechanism outperforms a random allocation scheme regarding QoS and satisfying delay.

Nguyen et al.[37] combined Reconfigurable Intelligent Surfaces (RIS) with UAVs to improve network overall performance and capacity. The primary aim of this study is to optimize the network's energy efficiency. To accomplish this objective, they introduced a coordinated optimization process involving allocating power for UAVs and the phase shift matrix for RIS. They introduced a DRL approach to address this continuous optimization challenge, particularly in contexts with time-varying channels. This DRL technique enables centralized decision-making while adapting to dynamic environmental conditions. Additionally, they introduced a parallel learning approach to reduce the latency associated with information transmission requirements in the centralized approach. Numerical results demonstrate the significant advantages of the implemented schemes over conventional methods regarding energy efficiency, flexibility, and processing time.

Cui et al.[38] introduced a Multi-Agent RL (MARL) mechanism to attain efficient long-term resource allocation policies in a multi-UAV context. In MARL, each UAV functions as an independent learning agent, and the resource allocation decisions are treated as actions taken by these UAV agents. The MARL approach allows agents to learn its optimal policy based on local observations independently but employs a QL-based common structure. The implemented mechanism demonstrates satisfactory performance, particularly compared to scenarios involving the complete information exchange among UAVs.

Chen et al.[39] took on the complex challenge of optimizing caching and resource allocation in a network where cache-supported UAVs serve Ground Users (GUs) operating across Unlicensed Long-Term Evolution (LTE-U) and licensed LTE networks. Their proposed model is centred around GUs with access to both types of bands and receiving content through both links. The comprehensive problem of spectrum allocation, content caching, and user association is structured as an optimization problem. To address this multifaceted challenge, they introduced a distributed Machine Learning (ML) based algorithm, called the Liquid State Machine (LSM). The LSM method empowers the cloud to predict the distribution of users' content requests, even with limited users and network information. Furthermore, it enables UAVs to autonomously select optimal resource allocation policies based on the network's current conditions.

In Ref. [40], Seid et al. introduced a novel Multi-Agent Federated RL (MAFRL) mechanism for efficiently allocating resources in a UAV-assisted healthcare context. The MAFRL algorithm tackles the issues related to resource allocation and computation offloading by presenting them as an optimization problem within the realm of Federated Learning (FL) that involves numerous participants. The primary objectives of MARFL are to minimize energy consumption and maintain QoS. Simulation results are performed on the heartbeat dataset, demonstrating that the MAFRL algorithm offers significant advantages over baseline learning algorithms. MAFRL enhances privacy, reduces cost, and enhances accuracy, making it a promising solution for resource allocation in healthcare systems.

In Ref. [41], Li et al. introduced an efficient resource allocation through DRL to ensure continuous network coverage. It has the unique capability to dynamically adjust neural network structures, making it effective in meeting coverage needs by jointly allocating power and subchannels to GUs. It is observed from the experiments that the proposed mechanism reduced rate variance by 66.7% and increased spectral efficiency by 34.7% as compared to benchmark algorithms.

In Ref. [42], Du et al. focused on a UAV-assisted MEC for providing data preprocessing services to IoT devices. A single UAV as an edge is considered in this context, which hovers to different locations in varying time slots for data collection and processing. The primary aim of this work is to maximize the energy efficiency of UAVs, encompassing the energy spent while hovering and during computation. This is achieved by optimizing several factors, including UAV hovering duration, task scheduling, and allocating resources for the IoT devices tasks by assuring QoS requirements. They introduced an iterative method to achieve more accurate sub-optimal solutions and policies. Simulation results validate the effectiveness of the implemented approach, demonstrating its superiority to other benchmark approaches.

Peng and Shen[43] worked on multi-domain resource management within UAVs-supported vehicular networks. The major objective is to support on-demand service provisioning by optimal resource allocation to vehicles in MEC-assisted UAVs and BSs context.

Without a central controller, the optimization resource allocation problem is tackled at the MEC using a distributed approach. The main aim is to optimize the number of tasks offloaded, considering their varying QoS demands. A DRL-based DDPG mechanism is proposed for optimal and efficient allocation of resources and intelligent vehicle association decisions by providing QoS assurance.

Wang et al.[44] employed UAVs as BSs for ensuring MEC capabilities to GUs. A DRL-based multi-agent method is proposed for optimal UAV path planning to maximize overall energy efficiency. In addition, UAVs load balancing and GUs task offloading load balancing. Additionally, fairness is considered to ensure UAV load balancing and balance in UE task offloading. Experimental results demonstrate superior performance of the implemented method to other baseline works, offering a more efficient and balanced solution for UAV-based MEC services.

The literature review highlights the significant role of UAVs as a cost-effective and reliable technology in B5G communication systems, particularly for enhancing network capacity and coverage in densely populated areas or emergency situations. Despite the numerous advantages of UAVs as ABS, several challenges persist in optimizing their performance, including optimal power allocation, resource allocation, link selection, energy efficiency, trajectory planning and management, QoS assurance, and throughput maximization. To address these challenges, researchers have turned to AI approaches, particularly RL methods, which hold promise for tackling these complex problems. While some studies have employed traditional RL algorithms, such as QL and DQN, for tasks like optimal trajectory planning, throughput maximization, resource allocation, and link selection, these approaches face limitations in handling large continuous state and action spaces, particularly in dynamic environments, like UAVs-assisted B5G networks.

In contrast, the DDPG method has emerged as a promising solution due to its ability to handle high-dimensional state-action spaces and complex environments. While previous studies have utilized DDPG for tasks, like optimal trajectory planning and link selection in static environments, optimal resource allocation in UAVs-assisted B5G networks remains an area of ongoing research.

Building upon the advantages of DDPG over traditional RL algorithms, like QL and DQN, we have developed a novel DDPG-based mechanism to address the multi-objective challenges of energy efficiency and throughput maximization in UAVs-assisted B5G networks. By leveraging DDPG's capabilities in handling complex environments and high-dimensional spaces, our proposed approach aims to overcome the limitations of existing methods and pave the way for enhanced performance and efficiency in UAVs-assisted B5G communication systems.

3 Proposed System

3.1 System overview

In this study, we propose an efficient system utilizing UAVs to enhance network capacity and improve user QoS. Our innovative UAV-assisted wireless network is designed to accommodate a wide range of emerging 5G and beyond network services. Figure 1 illustrates the overall architecture of our system, which includes multiple BSs, UAVs functioning as ABSs, a 5G core network, 5G users, and a centralized controller responsible for managing both the BSs and UAVs. The UAVs establish connectivity in densely populated areas or where certain BSs experience failures, effectively alleviating the strain on overloaded BSs resulting from a surge in users within a specific area. Thus, UAVs represent an optimal solution for addressing these issues. In our system, we have used UAVs to mitigate situations involving overloaded BSs and malfunctions. Nonetheless, incorporating UAVs in this context presents challenges, including maximizing data transmission rates, optimising power allocation, and enhancing energy efficiency. To address these challenges, we have introduced a DDPG-based model to achieve higher throughput and maximize energy efficiency, resulting in improved customer QoE.

3.2 System model and problem formulation

The system model consists of multiple UAVs serving as flying BSs, which are denoted as $S_{\text{UAV}} = \text{UAV}_1, \text{UAV}_2, \text{UAV}_3, \ldots, \text{UAV}_n$. It also includes a set of multiple BSs denoted as $S_{\text{BS}} = \text{BS}_1, \text{BS}_2, \ldots, \text{BS}_m$, and K GUs denoted as $S_{\text{GU}} = \text{GU}_1, \text{GU}_2, \text{GU}_3, \ldots, \text{GU}_k$. As presented in Fig. 1, the centralized controller is responsible for controlling the deployed UAVs and BSs. The GUs are randomly distributed to UAVs and BSs. In our approach, we suppose the centralized controller has all the information related to

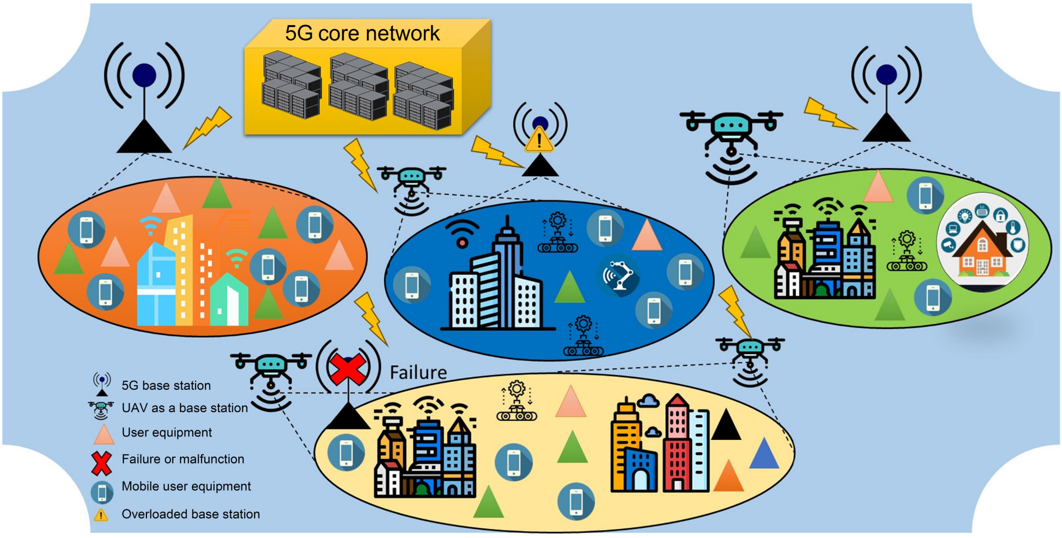

Fig. 1 Architecture of UAVs assisted B5G network for enhancing the network coverage and capacity in emergency cases or supporting increasing users in highly dense areas to maintain QoS.

the UAVs, BSs, and GUs, such as GUs location, transmission power, and Channel Quality Information (CQI). Based on information from the environment, the centralized controller can perform the operations of optimal BS selection for each user and control the resource allocation of BSs. In our setup, suppose that the UAVs are operating in the air and BSs are operating in a fixed location. For example, The location of UAV$_i$ is defined as $\mathrm{Loc}_{\mathrm{UAV}_n} = (X_n, Y_n, Z_n)$ where X_n, Y_n, and Z_n are the coordinates of n-th UAV. Similarly, BS location is defined as $\mathrm{Loc}_{\mathrm{BSm}} = (X_m, Y_m, Z_m)$ where X_m, Y_m, and Z_m are the coordinates information of m-th BS. Additionally, GUs are distributed randomly at the location of $\mathrm{Loc}_{\mathrm{GU}k} = (X_k, Y_k, Z_k)$ where X_k, Y_k, and Z_k are the coordinate information of k-th user.

In the B5G context, the system's real-time feedback mechanism is solely influenced by the state and actions taken in the most recent time slot. It does not depend on the past states, aligning with the fundamental characteristics of a Markov Decision Process (MDP). Leveraging this similarity to MDPs, we have employed a DRL-DDPG algorithm to address this optimization challenge effectively. The system's state and action spaces evolve through training and iterative processes, enabling the generation of approximation functions. As a result, we can reorganize our system into an MDP, which includes defining the state-action pairs and reward function.

The global state space of k GUs is defined as follows:

$$S^t = \mathrm{TR}_1^t,\ \mathrm{TR}_2^t,\ \mathrm{TR}_3^t,\ \mathrm{TR}_4^t,\ \ldots,\ \mathrm{TR}_n^t,\ \mathrm{PC}_{\mathrm{total}}^t \tag{1}$$

where TR_n^t shows each GU's current data transmission rate at timestamp t. $\mathrm{PC}_{\mathrm{total}}^t$ is the total power consumption at timestamp t by the UAVs and BSs. $\mathrm{PC}_{\mathrm{total}}^t$ is described in the following:

$$\mathrm{PC}_{\mathrm{total}}^t = \sum_{k}^{K} I^k(\mathrm{TP}_m^t) + \sum_{k}^{K}(1 - I^k)(\mathrm{TP}_n^t) \tag{2}$$

where TP_m^t is transmission power assigned to the k-th GU by the m-th BS and TP_n^t shows assigned transmission power by the n-th UAV to the R-th GU at timestamp t. Moreover, I^k is an indicator that presents the connection status of GU to BS and UAV.

The main goal of the proposed DRL mechanism is to maximize the data transmission rate TR_n^t by controlling the power allocation of UAVs and BSs to each GU. So, our action space A^t is to achieve optimal power distribution, as defined as follows:

$$A^t = \mathrm{TP}_1^t,\ \mathrm{TP}_2^t,\ \mathrm{TP}_3^t,\ \ldots,\ \mathrm{TP}_n^t \tag{3}$$

where A^t illustrates the action in time slot t. Eq. (4) illustrates transmission power TP_n^t at time slot t,

$$\mathrm{TP}_n^t = \mathrm{TP}_{\mathrm{Min}} \left[\frac{\mathrm{TP}_{\mathrm{Max}}}{\mathrm{TP}_{\mathrm{Min}}}\right]^{\frac{g}{l-1}} \tag{4}$$

where $\mathrm{TP}_{\mathrm{Min}}$ and $\mathrm{TP}_{\mathrm{Max}}$ present the minimum and maximum transmission power, respectively, and l shows the transmission power level.

We have formulated the reward function R^t based on the total data transmission rate TR_n^t, which is a very important parameter to maximize the performance of our system. The reward function R^t is as follows:

$$R^t = \sum_{k}^{K} I_t^k (TR_m^t) + \sum_{k}^{K} (1 - I_t^k)(TR_n^t) \qquad (5)$$

where TR_m^t illustrates the data transmission rate of k GU at timestamp t by the m-th BS and TR_n^t presents the data transmission rate of the k-th GU at timestamp t by the n-th UAV.

3.3 Deep deterministic policy gradient for optimal resource allocation

RL revolves around agents' interaction, environment, and the reward system. Agents operate within the environment, taking various actions A^t that lead to the generation of new states S'. In response to these evolving states, the learning agent obtains a response in the form of R^t reward, which can be positive or negative. This reward R^t influences the agent's decision-making process, as it seeks to determine the optimal state-action pairs, essentially crafting a strategy through exploration to optimize the total reward. Within the MDP framework, the agent engages with the environment by adhering to a decision strategy or policy denoted as Π. This policy Π essentially maps states to the corresponding actions. We aim to ascertain the optimal policy Π' to maximize the Q-value function as presented in the following[24]:

$$\Pi'(S) = \text{argmax}\,(Q\,(S,\, A)) \qquad (6)$$

The optimal Q' function is presented in the following[24]:

$$Q'(S,\, A) = \alpha\,[R^t + \gamma \max_{A'} Q\,(S',\, A')] \qquad (7)$$

where S' and A' represent the new state and action achieved after the agent's action, denoted as A, in state S. Additionally, α and γ are the Learning Rate (LR) and discount factor, respectively.

The most common RL algorithms are QL and DQN, but they suffer from high dimensional issues while training in a dynamic environment like our mechanism. Due to that, We have adopted the latest DDPG algorithm, which is a development of the actor-critic approach, leveraging DNNs to approximate value functions and policy[45]. DDPG distinguishes itself from traditional RL algorithms by its capability to tackle the challenges posed by high-dimensional optimization problems characterized by extensive state and action spaces. Moreover, DDPG excels in making effective decisions when dealing with continuous action spaces, such as in our communication system. The DDPG algorithm encompasses actor, critic, actor target θ^a, and critic target θ^c models, as presented in Fig. 2. The actor-network θ^a plays a crucial role in selecting the current action on the basis of the current state while also handling the acquisition of the subsequent state and reward information. In contrast, the critic network θ^c is tasked to compute the current Q value. We maintain actor target θ^a and critic-target θ^c networks, replicas of the actor θ^a and critic θ^c networks, respectively. Their network parameters are updated using a soft update technique to enhance training stability. The critic network θ^c is optimized in each iteration by minimizing the loss function followed by temporal difference error, as defined in the following:

$$\text{Loss}_{\theta^c} = \alpha\,[(Z^t - Q\,(S^t,\, A^t;\, \theta^c))^2] \qquad (8)$$

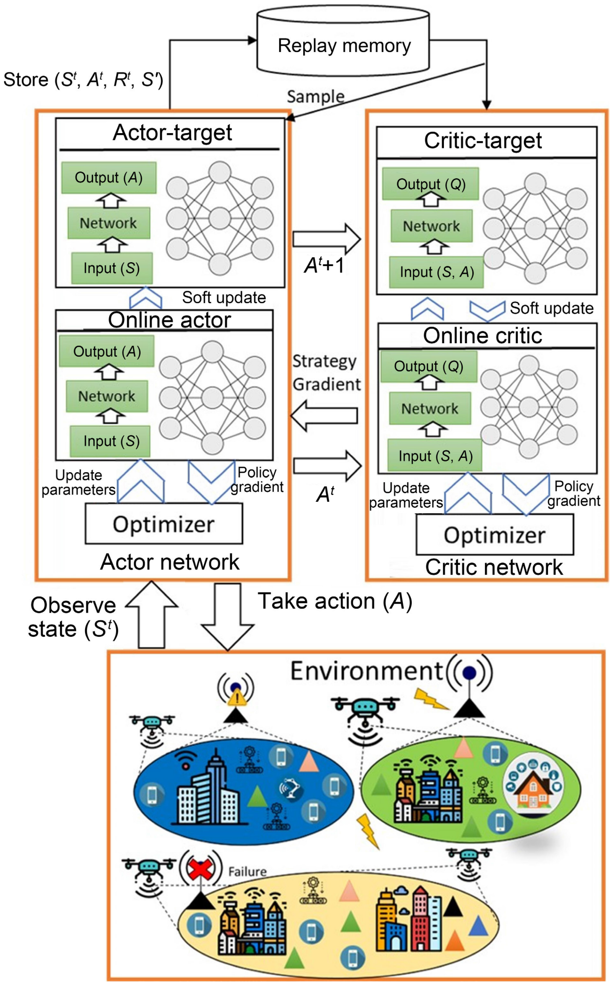

Fig. 2 Architecture of DRL-based DDPG mechanism for energy-efficient resource allocation in UAVs assisted future network.

On the other side, during training, the actor-network θ^a has to maximize the objective function, which follows a policy gradient method[46, 47]. We have embraced the most recent DDPG algorithm, which is a development of the actor-critic approach, leveraging DNNs to approximate value functions and policy[45].

Z^t is the expected return and is defined in the following:

$$Z^t = R^t + \alpha Q\left(S', \Pi\left(S'; \theta^{a'}\right); \theta^{c'}\right) \tag{9}$$

We use a strategy to keep Z^t stable during the training phase by slowly updating the target network parameters.

Algorithm 1 explains the process of the proposed DDPG-based mechanism for achieving energy-efficient optimal resource allocation policies. The DDPG agent interacts with the communication system environment by observing state S_t at time step t and takes an action A^t, receiving reward R^t and moving to new state S'. Every experience of an agent is stored in a memory buffer (S^t, A^t, R^t, S'). The memory buffer is initialized with limited memory, and in case memory becomes full, the oldest experience is deleted to store new experiences. We randomly select mini-batches from the memory buffer to train the actor θ^a and critic θ^c networks.

Algorithm 1 DDPG-based energy efficient dynamic resource allocation

```
Initialize memory replay buffer $M_b$ to $\text{Size}_{\max}$;
Initialize parameters of two actor networks $\theta^a$ and $\theta^{a'}$;
Initialize parameters of two critic networks $\theta^c$ and $\theta^{c'}$;
for ep = 1 : N do
    Set initial state and the cumulative reward at each episode $R_{\text{ep}}$ to zero
    for step t do
        Select action $A^t$ by $\theta^a$ actor online network;
        Perform action $A^t$, receive reward $R^t$, and move to the new $S'$ state;
        if $|M_b| < \text{Size}_{\max}$ then
            Store transition experience ($S^t$, $A^t$, $R^t$, $S'$) to $M_b$;
        else
            Remove oldest transition values ($S^t$, $A^t$, $R^t$, $S'$) from the memory buffer $M_b$;
            Randomly select mini_batch samples from $M_b$ and input to both networks;
            Update actor and critic online networks parameters using loss function $\text{Loss}_{\theta^c}$ and policy gradient;
            Update target networks $\theta^{a'}$ and $\theta^{c'}$;
            $R_{\text{ep}} = R_{\text{ep}} + R^t$;
        end if
    end for
end for
Final policy: Optimal resource allocation decision policy that maximizes energy efficiency and throughput
```

4 Simulation Result and Analysis

4.1 System settings

The proposed UAVs-assisted energy-efficient resource allocation mechanism is implemented through the DDPG model using Python and TensorFlow. Table 1 presents the details of the experimental and simulation parameters. The convergence of our DDPG algorithm using various LR values, specifically 0.001, 0.003, and 0.01, is depicted in Fig. 3. It is evident that the average cumulative reward exhibits an increasing trend and tends to stabilize after approximately 150 episodes for all three LR settings. The DDPG algorithm displays slow convergence with a small LR but converges more rapidly as the LR increases. However, it is important to note that using an LR of 0.1 may lead the DDPG algorithm to converge to sub-optimal values, and further increasing the LR does not necessarily yield better results. In our experiments, we obtain superior results with an LR of 0.001.

Table 1 Experimental parameters and specification of DDPG resource allocation mechanism.

Name of parameter	Specification
Bandwidth of BS	10 MHz
Bandwidth of UAV	15 MHz
Gaussian noise	−110 dBm
TP_{Max} of BS	20 dBM
TP_{Max} of UAV	20 dBM
Channel gain BS	−60 dB
Channel gain UAV	−50 dB
Path loss-index of BS	2
Path loss-index of UAV	2
Batch-size	256
Memory buffer size	5000
Initial α	0.01
γ	0.9

4.2 Experimental results and discussion

We have performed the experimental analysis of achieved throughput, energy consumption, and energy

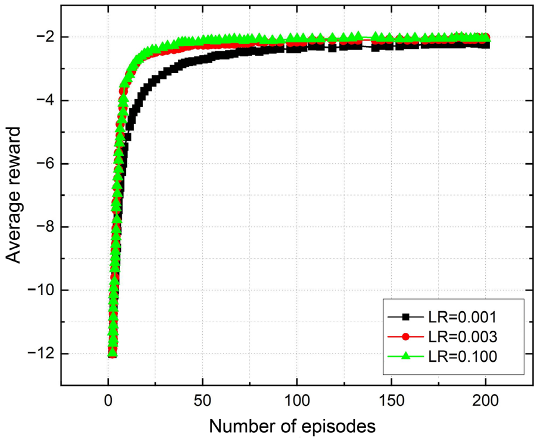

Fig. 3 Convergence of DDPG algorithm on different learning rates.

efficiency achieved through our DRL-based resource allocation mechanism for UAVs-assisted B5G network. Initially, we use one UAV and three BSs for our experimental setting. Energy consumption and throughput are two important parameters to validate the performance of a system, because energy consumption causes an increase in the total operational expenses, and low throughput violates the SLA and impacts user experience. Due to that, we have performed throughput analysis, energy efficiency, and energy consumption to validate our system. Figure 4 illustrates the results of the total throughput achieved in two different use cases, with UAVs and without UAVs and multiple BSs. It can be observed that by increasing BSs, the system's throughput increases in both cases. More importantly, the UAV-assisted case achieves better total throughput than the case without UAVs by using the same number of BSs.

On the other hand, Fig. 5 illustrates the energy consumption results achieved from our experiment using different numbers of BSs with and without UAVs. It can be seen that the UAV-assisted case uses higher energy consumption compared to those without UAVs. This is because it provides higher throughput to the communication system. So, it is visible from the results that the UAVs are very efficient in increasing the system throughput and coverage, and are very beneficial to maintaining QoS for users in dense areas or emergency cases.

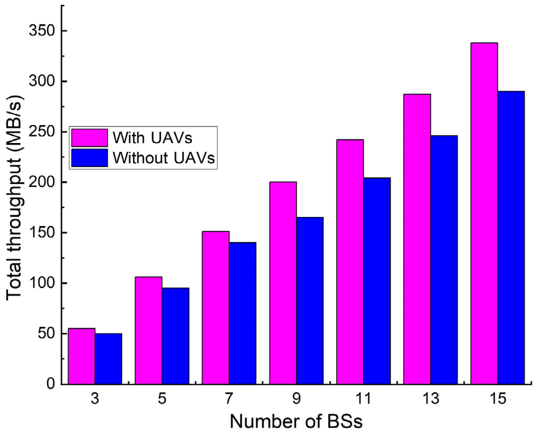

Fig. 4 Total throughput by using multiple base stations in two scenarios: With UAVs and without UAVs.

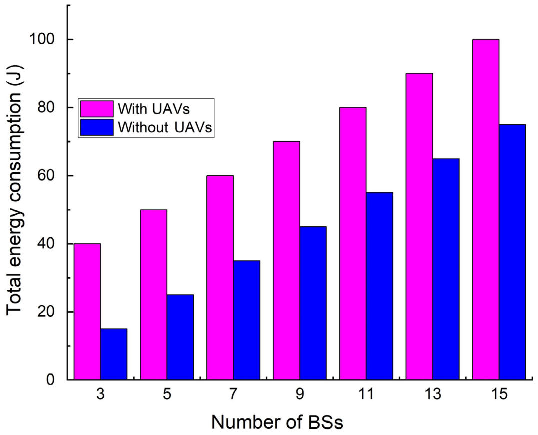

Fig. 5 Total energy consumption by using multiple base stations in two scenarios: With UAVs and without UAVs.

Figure 6 illustrates the comparative analysis of our system with existing techniques, such as QL, DQN, and throughput maximization for energy efficiency. Figure 6 shows the experimental results of energy efficiency achieved by all the approaches using different numbers of users. It is observed that our DDPG mechanism achieves higher energy efficiency than the other approaches. Conversely, Fig. 7 illustrates the energy efficiency results achieved by all four algorithms using multiple BSs with different numbers of users.

As the number of BSs increases, a noteworthy rise in energy efficiency is observed across all four considered methods. This uptrend in energy efficiency is primarily attributed to the decreasing user density resulting from the expanding BS coverage. In particular, the proposed DDPG mechanism exhibits a distinct pattern, maintaining a relatively stable energy efficiency, which tends to be higher in comparison to the other three algorithms. This implies that in situations with a surplus of BSs and a sparse user population, the application of our DDPG algorithm can notably boost system energy efficiency, all while maintaining a high

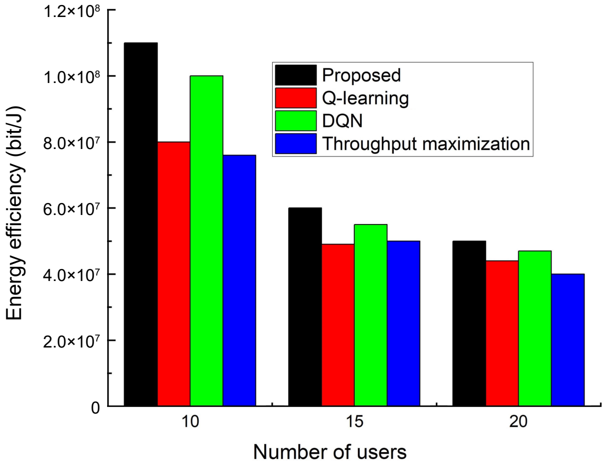

Fig. 6 Comparative analysis of energy efficiency with respect to number of users.

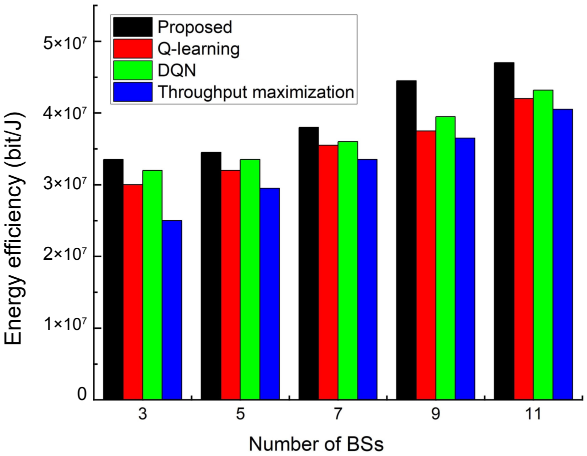

Fig. 7 Comparative analysis of the energy efficiency with respect to the number of BSs.

level of QoE. So, it is visible that our DDPG mechanism achieves more energy efficiency, provides higher throughput than all the methods, and shows satisfactory results.

In comparison to previous approaches, such as those by Du et al.[29], Chu et al.[30], and Omoniwa et al.[32], our Deep-EERA mechanism stands out for its comprehensive optimization of energy efficiency and throughput in B5G dynamic environments. Du et al.[29] focused solely on static BS radio resource allocation, while Chu et al.[30] optimized uplink rates without considering energy efficiency. Omoniwa et al.[32] achieved good results for energy efficiency but limited their scope to UAV trajectory design without integrating B5G BS communication environments. In contrast, Li et al.[33] addressed energy efficiency and downlink throughput in a UDN environment but faced challenges with static environments and dimensionality issues. Our Deep-EERA system, employing the DDPG method, excels by overcoming these challenges, learning optimal policies in dynamic B5G environments, and delivering superior energy savings and QoS improvements for B5G communication environments.

4.3 Limitations

Our study highlights the advantages and effectiveness of integrating UAVs into communication systems to address coverage, capacity, QoS, and QoE challenges. Our proposed DDPG-based resource allocation system demonstrates promising results in achieving optimal resource allocation objectives, particularly focused on energy conservation and enhanced QoS through throughput maximization in a B5G communication framework.

Despite the numerous advantages of our system, several limitations warrant consideration. Firstly, our mechanism requires further exploration to optimize link selection, which remains an area for future research and enhancement. Secondly, while we have validated the proposed mechanism in a simulation environment, the performance may vary when deployed in real-world scenarios due to factors, such as signal interference, hardware variations, and weather conditions. Additionally, fine-tuning the DDPG model is necessary to adapt to different network topologies and configurations, as its performance may be influenced by environmental variations. Lastly, scaling up the system to accommodate a larger number of UAVs, BSs, and GUs may introduce challenges related to decision-making, which require further investigation and optimization.

5 Conclusion

In conclusion, this study addresses the challenges associated with integrating UAVs as aerial BSs in 5G and beyond networks, aiming to enhance network coverage and capacity. We introduce an efficient DRL approach for resource allocation, prioritizing maximizing energy efficiency and throughput. By formulating resource allocation, including power allocation, as an optimization problem, we develope a DDPG algorithm tailored to dynamic environments characterized by high-dimensional data, making it well-suited for our UAV-assisted wireless network setting. Our DDPG agent is trained in a simulation environment where UAVs and BSs work together to serve users, enhancing network capacity and ensuring seamless connectivity. This agent effectively controlls power transmission for both BSs and UAVs, ultimately maximizing data rates while minimizing energy consumption. Through extensive experimentation focusing on energy efficiency, throughput, and energy consumption, we validate the effectiveness of the implemented mechanism.

The simulation results unequivocally reveal the superiority of the proposed DDPG mechanism over existing approaches, achieving higher throughput while maximizing energy efficiency. This work marks a significant step forward in leveraging UAVs as integral components of the B5G system, offering an efficient and promising solution to increase network capability in scenarios ranging from emergencies to high-demand situations. The application of DRL-based resource allocation strategies presents a robust framework for future advancements in wireless communication networks.

In our future research efforts, we aim to expand upon our Deep-EERA mechanism to address the complexities of UAVs-assisted B5G networks and tackle multi-objective challenges related to energy efficiency, trajectory planning, throughput maximization, and optimal link selection.

To achieve this, we plan to delve deeper into the intricacies of UAVs-assisted B5G networks, considering various real-world scenarios and network configurations. Furthermore, we intend to implement and validate our Deep-EERA mechanism in real-time B5G network environments, where we will closely examine and address the practical challenges and constraints encountered in real-world settings.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (No. 62271063), the National Key Laboratory of Science and Technology on Vacuum Electronics, and the Director Fund of Beijing Key Laboratory of Space-ground Interconnection and Convergence.

References

[1] A. Dogra, R. K. Jha, and S. Jain, A survey on beyond 5G network with the advent of 6G: Architecture and emerging technologies, *IEEE Access*, vol. 9, pp. 67512–67547, 2021.

[2] 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, 2019.

[3] K. Abbas, T. A. Khan, M. Afaq, and W. C. Song, Network slice lifecycle management for 5G mobile networks: An intent-based networking approach, *IEEE Access*, vol. 9, pp. 80128–80146, 2021.

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

[5] L. Gupta, R. Jain, and G. Vaszkun, Survey of important issues in UAV communication networks, *IEEE Commun. Surv. Tutor.*, vol. 18, no. 2, pp. 1123–1152, 2016.

[6] P. W. Khan, G. Xu, M. A. Latif, K. Abbas, and A. Yasin, UAV's agricultural image segmentation predicated by Clifford geometric algebra, *IEEE Access*, vol. 7, pp. 38442–38450, 2019.

[7] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, et al., Designing and implementing future aerial communication networks, *IEEE communications magazine*, vol. 54, no. 5, pp. 26–34, 2016.

[8] S. Chandrasekharan, K. Gomez, A. Al-Hourani, S. Kandeepan, T. Rasheed, L. Goratti, L. Reynaud, D. Grace, I. Bucaille, T. Wirth, et al., Designing and implementing future aerial communication networks, *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 26–34, 2016.

[9] Z. Xiao, P. Xia, and X. G. Xia, Enabling UAV cellular with millimeter-wave communication: Potentials and approaches, *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 66–73, 2016.

[10] A. Sharma, P. Vanjani, N. Paliwal, C. M. W. Basnayaka, D. N. K. Jayakody, H. C. Wang, and P. Muthuchidambaranathan, Communication and networking technologies for UAVs: A survey, *J. Netw. Comput. Appl.*, vol. 168, p. 102739, 2020.

[11] P. W. Khan, K. Abbas, H. Shaiba, A. Muthanna, A. Abuarqoub, and M. Khayyat, Energy efficient computation offloading mechanism in multi-server mobile edge computing—an integer linear optimization approach, *Electronics*, vol. 9, no. 6, p. 1010, 2020.

[12] R. Shahzadi, M. Ali, H. Z. Khan, and M. Naeem, UAV assisted 5G and beyond wireless networks: A survey, *J. Netw. Comput. Appl.*, vol. 189, p. 103114, 2021.

[13] Y. Gao, J. Cao, P. Wang, J. Wang, M. Zhao, S. Cheng, S. Hu, and W. Lu, UAV based 5G wireless networks: A practical solution for emergency communications, in *Proc. 2020 XXXIIIrd General Assembly and Scientific Symp. Int. Union of Radio Science*, Rome, Italy, 2020, pp. 1–4.

[14] Q. Wu, Y. Zeng, and R. Zhang, Joint trajectory and communication design for multi-UAV enabled wireless networks, *IEEE Trans. Wirel. Commun.*, vol. 17, no. 3, pp. 2109–2121, 2018.

[15] 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. Serv. Manage., vol. 18, no. 4, pp. 4531–4547, 2021.

[16] Q. Wu, J. Xu, Y. Zeng, D. W. K. Ng, N. Al-Dhahir, R. Schober, and A. L. Swindlehurst, A comprehensive overview on 5G-andbeyond networks with UAVs: From communications to sensing and intelligence, *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 2912–2945, 2021.

[17] K. Abbas, J. Hong, N. Van Tu, J. H. Yoo, and J. W. K. Hong, Autonomous DRL-based energy efficient VM consolidation for cloud data centers, *Phys. Commun.*, vol. 55, p. 101925, 2022.

[18] H. Ye, G. Y. Li, and B. H. F. Juang, Deep reinforcement learning based resource allocation for V2V communications, *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3163–3173, 2019.

[19] X. He, K. Wang, H. Huang, T. Miyazaki, Y. Wang, and S. Guo, Green resource allocation based on deep reinforcement learning in content-centric IoT, *IEEE Trans. Emerg. Top. Comput.*, vol. 8, no. 3, pp. 781–796, 2020.

[20] X. Liao, J. Shi, Z. Li, L. Zhang, and B. Xia, A model-driven deep reinforcement learning heuristic algorithm for resource allocation in ultra-dense cellular networks, *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 983–997, 2020.

[21] H. Li, H. Gao, T. Lv, and Y. Lu, Deep Q-learning based dynamic resource allocation for self-powered ultra-dense networks, in *Proc. 2018 IEEE Int. Conf. Communications Workshops* (*ICC Workshops*), Kansas City, MO, USA, 2018, pp. 1–6.

[22] Z. Wang, V. Aggarwal, and X. Wang, Power allocation for energy harvesting transmitter with causal information, *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 4080–4093, 2014.

[23] H. Zhang, S. Huang, C. Jiang, K. Long, V. C. M. Leung, and H. V. Poor, Energy efficient user association and power allocation in millimeter-wave-based ultra dense networks with energy harvesting base stations, *IEEE J. Sel. Areas Commun.*, vol. 35, no. 9, pp. 1936–1947, 2017.

[24] R. S. Sutton and A. G. Barto, *Reinforcement Learning*: *An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.

[25] C. Savaglio, P. Pace, G. Aloi, A. Liotta, and G. Fortino, Lightweight reinforcement learning for energy efficient communications in wireless sensor networks, *IEEE Access*, vol. 7, pp. 29355–29364, 2019.

[26] Z. M. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems, *IEEE Commun. Surv. Tutor.*, vol. 19, no. 4, pp. 2432–2455, 2017.

[27] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, Playing atari with deep reinforcement learning, arXiv preprint arXiv: 1312.5602, 2013.

[28] X. Li, J. Fang, W. Cheng, H. Duan, Z. Chen, and H. Li, Intelligent power control for spectrum sharing in cognitive radios: A deep reinforcement learning approach, *IEEE Access*, vol. 6, pp. 25463–25473, 2018.

[29] Y. Du, F. Zhang, and L. Xue, A kind of joint routing and resource allocation scheme based on prioritized memories-deep Q network for cognitive radio ad hoc networks, *Sensors*, vol. 18, no. 7, p. 2119, 2018.

[30] M. Chu, H. Li, X. Liao, and S. Cui, Reinforcement learning-based multiaccess control and battery prediction with energy harvesting in IoT systems, *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2009–2020, 2019.

[31] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. S. Oh, Semisupervised deep reinforcement learning in support of IoT and smart city services, *IEEE Internet Things J.*, vol. 5, no. 2, pp. 624–635, 2018.

[32] B. Omoniwa, B. Galkin and I. Dusparic, Optimizing energy efficiency in UAV-assisted networks using deep reinforcement learning, *IEEE Wireless Communications Letters*, vol. 11, no. 8, pp. 1590−1594, 2022.

[33] Y. Li, X. Zhao, and H. Liang, Throughput maximization by deep reinforcement learning with energy cooperation for renewable ultradense IoT networks, *IEEE Internet Things J.*, vol. 7, no. 9, pp. 9091–9102, 2020.

[34] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, Deterministic policy gradient algorithms, in *Proc. 31*st *Int. Conf. Int. Conf. Machine Learning*, Beijing, China, 2014, pp. I-387–I-395.

[35] K. Li, Y. Emami, W. Ni, E. Tovar, and Z. Han, Onboard deep deterministic policy gradients for online flight resource allocation of UAVs, *IEEE Netw. Lett.*, vol. 2, no. 3, pp. 106–110, 2020.

[36] H. Peng and X. S. Shen, DDPG-based resource management for MEC/UAV-assisted vehicular networks, in *Proc. 2020 IEEE 92*nd *Vehicular Technology Conf.* (*VTC2020-Fall*), Victoria, Canada, 2020, pp. 1–6.

[37] K. K. Nguyen, S. R. Khosravirad, D. B. Da 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. Top. Signal Process.*, vol. 16, no. 3, pp. 358–368, 2022.

[38] J. Cui, Y. Liu, and A. Nallanathan, Multi-agent reinforcement learning-based resource allocation for UAV networks, *IEEE Trans. Wirel. Commun.*, vol. 19, no. 2, pp. 729–743, 2020.

[39] M. Chen, W. Saad, and C. Yin, Liquid state machine learning for resource allocation in a network of cache-enabled LTE-U UAVs, in *Proc. GLOBECOM 2017-2017 IEEE Global Communications Conf.*, Singapore, 2017, pp. 1–6.

[40] A. M. Seid, A. Erbad, H. N. Abishu, A. Albaseer, M. Abdallah, and M. Guizani, Multiagent federated reinforcement learning for resource allocation in UAV-enabled internet of medical things networks, *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19695–19711, 2023.

[41] J. Li, C. Zhou, J. Liu, M. Sheng, N. Zhao, and Y. Su, Reinforcement learning-based resource allocation for coverage continuity in high dynamic UAV communication networks, *IEEE Trans. Wirel. Commun.*, vol. 23, no. 2, pp. 848–860, 2024.

[42] Y. Du, K. Wang, K. Yang, and G. Zhang, Energy-efficient resource allocation in UAV based MEC system for IoT devices, in *Proc. 2018 IEEE Global Communications Conf.* (*GLOBECOM*), Abu Dhabi, United Arab Emirates, 2018, pp. 1–6.

[43] H. Peng and X. Shen, Multi-agent reinforcement learning based resource management in MEC- and UAV-assisted vehicular networks, *IEEE J. Sel. Areas Commun.*, vol. 39, no. 1, pp. 131–141, 2021.

[44] Z. Wang, H. Rong, H. Jiang, Z. Xiao, and F. Zeng, A load-balanced and energy-efficient navigation scheme for UAV-mounted mobile edge computing, *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 5, pp. 3659–3674, 2022.

[45] 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, 2021.

[46] H. Ju and B. Shim, Energy-efficient multi-UAV network using multi-agent deep reinforcement learning, in *Proc. 2022 IEEE VTS Asia Pacific Wireless Communications Symp.* (*APWCS*), Seoul, Republic of Korea, 2022, pp. 70–74.

[47] T. Liu, T. Zhang, J. Loo, and Y. Wang, Deep reinforcement learning-based resource allocation for UAV-enabled federated edge learning, *J. Commun. Inf. Netw.*, vol. 8, no. 1, pp. 1–12, 2023.

Shabeer Ahmad received the BEng degree in telecommunication from Iqra National University, Pakistan, and the MEng degree in electronics and communication engineering from Beijing University of Posts and Telecommunications, China. He is currently a PhD candidate in electronic science and technology at Beijing University of Posts and Telecommunications, Beijing, China. His research interests include the application of machine learning in wireless communication, resource allocation in UAV-enabled wireless networks, software-defined networking, cloud computing, and vehicular ad-hoc network.

Ali Nauman received the MEng degree in wireless communications from Institute of Space Technology, Pakistan in 2016, and the PhD degree in information and communication engineering from Yeungnam University, Republic of Korea in 2022. Currently, he is working as an assistant professor at Department of Information and Communication Engineering, Yeungnam University, Republic of Korea. He has contributed to five patents, authored/co-authored 3 book chapters, and more than 20 technical articles in leading journals and peer-reviewed conferences. The main domain of his research is in the field of artificial intelligence-enabled wireless networks for tactile healthcare, multimedia, and industry 5.0. His research interests include resource allocation for 5G and Beyond-5G (B5G) networks, Device-to-Device communication (D2D), Internet-of-Everything (IoE), URLLC, Tactile Internet (TI), and Artificial Intelligence (AI).

Jinling Zhang received the BS and MS degrees in semiconductor physics and theoretical physics from Inner Mongolia University, Hohhot, China in 1990 and 1993, respectively, and the PhD degree in electromagnetic field and microwave technology from Beijing University of Posts and Telecommunications, Beijing, China in 2010. From 1993 to 2001, she was a lecturer at Physics Department, Inner Mongolia University, China. From 2002 to 2014, she was an assistant professor, and since 2014 she has been a professor at School of Electronic Engineering, Beijing University of Posts and Telecommunications, China. Her research interests include microwave millimeter wave and terahertz communication devices and systems, microwave power transmission, electromagnetic compatibility, information mining, and intelligent information system design.

Adil Khan received the BEng degree in telecommunication from Iqra National University, Pakistan, the MEng degree in electronics and communication engineering from Beijing University of Posts and Telecommunications, Beijing, China, and the PhD degree in electronic science and technology from Beijing University of Posts and Telecommunication, China. His research interests include the application of machine learning in wireless communication, performance analysis of UAV-enabled wireless networks, mobile edge computing, and vehicular ad-hoc networks.

Khizar Abbas received the BEng degree in software engineering from the Government College University Faisalabad (GCUF), Pakistan in 2014, the MEng degree in computer science from the University of Agriculture Faisalabad, Pakistan in 2017, and the PhD degree in computer engineering from Jeju National University, Republic of Korea in 2022. He is currently a research assistant professor at System Security Lab, Department of Computer Science, Hanyang University, Seoul, Republic of Korea. Before joining Hanyang University, He worked as a postdoctoral researcher at Distributed Processing and Network Management Laboratory, Department of Computer Science and Engineering, POSTECH, Republic of Korea. Prior to this, He worked as a visiting lecturer at Department of Computer Science, Government College University Faisalabad, Pakistan. His research interests include software-defined networks, B5G, network slicing, network function virtualization, mobile edge computing, network orchestration and management, network security, blockchain for networks, artificial intelligence for B5G networks, machine learning, and reinforcement learning.

Babar Hayat received the BEng degree in telecommunication from Iqra National University, Pakistan, and the MEng degree in electronics and communication engineering from Beijing University of Posts and Telecommunications, China. He is currently a PhD candidate in electronic science and technology at Beijing University of Posts and Telecommunications, Beijing, China. His research interests include wireless communication and metasurface-based antenna design and analysis of metamaterial and metasurface for antenna applications, such as polarization conversion, phased array antennas, ultrawideband antennas, transmit arrays, and reflect array antennas.