Received 9 March 2024; accepted 29 March 2024. Date of publication 9 April 2024; date of current version 24 April 2024.

Digital Object Identifier 10.1109/OJCOMS.2024.3386595

DDPG-Based Optimization for Zero-Forcing Transmission in UAV-Relay Massive MIMO Networks

MAI T. P. LE [1], VIEN NGUYEN-DUY-NHAT [1], HIEU V. NGUYEN [1], AND OH-SOON SHIN [2] (Member, IEEE)

[1]Faculty of Electronics and Telecommunication Engineering, University of Science and Technology, The University of Danang, Da Nang 50000, Vietnam

[2]School of Electronic Engineering, Soongsil University, Seoul 06978, South Korea

CORRESPONDING AUTHOR: M. T. P. LE (e-mail: lpmai@dut.udn.vn)

This work was supported in part by the Ministry of Education and Training under Project B2024.DNA.19, and in part by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government, Ministry of Science and ICT (MSIT) under Grant RS-2023-00208995.

ABSTRACT This study explores the advantages of employing an unmanned aerial vehicle (UAV) in a massive multiple-input multiple-output (MIMO) network with zero-forcing processing at the base station (BS). Considering potential inaccuracies in channel estimation, we derive a closed-form expression for lower bounds on spectral efficiency in the massive MIMO system, utilizing the UAV as an aerial relay. Subsequently, we formulate a comprehensive optimization problem that encompasses UAV placement and user power allocation in the downlink network, aiming to maximize the data rate for terrestrial users. To address the optimization problem, we propose a novel deep learning-based algorithm that jointly optimizes UAV positioning and power allocation. Finally, we present numerical results that not only validate our theoretical framework and but also demonstrates the effectiveness of the proposed approach.

INDEX TERMS DDPG, massive MIMO, spectral efficiency analysis, unmanned aerial vehicles, Wishart matrices, zero-forcing.

I. INTRODUCTION

THE COMMERCIALIZATION of fifth-generation (5G) communications and the anticipated emergence of sixth-generation (6G) have motivated considerable research efforts towards integrating Unmanned Aerial Vehicles (UAVs), commonly known as drones, into wireless communication networks [1], [2]. This integration enables reliable UAV command and control, as well as communications for mission-related payloads [3]. In this context, UAVs, with assigned missions, can seamlessly connect to wireless networks, functioning as strategic aerial access points (APs), base stations (BSs), or relays. This innovative approach, known as UAV-assisted communications, leverages UAVs to enhance the performance of terrestrial wireless communications from an aerial vantage point [4]. Essentially, UAVs play a transformative role by serving as aerial entities that not only fulfill their designated missions but also enhance and diversify communication capabilities through integration with cellular networks [5].

On the other hand, massive multiple-input multiple-output (MIMO) technology has emerged as a critical component in current 5G networks, offering improved spectral efficiency (SE) and energy efficiency (EE) by employing adaptive beamforming and spatial multiplexing techniques [6]. A massive MIMO base station (BS) is characterized by the presence of hundreds of individually controllable antenna-integrated radios, enabling the simultaneous servicing of numerous user equipments (UEs) on the same time-frequency resource [7]. This technology effectively addresses the challenges in 5G systems by efficiently handling massive data traffic and accommodating a large number of UEs. Leading industry players such as Ericsson, Nokia AirScale, and Huawei have already begun commercializing massive MIMO technology, with deployments featuring 64-antenna configurations at the BS [8].

Given the pivotal role of massive MIMO in existing and future wireless networks, considerable research attention has been dedicated to exploring the integration of UAVs

within massive MIMO-based networks, as discussed in the following section.

A. RELATED WORK

The integration of UAVs and massive MIMO technology in recent work can be mainly divided into three categories according to the function of UAV as: flying BS [9], [10], [11], flying UEs [12], [13], or flying relays [7], [14]. A substantial portion of the existing research on UAV-assisted massive MIMO predominantly explores the use of UAVs as either flying BSs or UEs, focusing on single-hop communications.

However, limited attention has been given to the third category involving relaying structures, where UAVs function as aerial relays within a network comprising two links: ground BS massive MIMO to aerial relay and aerial relay to UEs. This study uniquely focuses on zero-forcing processing within a UAV-assisted relay massive MIMO network, employing deep learning-based approaches to enhance system performance. Further details on the related work are presented subsequently.

1) UAV-RELAY MASSIVE MIMO NETWORK

The deployment of large MIMO arrays in UAV-assisted communication for 6G systems presents several challenges due to the substantial load and size of these systems [15]. These challenges primarily revolve around addressing the limited power resources and stringent constraints on energy consumption, making ground base station (BS) deployment more practical in practice. However, there are notable examples, such as [5], [7], [14], [16], [17], [18], [19], where UAVs are utilized as platforms for aerial intelligent reflecting surfaces (IRS) [5], [16] or as aerial-mounted relays [14], [17], [18], [19]. Flying relays, in comparison to conventional terrestrial relays, offer significant benefits like 360-degree coverage, three-dimensional mobility, and on-demand deployment capabilities [1]. Additionally, UAV relays can navigate above obstacles or affected regions, establishing line-of-sight (LoS) connections with ground UEs and BSs [14]. This inherent characteristic enhances their appeal for delivering high transmission rates and ensuring reliable wireless connectivity [20].

It is worth to note that the aforementioned investigations focus primarily on UAV-assisted massive MIMO systems operating in the millimeter-wave (mmWave) frequency range. The short wavelength of mmWave signals enables the packing of numerous antenna elements or IRS elements into a limited physical area on UAV-mounted relays. This factor has contributed to the growing interest in UAV-assisted communication.

In [14], the optimization of UE association in a UAV relay-assisted mmWave massive MIMO system was explored, where a hybrid beamforming was proposed to mitigate inter-user interference. Hybrid beamforming is particularly advantageous in mmWave UAV-assisted communication, as it utilizes a reduced number of radio frequency (RF) chains to achieve low-dimensional digital beamformers.

Furthermore, [18] investigated the problem of maximizing the total achievable rate in a mmWave UAV-assisted multi-user massive MIMO system. The study jointly considered hybrid beamforming, UAV relay positioning, and power allocation. In [19], a solution was provided for the UAV-assisted hybrid precoding problem, aiming to achieve performance comparable to fully digital beamforming benchmark schemes in terms of sum-mean squared error. Although mmWave communication can effectively meet the high-throughput and low-latency requirements of various UAV application scenarios, it is more suitable for short-range connections in dense urban areas or specific hotspot locations. This is due to its vulnerability to signal attenuation and blockage by obstacles. Consequently, a significant research gap exists in the exploration the UAV-assisted massive MIMO model within the current sub-6GHz range.

2) AI-BASED APPROACHES FOR UAV-ASSISTED NETWORK

The optimization of UAV-assisted wireless massive MIMO networks poses significant challenges due to the mobility of UAVs, the unpredictable wireless medium, and the real-time decision-making requirements. Traditional optimization methods struggle to address these challenges effectively, leading to the emergence of artificial intelligence (AI) as a potential solution. In recent research, various AI-based approaches, including deep learning [21], reinforcement learning (RL) [22], and deep reinforcement learning (DRL) [23], [24], have been proposed to enhance the performance of UAV-assisted communication.

DRL algorithms, in particular, are well-suited for UAV-based networks with imperfect channel state information (CSI). These algorithms enable UAVs to acquire channel knowledge through iterative interactions and adapt their action strategies accordingly. For example, in [25], a DRL-based algorithm was proposed to optimize UAV altitude, aiming to enhance the average Quality of Service (QoS) for the UAV acting as an aerial UE. In [24], an extended deep deterministic policy gradient (DDPG) algorithm was employed to solve the joint optimization problem of maximizing sum data rate and harvested energy while minimizing the UAV's energy consumption in a UAV-based wireless powered IoT network.

In the context of UAVs integrated with massive MIMO, the authors in [26] utilized a Deep Q-Network (DQN) to optimize UAV navigation, where UAVs acted as aerial UEs. The DQN was used to select an optimal policy. Recent work [18] has demonstrated significant performance enhancements through DRL-based approaches for UAV-relay massive MIMO scenarios. However, it is important to note that the exploration of DRL-based wireless communications is still in its early stages, and further research is warranted to fully realize its potential.

B. MOTIVATION AND CONTRIBUTION

Most of related work on UAV-relay massive MIMO network has placed excessive emphasis on mmWave range, as noted in [5], [7], [14], [17], [18], [19]. Although some studies have explored AI-based solutions, their scope has been confined to single-hop communication scenarios, which involve a direct link between either a flying massive MIMO base station and ground UEs or a ground massive MIMO base station and flying UEs. However, due to UE mobility and complex nature of the urban environment with dense high-rise buildings and other obstacles, blockage emerges as a significant drawback of massive MIMO. This issue results in the obstruction of LoS transmission between ground BS and UEs [27]. In contrast, our work focuses on a UAV-assisted massive MIMO system where the UAV serves as an aerial relay to enhance connectivity in scenarios where the direct transmission from the BS to UEs is obstructed. We assume imperfect channel knowledge, and direct links between the BS and UEs are considered unavailable. Prior investigations [28], [29], [30] have analyzed the performance of the low-complexity maximal ratio transmission (MRT) scheme in UAV-relay massive MIMO networks. However, to the best of our knowledge, no prior research has explored the application of favorable zero-forcing (ZF) processing in the specific system under consideration [30]. This research gap can be attributed, in part, to the inherent challenges associated with handling products of Wishart matrices and the complexity of signal processing in a two-hop communication scenario, especially when incorporating channel estimation in a network with imperfect CSI.

This study aims to address this research gap by investigating the fundamental limits of the ZF technique in a UAV-enhanced massive MIMO network. Furthermore, to mitigate the effects of channel estimation errors, we employ the deep deterministic policy gradient (DDPG) algorithm to jointly optimize UAV positioning and maximize the system sum rate. Our contribution lies in the specialized application of DDPG to optimize ZF in a UAV-relay network, which presents unique challenges such as high mobility, dynamic environments, and real-time decision-making. We believe that this focused approach fills a significant gap in the literature, as it addresses a critical aspect of UAV network design that is increasingly relevant in the era of 5G and beyond.

Through extensive simulations, we demonstrate that our proposed DDPG-based algorithm significantly enhances the system sum rate even in the presence of imperfect CSI. Notably, our approach outperforms conventional convex optimization methods, highlighting its effectiveness in addressing challenges associated with imperfect CSI. These findings open up promising avenues for optimizing UAV-enhanced massive MIMO networks. In summary, we emphasize the following key advancements in our research:

- First, we provide a downlink performance analysis of a UAV-relay massive MIMO system under the assumption of imperfect CSI, adopting the minimum mean square error (MMSE) channel estimate. Closed-form expressions for the SE lower bounds are derived, specifically for the ZF precoding scheme. Notably, these expressions mark the first exploration of ZF within the proposed UAV-relay massive MIMO network with imperfect CSI.
- To maximize the downlink sum rate, we formulate a joint optimization involving power allocation and UAV placement. This optimization can be solved by using a sequence of convex problems or sub-problems. While traditional approaches typically employ algorithms with polynomial computational complexity, it is important to note that the formulated problem is specifically tailored for a UAV-based system operating in a dynamic environment, requiring adaptive behavior over time. Recognizing the effectiveness of deep learning in handling continuous action spaces, we propose a low-complexity algorithm that leverages the DDPG methodology. By integrating ZF beamforming with the DDPG-based algorithm, we aim to address various challenges, including high mobility, dynamic environments, and the need for real-time decision-making.
- To validate the accuracy of the derived SE performance, we conduct a comprehensive set of Monte Carlo simulations for ZF processing in the downlink data transmission. We compare these results with those obtained from analytical analysis. Additionally, we present analytical and simulation-based performance evaluations of MRT processing for the purpose of comparison. Both scenarios, considering perfect and imperfect CSI, are investigated for both ZF and MRT precoding schemes.
- Finally, we demonstrate the effectiveness of the proposed DDPG-based algorithm, aligning with the outcomes of our theoretical analysis. By analyzing the training results, we illustrate that the optimal policy derived from the DDPG algorithm surpasses the flexibility offered by traditional rule-based policies. The algorithm effectively explores various locations and identifies optimal positions, leading to the maximization of the system sum rate. The results indicate a significant improvement in the system sum rate, particularly in scenarios where perfect CSI is unavailable. Notably, the proposed DDPG algorithm, as validated by simulation results, achieves this enhancement while catering to a relatively large number of UEs and ground BS antennas, while still maintaining a moderate number of UAV-relay antennas.

C. ORGANIZATION

The rest of this paper is organized as follows. Section II introduces the system model, providing a foundation for the ensuing discussion. A comprehensive analysis of downlink spectral efficiency for ZF transmission is presented in Section III. Section IV introduces the DDPG-based

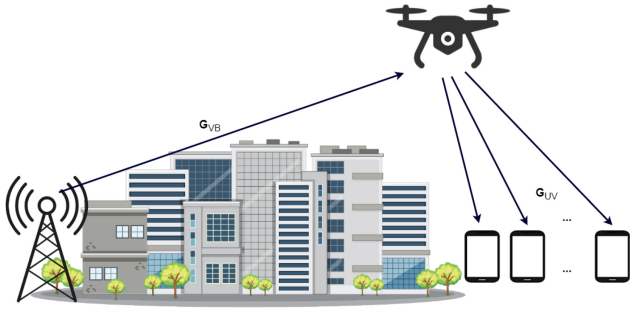

FIGURE 1. The proposed UAV-relay massive MIMO system model.

algorithm, addressing the joint optimization of power allocation and UAV location while maximizing the system sum rate. In Section V, we showcase numerical results that substantiate the validity of our theoretical findings, encompassing the potential gains from the proposed algorithm. Finally, conclusions are drawn in Section VI.

Notation: Bold lowercase letters denote column vectors, while bold uppercase letters represent matrices. The expectation, absolute value, and Euclidean norm are represented by $\mathbb{E}\{\cdot\}$, $|\cdot|$, and $\|\cdot\|$, respectively. The Hermitian transpose of $\mathbf{x}$ is denoted by $\mathbf{x}^{\mathrm{H}}$, and the trace of $\mathbf{x}$ is represented as $\mathrm{tr}(\mathbf{x})$. The variance operator is denoted as $\mathrm{Var}(\cdot)$. The notation $x \sim \mathcal{CN}(0, 1)$ indicates that the variable x follows a circularly symmetric complex Gaussian distribution with zero mean and unit variance.

II. SYSTEM MODEL

As illustrated in Fig. 1, we consider a downlink UAV-assisted massive MIMO network, where the BS is equipped with massive MIMO array comprising N antennas and serves K single-antenna UEs. We assume that direct transmissions between the ground BS and UEs are unavailable. This is due to the challenging characteristics of the urban environment, such as signal blockages and severe multi-path fading, which makes it challenging for UEs to establish reliable LoS links with the ground BS [30]. Therefore, a UAV with N_r antennas serves as an aerial relay, augmenting the connection between obstructed links from the BS to UEs and facilitating high-speed data transfer. In this network, a UAV with N_r antennas functions as an aerial relay, enhancing the connection between obstructed links from the BS to UEs and enabling high-speed data transfer. Global Positioning System (GPS) coordinates precisely define the locations of network nodes, represented as $\mathbf{c}_{\mathtt{X}} \triangleq [x_{\mathtt{X}}\ y_{\mathtt{X}}\ z_{\mathtt{X}}] \in \mathbb{R}^3$. Here, $\mathtt{X} \in \mathcal{N} \triangleq \{\mathtt{B}, \mathtt{V}, \{\mathtt{U}_k\}_{k\in\mathcal{K}}\}$ denote the BS, the UAV, and the UE k, where $k \in \mathcal{K} = \{1, 2, \ldots, K\}$, respectively. The LoS distance in three-dimensional space between any two nodes is calculated as $d(\mathbf{c}_{\mathtt{A}}, \mathbf{c}_{\mathtt{B}}) \triangleq \|\mathbf{c}_{\mathtt{A}} - \mathbf{c}_{\mathtt{B}}\|$, where $\mathtt{A}$, $\mathtt{B}$ are elements of the set $\mathcal{N}$. Consequently, their horizontal ground distance is expressed as $d_{\mathtt{g}}(\mathbf{c}_{\mathtt{A}}, \mathbf{c}_{\mathtt{B}}) \triangleq \sqrt{(x_{\mathtt{A}} - x_{\mathtt{B}})^2 + (y_{\mathtt{A}} - y_{\mathtt{B}})^2}$. It is important to note that we assume the BS has access to UE locations, obtained through data mining processes involving social networks, such as the Twitter API [31]. Furthermore, it is crucial to emphasize that the UAV operates within a specified range of altitudes, specifically, $Z^{\min} \le z_{\mathtt{V}} \le Z^{\max}$.

A. CHANNEL MODEL

In our system model, we consider the presence of both large-scale and small-scale fading in all channels. The channel between the BS and the UAV can be expressed as an $N \times N_r$ matrix $\mathbf{G}_{\mathrm{VB}} = \mathbf{H}_{\mathrm{VB}}\mathbf{D}_{\mathrm{VB}}^{1/2}$, where $\mathbf{H}_{\mathrm{VB}} \in \mathbb{C}^{N\times N_r}$ represents the small-scale fading matrix. The elements of $\mathbf{H}_{\mathrm{VB}}$ follow independent and identically distributed (i.i.d.) complex normal $\mathcal{CN}(0, 1)$ random variables. The large-scale fading effect between the BS and the UAV is accounted for by the diagonal matrix $\mathbf{D}_{\mathrm{VB}} = \beta_{\mathrm{VB}}\mathbf{I}_{N_r}$, where β_{VB} represents this fading effect. Similarly, the channel between the UAV and each UE can be represented by an $N_r \times K$ matrix $\mathbf{G}_{\mathrm{UV}} = \mathbf{H}_{\mathrm{UV}}\mathbf{D}_{\mathrm{UV}}^{1/2}$. Here, $\mathbf{H}_{\mathrm{UV}} \in \mathbb{C}^{N_r\times K}$ represents the small-scale fading matrix, with elements following i.i.d. complex normal $\mathcal{CN}(0, 1)$ random variables. The large-scale fading effect between the UAV and the UEs is accounted for by the diagonal matrix $\mathbf{D}_{\mathrm{UV}} \in \mathbb{C}^{K\times K}$, where the k-th entry is denoted as $\beta_{\mathrm{UV},k}$.

The air-to-air channel of the BS-UAV link is characterized by LoS transmission, with minimal small-scale fading and non-line-of-sight (NLoS) components. This channel is described by the free-space path loss model [32], where β_{VB} can be expressed as [33]

$$\beta_{\mathrm{VB}} = \beta_0 d(\mathbf{c}_{\mathtt{B}}, \mathbf{c}_{\mathtt{V}})^{-2}, \tag{1}$$

with β_0 representing the channel's power gain at a reference distance.

The connectivity between UAVs and ground-based UEs is influenced by various urban environmental factors. In our analysis, we incorporate the large-scale fading expression derived from [34] to capture these effects. This expression takes into account several elements, including the path loss denoted as $\mathtt{PL}_k^0$, where the path loss exponent is represented by α_k. Moreover, the expression considers the carrier frequency $f_{\mathtt{c}}$ and the speed of light c, along with supplementary factors for LoS and NLoS links, denoted as $\mu^k{}_{\mathtt{LoS}}$ and $\mu^k{}_{\mathtt{NLoS}}$, respectively.

In mathematical term, the large-scale fading expression for the link between UAV and UE k is defined as

$$\beta_{\mathrm{UV},k} = \mathtt{PL}_k^0 + \mu^k_{\mathtt{LoS}} P^k_{\mathtt{LoS}} + \mu^k_{\mathtt{NLoS}} P^k_{\mathtt{NLoS}}, \tag{2}$$

where the distance-related path loss $\mathtt{PL}_k^0$ is defined as

$$\mathtt{PL}_k^0 = 10\alpha_k \log_{10}\left(\frac{4\pi f_{\mathtt{c}} d(\mathbf{c}_{\mathtt{U}_k}, \mathbf{c}_{\mathtt{V}})}{c}\right).$$

Here, $P^k_{\mathtt{NLoS}}$ represents the probability of NLoS, whereas its complementary probability $P^k_{\mathtt{LoS}} = 1 - P^k_{\mathtt{NLoS}}$ is determined according to [35]:

$$P^k_{\mathtt{LoS}} = \frac{1}{1 + a\exp\left[-b\left(\arctan\left(\frac{z_{\mathtt{V}}}{d(\mathbf{c}_{\mathtt{U}_k}, \mathbf{c}_{\mathtt{V}})}\right) - a\right)\right]},$$

where the constants a and b are parameters characterizing the environment.

B. UPLINK TRAINING PHASE

We adopt a practical assumption in our system model, where the UAV-relay does not possess knowledge of the small-scale fading channels. Instead, it leverages information about the large-scale fading to transmit an amplified version of the received signal to the destination. This approach restricts the channel estimation process to be performed solely at the destination. To implement this assumption, a two-hop channel estimation procedure is required, involving training for both the UEs-UAV link and the UAV-BS link. We assume that the link between the UEs and the BS through the UAV-relay is synchronized, with negligible processing time. Let τ represent the number of symbols per coherence block (CB), where τ_p pilot symbols are allocated for the training procedure ($\tau_p < \tau$). We denote the $K \times \tau_{\mathrm{p}}$ matrix $\Phi \triangleq [\boldsymbol{\phi}_1^{\mathrm{T}} \boldsymbol{\phi}_2^{\mathrm{T}} \dots \boldsymbol{\phi}_K^{\mathrm{T}}]^{\mathrm{T}}$ as the pilot matrix allocated to the K UEs. It is important to note that all pilot sequences within Φ are orthogonal to each other, ensuring that $\Phi\Phi^{\mathrm{H}} = \mathbf{I}_K$.

For the link between UEs and UAV-relay: During the training phase, the UAV-relay receives a $N_r \times \tau_{\mathrm{p}}$ matrix which is denoted by $\mathbf{Y}_{\mathrm{Vp}}$:

$$\mathbf{Y}_{\mathrm{Vp}} = \sqrt{P_{\mathrm{Up}}}\mathbf{G}_{\mathrm{UV}}\Phi + \mathbf{N}_{\mathrm{Vp}}, \tag{3}$$

where P_{Up} represents the pilot transmit power from each UE and the AWGN noise $\mathbf{N}_{\mathrm{Vp}} \in \mathbb{C}^{N_r \times \tau_{\mathrm{p}}}$ has i.i.d $\mathcal{CN}(0, 1)$ elements.

For the link between UAV-relay and BS: In this scenario, the UAV serves as an amplify-and-forward aerial relay. Initially, the pilot signals undergo amplification at the UAV, achieved by a normalization factor α_{p}, before being transmitted to the BS. The received $N \times \tau_{\mathrm{p}}$ pilot signals at the BS are mathematically expressed as:

$$\begin{aligned}\mathbf{Y}_{\mathrm{p}} &= \sqrt{\alpha_{\mathrm{p}}}(\mathbf{G}_{\mathrm{VB}}\mathbf{Y}_{\mathrm{Vp}}) + \mathbf{N}_{\mathrm{Bp}},\\ &= \sqrt{\alpha_{\mathrm{p}}}\mathbf{G}_{\mathrm{VB}}(\sqrt{P_{\mathrm{Up}}}\mathbf{G}_{\mathrm{UV}}\Phi + \mathbf{N}_{\mathrm{Vp}}) + \mathbf{N}_{\mathrm{Bp}},\\ &= \sqrt{\alpha_{\mathrm{p}}P_{\mathrm{Up}}}\mathbf{G}\Phi + \mathbf{N}_{\mathrm{p}}.\end{aligned} \tag{4}$$

Here, $\mathbf{G} \triangleq \mathbf{G}_{\mathrm{VB}}\mathbf{G}_{\mathrm{UV}}$ represents the equivalent channel matrix. The resulting noise matrix is given by $\mathbf{N}_{\mathrm{p}} = \sqrt{\alpha_{\mathrm{p}}}\mathbf{G}_{\mathrm{VB}}\mathbf{N}_{\mathrm{Vp}} + \mathbf{N}_{\mathrm{Bp}}$, where the AWGN matrix $\mathbf{N}_{\mathrm{Bp}} \in \mathbb{C}^{N \times \tau_{\mathrm{p}}}$ consists of i.i.d $\mathcal{CN}(0, 1)$ entries.

Denote $\mathbf{g}_k \in \mathbb{C}^{N \times 1}$ as the channel vector of UE k, then $\mathbf{g}_k$ can be extracted as a column vector from the equivalent channel matrix $\mathbf{G}$. Building upon the derivations in [30], the minimum mean-squared error (MMSE) channel estimate matrix at the massive MIMO BS is provided as

$$\hat{\mathbf{G}} = \frac{\nu_1}{\nu_2\sqrt{\alpha_p P_{\mathrm{Up}}}}\mathbf{Y}_{\mathrm{p}}\tilde{\mathbf{D}}\Phi^{\mathrm{H}}\mathbf{D}_{\mathrm{UV}}, \tag{5}$$

where $\tilde{\mathbf{D}} = (\mathbf{I}_{\tau_{\mathrm{p}}} + \frac{\nu_1}{\nu_2}\Phi\mathbf{D}_{\mathrm{UV}}\Phi^{\mathrm{H}})^{-1}$, $\nu_1 \triangleq NN_r\beta_{\mathrm{VB}}$, and $\nu_2 \triangleq \frac{N(\alpha_p N_r\beta_{\mathrm{VB}}+1)}{\alpha_{\mathrm{p}}P_{\mathrm{Up}}}$.

Remark 1: From (5), the channel estimate $\hat{\mathbf{g}}_k \in \mathbb{C}^{N \times 1}$ for UE k is given as

$$\hat{\mathbf{g}}_k = \nu_k\mathbf{Y}_{\mathrm{p}}\boldsymbol{\phi}_k, \tag{6}$$

where

$$\nu_k = \frac{\sqrt{\alpha_{\mathrm{p}}P_{\mathrm{Up}}}N_r\beta_{\mathrm{VB}}\beta_{\mathrm{UV},k}}{\alpha_{\mathrm{p}}N_r\beta_{\mathrm{VB}}(P_{\mathrm{Up}}\tau_{\mathrm{p}}\beta_{\mathrm{UV},k} + 1) + 1}. \tag{7}$$

According to the orthogonal property of the MMSE estimator [36], the two vectors $\hat{\mathbf{g}}_k$ and $\tilde{\mathbf{g}}_k$ are mutually uncorrelated. Here, $\tilde{\mathbf{g}}_k$ represents the channel estimation error, defined as $\tilde{\mathbf{g}}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k$. It is important to note that $\mathbf{g}_k$ and $\tilde{\mathbf{g}}_k$ are uncorrelated but not independent. This occurs because the two-hop vector channel $\mathbf{g}_k$ is derived by multiplying a Gaussian matrix and a Gaussian vector, and as a result, it does not exhibit the characteristics of a Gaussian vector. This leads to

$$\mathbb{E}\{\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\} = N_r\beta_{\mathrm{VB}}\beta_{\mathrm{UV},k}\mathbf{I}_N. \tag{8}$$

Meanwhile, the variance of the UE k channel estimate is obtained as

$$\mathbb{E}\{\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\mathrm{H}}\} = \nu_k^2\kappa_k\,\mathbf{I}_N, \tag{9}$$

resulting in

$$\begin{aligned}\mathbb{E}\{\tilde{\mathbf{g}}_k\tilde{\mathbf{g}}_k^{\mathrm{H}}\} &= \mathbb{E}\{\mathbf{g}_k\mathbf{g}_k^{\mathrm{H}}\} - \mathbb{E}\{\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\mathrm{H}}\}\\ &= (N_r\beta_{\mathrm{VB}}\beta_{\mathrm{UV},k} - \nu_k^2\kappa_k)\mathbf{I}_N.\\ &= (\beta_k - \eta_k)\mathbf{I}_N,\end{aligned} \tag{10}$$

where

$$\begin{aligned}\beta_k &\triangleq N_r\beta_{\mathrm{VB}}\beta_{\mathrm{UV},k},\\ \eta_k &\triangleq \nu_k^2\kappa_k,\\ \kappa_k &= \tau_{\mathrm{p}}^2\alpha_{\mathrm{p}}P_{\mathrm{Up}}N_r\beta_{\mathrm{VB}}\beta_{\mathrm{UV},k} + \tau_{\mathrm{p}}(\alpha_{\mathrm{p}}N_r\beta_{\mathrm{VB}} + 1).\end{aligned} \tag{11}$$

The detailed proofs are referred to [30, Appendix B].

C. DOWNLINK DATA TRANSMISSION PHASE

Similar to the uplink training phase, the downlink data transmission also comprises two hops, corresponding to the BS-UAV link and the subsequent UAV-UEs link.

First, the transmitted signal from the BS arrives at the UAV as

$$\mathbf{y}_{\mathrm{Vd}} = \sqrt{P_0}\mathbf{G}_{\mathrm{VB}}^{\mathrm{H}}\mathbf{W}\mathbf{x} + \mathbf{n}_{\mathrm{Vd}}, \tag{12}$$

where the $N \times K$ matrix $\mathbf{W}$ is the designed precoding matrix, P_0 represents the transmit power of the BS, and the signal transmitted, denoted as $\mathbf{x} \in \mathbb{C}^{K \times 1}$, satisfies the condition $\mathbb{E}\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$. Additionally, the received signal includes Gaussian noise, characterized by $\mathbf{n}_{\mathrm{Vd}} \in \mathbb{C}^{N_r \times 1}$, with each of its entries following i.i.d. complex normal $\mathcal{CN}(0, 1)$ random variables.

After amplification at the UAV-relay by a factor of α_{d}, the signals are delivered to the KUEs as

$$\mathbf{y}_{\mathrm{d}} = \mathbf{G}_{\mathrm{UV}}^{\mathrm{H}}\mathbf{r}_{\mathrm{Vd}} + \mathbf{n}_{\mathrm{U}} = \sqrt{\alpha_{\mathrm{d}}P_0}\mathbf{G}^{\mathrm{H}}\mathbf{W}\mathbf{x} + \tilde{\mathbf{n}}_{\mathrm{U}}. \tag{13}$$

Here, $\mathbf{n}_{\mathrm{U}} \in \mathbb{C}^{K \times 1}$ represents the AWGN with elements being i.i.d $\mathcal{CN}(0, 1)$ random variables, resulting in $\tilde{\mathbf{n}}_{\mathrm{U}} = \sqrt{\alpha}_{\mathrm{d}}(\mathbf{G}_{\mathrm{UV}}^{\mathrm{H}})\mathbf{n}_{\mathrm{Vd}} + \mathbf{n}_{\mathrm{U}}$. To meet the UAV power

constraint $\mathbb{E}\{\|\mathbf{r}_{\mathrm{Vd}}\|^2\} \leq P_{\mathrm{Vd}}$, the amplification factor needs to satisfy the following condition [30]

$$\alpha_{\mathrm{d}} \leq \frac{P_{\mathrm{Vd}}}{N_r(P_0\beta_{\mathrm{VB}}\mathrm{Tr}(\mathbf{W}\mathbf{W}^{\mathrm{H}})+1)}. \tag{14}$$

III. PERFORMANCE ANALYSIS AND PROBLEM FORMULATION

A. DOWNLINK SPECTRAL EFFICIENCY WITH ZERO-FORCING TRANSMISSION

Since the channel consists of two hops, the channel estimate can be decomposed as $\hat{\mathbf{G}} = \mathbf{Z}_1\mathbf{Z}_2\mathbf{D}_g^{1/2}$, where elements of $\mathbf{Z}_1 \in \mathbb{C}^{N\times N_r}$ and $\mathbf{Z}_2 \in \mathbb{C}^{N_r\times K}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables, and $\mathbf{D}_g^{1/2} = \mathrm{diag}(\eta_1,\ldots,\eta_K)$ is referred to (11).

Lemma 1: Let[1]

$$\mathbf{A} = \hat{\mathbf{G}}\left(\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}\mathbf{D}_g^{1/2}. \tag{15}$$

Following the approach in [37], the precoding matrix is designed as

$$\mathbf{W}_{\mathrm{ZF}} = \sqrt{c}\mathbf{A}\mathbf{D}_p^{1/2}, \tag{16}$$

where the diagonal elements of $\mathbf{D}_p = \mathrm{diag}(p_1,\ldots,p_K)$ are the signal power of K UEs, and the scaling factor c is designed to ensure that the total transmit power is no greater than unity, i.e., $\mathbb{E}\{||\mathbf{W}_{\mathrm{ZF}}\mathbf{x}||^2\} \leq 1$ [37]. To this end, we obtain the closed-form expression of c as follows:

$$c = \frac{(N-N_r)(N_r-K)}{N_r}. \tag{17}$$

Proof: Please refer to Appendix A for the derivation of (17).

Collectively, the $K\times 1$ received vector in (13) with ZF precoding becomes

$$\begin{aligned}\mathbf{y}_d &= \sqrt{\alpha_{\mathrm{d}}P_0}\mathbf{G}^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x} + \tilde{\mathbf{n}}_{\mathrm{U}}\\ &= \sqrt{\alpha_{\mathrm{d}}P_0}\hat{\mathbf{G}}^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x} + \sqrt{\alpha_{\mathrm{d}}P_0}\tilde{\mathbf{G}}^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x} + \tilde{\mathbf{n}}_{\mathrm{U}}\\ &= \sqrt{c\alpha_{\mathrm{d}}P_0}\mathbf{D}_g^{1/2}\mathbf{D}_p^{1/2}\mathbf{x} + \sqrt{\alpha_{\mathrm{d}}P_0}\tilde{\mathbf{G}}^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x} + \tilde{\mathbf{n}}_{\mathrm{U}}.\end{aligned} \tag{18}$$

The received signal at UE k is expressed as

$$\mathbf{y}_{dk} = \sqrt{c\alpha_{\mathrm{d}}P_0\eta_k p_k} + \sqrt{\alpha_{\mathrm{d}}P_0}\tilde{\mathbf{g}}_k^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x} + \tilde{n}_{\mathrm{Uk}}. \tag{19}$$

Since the variance of $\tilde{n}_{\mathrm{Uk}}$ is $\sigma_k^2 = \alpha_{\mathrm{d}}N_r\beta_{\mathrm{UV},k} + 1$ [30], it is left to compute the variance of the second term in (19). Adopting the approach in [37], using (10) leads to

$$\begin{aligned}\mathrm{Var}\{\tilde{\mathbf{g}}_k^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x}\} &= \mathbb{E}\{\mathbf{x}^H\mathbf{W}_{\mathrm{ZF}}^H\tilde{\mathbf{g}}_k\tilde{\mathbf{g}}_k^H\mathbf{W}_{\mathrm{ZF}}\mathbf{x}\}\\ &= (\beta_k-\eta_k)\mathbb{E}\left\{||\mathbf{W}_{\mathrm{ZF}}\mathbf{x}||^2\right\}\\ &= (\beta_k-\eta_k)\sum_{i=1}^{K}p_i.\end{aligned} \tag{20}$$

[1]Given that an arbitrary $N\times N$ matrix typically requires $\mathcal{O}(N^3)$ floating-point operations (flops), the complexity for matrix inversion of $\hat{\mathbf{G}}^H\hat{\mathbf{G}}$ in (15) would be $\mathcal{O}(K^3)$, reminding that $K \leq N_r$.

Lemma 2: From (19), the signal-to-interference-plus-noise ratio (SINR) of UE k is bounded by the following closed-form expression:

$$\gamma_k(\mathbf{p},\mathbf{c}_{\mathrm{V}}) \triangleq \frac{c\alpha_{\mathrm{d}}P_0p_k\eta_k}{\alpha_{\mathrm{d}}P_0\sum_{i=1}^{K}p_i(\beta_k-\eta_k)+\sigma_k^2}, \tag{21}$$

where $\mathbf{p} = [p_1,\ldots,p_K]$ is the vector of UE power control coefficients. The respective lower bound of the achievable rate of UE k is given by

$$R_k(\mathbf{p},\mathbf{c}_{\mathrm{V}}) = \frac{\tau-2\tau_{\mathrm{p}}}{2\tau}\ln\left(1+\gamma_k(\mathbf{p},\mathbf{c}_{\mathrm{V}})\right), \tag{22}$$

where the utilization of the UAV in a half-duplex amplify-and-forward relaying mode leads to the emergence of the pre-log factor $\frac{\tau-2\tau_{\mathrm{p}}}{2\tau}$.

From (22), the sum rate of all active UEs can be computed as

$$R_\Sigma(\mathbf{p},\mathbf{c}_{\mathrm{V}}) = \sum_{k=1}^{K}R_k(\mathbf{p},\mathbf{c}_{\mathrm{V}}). \tag{23}$$

B. PROBLEM FORMULATION FOR SUM RATE MAXIMIZATION

In the following, we present the conventional problem formulation for maximizing the downlink sum rate R_Σ in (23), taking into account the system contraints.The problem is introduced as follows:

$$\underset{\mathbf{p},\mathbf{c}_{\mathrm{V}}}{\mathtt{maximize}}\ R_\Sigma(\mathbf{p},\mathbf{c}_{\mathrm{V}}), \tag{24a}$$

$$\mathtt{s.t.}\ \sum_{k=1}^{K}p_k \leq 1, \tag{24b}$$

$$(14), \tag{24c}$$

$$Z^{\min} \leq z_{\mathrm{V}} \leq Z^{\max}. \tag{24d}$$

The constraints (24b) and (24c) are imposed to satisfy the signal and the UAV power contraints, respectively, while constraint (24d) ensures that the UAV operates within the permissible altitude range. Here, we assume that at the highest altitude, $Z_{\mathrm{V}}^{\max}$, the UAV's coverage area is sufficient to serve all UEs.

The problem stated in (24) can be tackled using traditional optimization methods, involving the solution of a sequence of convex problems or sub-problems. Such an approach typically employs algorithms characterized by polynomial computational complexity. It is crucial to note that the formulated problem is tailored for a UAV-based system operating in a dynamic environment, requiring adaptive behavior over time. In light of this, the application of deep learning emerges as a promising option. Among the potential candidates, the DDPG-based algorithm, designed for continuous action spaces, is anticipated to enhance system performance. In the following, we will introduce the DDPG algorithm and its application in addressing the formulated problem.

IV. DEEP DETERMINISTIC POLICY GRADIENT (DDPG)-BASED ALGORITHM FOR MAXIMIZING SYSTEM SUM RATE

In this section, we begin with a concise overview of reinforcement learning and DDPG. Subsequently, we introduce the optimization problem, focusing on the maximization of the sum rate. We then employ DDPG to address the proposed problem by jointly optimizing the UAV location and the power allocation for UEs.

A. INTRODUCTION TO REINFORCEMENT LEARNING

Markov decision process (MDP) is formally defined as a quintuple $(\mathcal{S}, \mathcal{A}, \mathcal{R}, p, \gamma)$, where $\mathcal{S}$ represents the state space and A denotes the action space. The reward function, defined over the Cartesian product space $\mathcal{S} \times \mathcal{A}$ and mapping to a subset of real numbers, is denoted as $\mathcal{R} : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{P}(\mathbb{R})$. Additionally, we have the transition probability function $p : \mathcal{S} \times \mathcal{A} \rightarrow \mathcal{P}(\mathcal{S})$, where $\mathcal{P}(\mathcal{S})$ signifies the set encompassing all possible probability measures on $\mathcal{S}$. Lastly, the discount factor, denoted as $0 < \gamma < 1$, represents a critical component in the formulation.

The reward, crucially dependent on the state and action, is denoted as $r^{(t)} = \mathcal{R}(s^{(t)}, a^{(t)}, s^{(t+1)})$ at each time step t. The return is defined as the sum of discounted future rewards:

$$R^{(t)} = \sum_{\tau=1}^{\inf} \gamma^{\tau} r^{(t+\tau+1)}. \tag{25}$$

A policy π defines a probability distribution $\pi(.|s)$ over the action space $\mathcal{A}$ for each state $s \in \mathcal{S}$ $(a^{(t)} \sim \pi(.|s))$. If we let θ denotes the policy parameters (the weights and biases of a neural network), it can be expressed as

$$a^{(t)} \sim \pi_{\theta}(.|s). \tag{26}$$

The goal of reinforcement learning is to discover a policy that maximizes the expected return from the initial distribution $J = \mathbb{E}_{r,s,a}\{R^{(1)}\}$:

$$\pi^* = \arg\max_{\pi} J(\pi), \tag{27}$$

where π^* represents the optimal policy. Given a specific policy π, the value of the state-action function $\mathcal{Q}_{\pi}(s^{(t)}, a^{(t)})$ at the time step t is expressed as:

$$\mathcal{Q}_{\pi}\left(s^{(t)}, a^{(t)}\right) = \mathbb{E}\left[R^{(t)} | s^{(t)} = s, a^{(t)} = a\right]. \tag{28}$$

The $\mathcal{Q}$ function follows the Bellman equation and is expressed as

$$\mathcal{Q}_{\pi}\left(s^{(t)}, a^{(t)}\right) = \mathbb{E}\left[r^{(t+1)} | s^{(t)} = s, a^{(t)} = a\right] + \gamma \sum_{s' \in \mathcal{S}} P^{a}_{ss'} \left(\sum_{a' \in \mathcal{A}} \pi\left(s', a'\right) \mathcal{Q}_{\pi}\left(s', a'\right)\right) \tag{29}$$

where $P^{a}_{ss'} = P_r(s^{(t+1)} = s' | s^{(t)} = s, a^{(t)} = a,)$ represents the transition probability from state s to state s' through the action a.

The Q-learning algorithm seeks to discover the optimal action-value function $\mathcal{Q}^*_{\pi}$ by employing the following strategy:

$$\mathcal{Q}^*_{\pi}\left(s^{(t)}, a^{(t)}\right) = r^{(t+1)}\left(s^{(t)} = s, a^{(t)} = a, \pi = \pi^*\right) + \gamma \sum_{s' \in \mathcal{S}} P^{a}_{ss'} \max_{a' \in \mathcal{A}} Q^*\left(s', a'\right). \tag{30}$$

To derive the optimal $\mathcal{Q}^*(s^{(t)}, a^{(t)})$, we can use a recursive algorithm that does not require knowledge of the exact reward model or the state transition model. The update rule for the $\mathcal{Q}$ function is as follows:

$$\mathcal{Q}^*_{\pi}\left(s^{(t)}, a^{(t)}\right) < -(1-\alpha)\mathcal{Q}^*_{\pi}\left(s^{(t)}, a^{(t)}\right) + \alpha\left(r^{(t+1)} + \gamma \max_{a' \in \mathcal{A}} \mathcal{Q}_{\pi}\left(s^{(t+1)}, a'\right)\right), \tag{31}$$

where α is the learning rate for updating the $\mathcal{Q}$ function.

A significant challenge in using neural networks to approximate the Q function is the high correlation among states over time. This correlation can reduce the diversity and randomness of states, as they all stem from the same episode. To address this challenge, experience replay is employed, introducing a buffer window to store a subset of recent states. This approach significantly enhances the effectiveness of deep reinforcement learning (DRL). Instead of updating the Q-value function solely based on the most recent state, the deep neural network (DNN) is trained using a batch of states randomly selected from the experience replay buffer. This mechanism ensures that DRL can facilitate more stable and diverse learning of the Q-value function.

In the context of DRL, the Q-function is precisely defined with the assistance of a parameter vector denoted as θ, i.e.,

$$\mathcal{Q}\left(s^{(t)}, a^{(t)}\right) \triangleq \mathcal{Q}\left(\theta | s^{(t)}, a^{(t)}\right). \tag{32}$$

Here, θ serves as a repository of acquired knowledge, enabling the DNN to approximate the Q-function for various state-action pairs, thereby enhancing the agent's decision-making capabilities. The value of θ undergoes iterative updates through stochastic optimization algorithms, expressed as

$$\theta^{(t+1)} = \theta^{(t)} - \mu \Delta_{\theta} \mathcal{L}(\theta), \tag{33}$$

where μ and Δ_{θ} denote the learning rate for the update and the gradient of the loss function $\mathcal{L}(\theta)$ with respect to θ, respectively.

B. INTRODUCTION TO DEEP DETERMINISTIC POLICY GRADIENT

The DDPG algorithm is recognized a robust method in the realm of DRL, constituting a subset of MDPs. It excels at tackling challenges presented by continuous action space problems, extending traditional reinforcement learning into domains such as robotics, finance, and control systems. DDPG operates in an off-policy manner within the MDP framework, allowing agents to optimize cumulative rewards

through discrete-time interactions with environments, guided by states, actions, transition probabilities, and rewards [38].

Let $\theta_c^{(target)}$ denote the target critic network. The Bellman update target is then expressed as

$$y = r^{(t)} + \gamma \mathcal{Q}\left(\theta_c^{(target)} | s^{(t+1)}, a'\right). \tag{34}$$

The mean squared error (MSE) quantifies the discrepancy between the Q-values at time step t and the Bellman update target, estimated using the reward $r^{(t)}$ and the Q-values at time step $(t+1)$. This minimization process addresses the one-step temporal difference and is formally expressed by the following equation:

$$\mathcal{L}(\theta_c^{(train)}) = \mathbb{E}_{(s,a,r,s')\mathcal{D}} \sim \left(\mathcal{Q}\left(\theta_c^{(train)} | s^{(t+1)}, a'\right) - y\right)^2, \tag{35}$$

where the replay buffer $\mathcal{D}$ is defined as a collection of transitions (s, a, r, s'). In algorithms following the DDPG framework, the target network undergoes an update during each main network update through a Polyak averaging process:

$$\theta_a^{(target)} \leftarrow \rho\theta_a^{(train)} + (1-\rho)\theta_a^{(target)}, \tag{36}$$
$$\theta_c^{(target)} \leftarrow \rho\theta_c^{(train)} + (1-\rho)\theta_c^{(target)}, \tag{37}$$

where ρ is a hyperparameter between 0 and 1 (usually close to 1). An exploratory policy π' is constructed by adding a noise process $\mathcal{N}$ into our the actor policy:

$$\pi'\left(s^{(t)}\right) = \pi\left(s^{(t)} | \theta_a^{\pi}\right) + \mathcal{N}. \tag{38}$$

C. DDPG-BASED JOINT OPTIMIZATION OF POWER ALLOCATION AND UAV LOCATION

1) ALGORITHM IMPLEMENTATION

This subsection present the algorithm design based on DDPG. Specifically, the power allocation and UAV location are jointly optimized as discovering the changes of three-fold components: state/observation space, action space, and reward design, as described below.

- State/Observation Space: At each time step t, an observation is formed using the current environmental state $s^{(t)}$, encompassing the channels from the BS to the UAV and from the UAV to the K UEs. Subsequently, the state space at time step t can be defined as

$$\begin{aligned} s^{(t)} \triangleq \Big[& p_1^{(t-1)}, \ldots, p_K^{(t-1)}, x_V^{(t-1)}, y_V^{(t-1)}, z_V^{(t-1)}, \\ & d(c_{U_1}, c_V)^{(t-1)}, \ldots, d(c_{U_K}, c_V)^{(t-1)}, \\ & d(c_B, c_V)^{(t-1)}, \\ & R_1^{(t-1)}, \ldots, R_K^{(t-1)} \Big]. \end{aligned} \tag{39}$$

- Action Space: The action space at time step t is determined by

$$a^{(t)} = [p_1^{(t)}, \ldots, p_K^{(t)}, x_V^{(t)}, y_V^{(t)}, z_V^{(t)}]. \tag{40}$$

Algorithm 1 Algorithm to Solve Problem (24)

Input:
  The learning rate μ_a, μ_c.
  The soft update coefficient τ_c.
  The discount factor γ and batch size.
  The locations of UEs and UAV.
Initialization:
  Establish the replay buffer $\mathcal{D}$.
  Initialize the critic network $Q(s, a|\theta^Q)$ and the actor $\pi(s|\theta_a^{\pi})$.
  Set the training actor and critic network parameters $\theta_a^{(train)}$ and $\theta_c^{(train)}$.
  Set the target actor and critic network parameters $\theta_a^{(target)}$ and $\theta_c^{(target)}$.
  Set the number of episodes n_{Epi} and time steps n_{TS}.
for $epi = 1 : n_{\text{Epi}}$ **do**
  Compute the distances among nodes.
  Set $p_k = 1/K, k = 1, \ldots, K$.
  Construct the initial state $s^{(0)}$.
  for $t = 1 : n_{\text{TS}}$ **do**
    Select action from the actor network according to the current policy.
    Get $[p_1^{(t)}, \ldots, p_K^{(t)}, x_V^{(t)}, y_V^{(t)}, z_V^{(t)}]a^{(t)}$ from $a^{(t)}$ as in (40).
    Execute action $a^{(t)}$ and observe reward $R^{(t+1)}$ by (41).
    Observe and obtain the state $s^{(t+1)}$ as in (39).
    Add the experience $(s^{(t)}, a^{(t)}, R^{(t+1)}, s^{(t+1)})$ into the replay buffer.
    Update $\mathcal{Q}(s^{(t)}, a^{(t)})$ in (32) by minimizing the loss function in (45).
    Soft update the actor and critic target networks as in (47) and
  end for
end for
Output: optimal action a.

- Reward: Obtained after executing the action $a^{(t)}$, the reward at time step t is defined at time step t as the sum rate:

$$r^{(t)} = R_\Sigma(p, c_v). \tag{41}$$

- Constraints: To satisfy constraints (24b) and (24c), the sigmoid operator is used in the output of the actor network, and at this time, the action's values are in the range [0,1]. The constraint (24c) is easily satisfied by scaling the values in the range $c_v = c_v(c_{v_{max}} - c_{v_{min}}) + c_{v_{min}}$. Meanwhile, constraint (24b) is handled by normalizing $p_k = p_k / \sum_{k=1}^{K}(p_k)$.

Summarily, the DDPG-based algorithm is described in Algorithm 1, where each iteration of the episode computes the power allocation and UAV location corresponding to one channel realization.

2) COMPLEXITY ANALYSIS

The proposed algorithm, is developed within the DDPG framework, exhibits the following complexities:

- The main algorithm efficiently manages the states and actions to maximize rewards at each time step t. In the worst case, this procedure takes $\mathcal{O}(|s^{(t)}| \cdot |a^{(t)}|)$, where $|x|$ denotes the number of elements in vector x. Here, $|s^{(t)}| = 3K + 4$ and $|a^{(t)}| = K + 3$.
- Achieving convergence for $s^{(t)}$ and $a^{(t)}$, involves iterating over t, requiring n_{TS} iterations of the main algorithm.

The overall complexity of Algorithm 1 is determined as $\mathcal{O}(n_{\text{TS}}(3K+4)(K+3))$, which is equivalent to $\mathcal{O}(n_{\text{TS}}K^2)$. It should be noted that the proposed algorithm operates within the DRL framework, where a substantial amount of data is required for the reinforcement learning process. Therefore, it takes several CBs for the DRL process to converge and reach an optimal solution. The outermost loop, n_{Epi} of CBs, is needed to find a satisfactory solution, rather than the loop inside the proposed algorithm. The convergence rate for the reinforcement learning process will be further examined in Section V.

V. NUMERICAL RESULTS AND DISCUSSIONS

We will now validate the accuracy of our analytical results concerning the UAV-relay massive MIMO system with imperfect CSI and assess the efficiency of the proposed algorithm in solving the formulated optimization problem. It should be noted that our analysis considers an imperfect CSI transmission protocol, where the UAV-relay functions without small-scale fading information. Instead, it relies exclusively on knowledge about large-scale fading to amplify and forward the received signal to the UEs. To evaluate the system performance, we perform extensive simulations using Python, involving 1000 channel realizations. In order to ensure the robustness and reliability of our numerical results, we execute 100 episodes for each channel realization. Moreover, we simulate each episode over a total of 100 steps, allowing for a comprehensive assessment of the system's performance and capturing any dynamics or variations that may arise during the simulation.

A. SIMULATION PARAMETERS

As described in Section II, our system model features a BS equipped with a massive MIMO array with N antennas transmitting data to K UEs through a UAV-relay equipped with N_r antennas, considering a scenario where direct transmission is obstructed. For the simulation setup, we initially randomize the positions of the UEs within the ranges $(x, y, z) \in (0{:}500, -50{:}50, 0{:}1.5)$ [m]. The ground-based BS is set fixed at the origin position, with an assumed height of 20 m, i.e., $(0, 0, 20)$ [m]. The initial position of the UAV is set to $(350, 0, 50)$ [m], where the UAV's altitude is restricted within the allowable range of {40:100} [m] [39].

The uplink training power is uniformly set to 10 dBm for each UE and 23 dBm for the UAV-relay in all cases. For downlink data transmission, the BS is assigned a maximum power of 43 dBm, while the UAV operates at 23 dBm. Unless explicitly stated otherwise, we present the numerical results based on a fixed signal-to-noise ratio (SNR) of 15 dB. It is important to note that the number of uplink pilot sequences is assumed to be equal to the number of UEs, ensuring that each UE is assigned a unique pilot sequence for accurate channel estimation. For further details on the simulation parameters, please refer to Table 1.

TABLE 1. Simulation parameters.

Parameter	Value
System bandwidth	20 MHz
Noise power spectral density at UEs	-174 dBm/Hz
Power budget at BS, P_0	23 dBm
Power budget at UAV, P_{Vd}	10 dBm
Samples per CB, τ_c	200

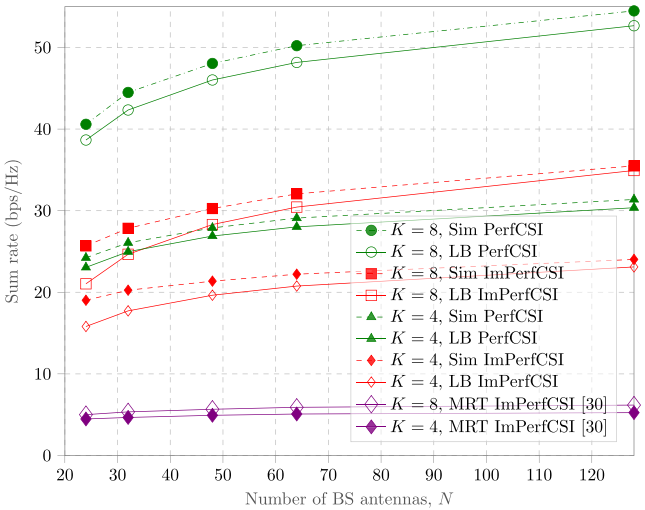

FIGURE 2. Average sum rate as a function of the number of BS antennas (N) for different values of UEs $K = \{4,8\}$ with $N_r = 16$.

B. ON THE VALIDITY OF ZF PERFORMANCE ANALYSIS

To validate the analytical derivations presented in Section III-A, we first examine the system sum rate employing ZF processing as a function of the number of BS antennas (N), while keeping the number of UAV-relay antennas fixed at $N_r = 16$. The corresponding plot in Figure 2 showcases the results of this analysis. We investigate two scenarios: perfect CSI at the BS (labeled as 'PerfCSI' and depicted by the green curves) and imperfect CSI (specifically, no CSI at the UAV-relay, labeled as 'ImPerfCSI' and depicted by the red curves). For both cases, we present simulation results (labeled as 'Sim') as well as theoretical lower bounds (labeled as 'LB').

In our analysis, we consider two different numbers of UEs: $K = 4$ and $K = 8$, while adhering to the constraint that the number of UAV-relay antennas is greater than or

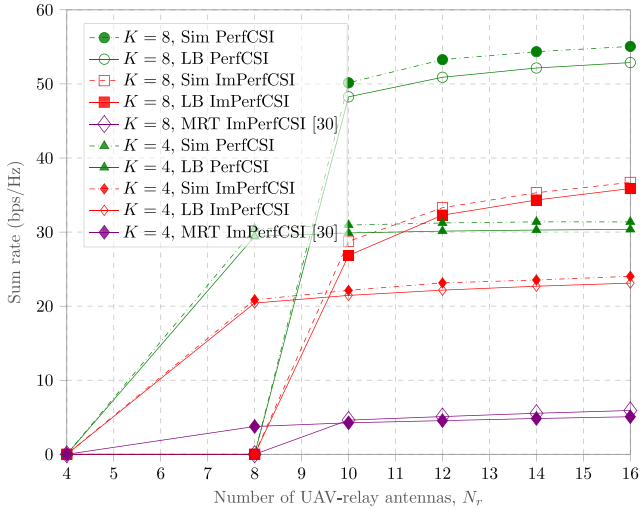

FIGURE 3. Average sum rate as a function of the number of the UAV-relay antennas (N_r) for different values of number of UEs $K = \{4,8\}$ with fixed $N = 128$.

equal to the number of UEs, i.e., $N_r \geq K$. To verify the benefits of ZF technique, we also include here the analysis of MRT processing from [30] for comparison purposes. The corresponding results are depicted by the purple curves, taking into account the presence of imperfect CSI.

Figure 2 clearly illustrates the substantial gain in sum rate achieved by employing ZF processing compared to MRT, particularly in scenarios with a high number of UEs. For instance, at $N = 64$, the sum rate difference between ZF and MRT is approximately 15 bps/Hz for $K = 4$, while this difference expands to about 27 bps/Hz for $K = 8$. Moreover, the impact of CSI unawareness on system performance, especially with an increased number of UEs, becomes apparent. At $N = 64$ and $K = 8$, there is a substantial discrepancy between the curves representing perfect CSI and imperfect CSI. The gap between these curves is approximately 28 bps/Hz, nearly four times larger than the gap observed for $K = 4$, which is approximately 7 bps/Hz. This trend emphasizes the amplifying effect of imperfect CSI on the disparity between the system performance assumed under perfect CSI and the actual performance considering the reality of imperfect CSI.

To further verify the accuracy of the analytical derivations for ZF processing, Figure 3 depicts the system sum rate as a function of the number of UAV-relay, with the number of BS antennas fixed at $N = 128$. Due to the constraint $N_r \geq K$, the starting points for $K = 4$ and $K = 8$ coincide at the same values of N_r, respectively. Consistent with the trends observed in Figure 2, similar patterns emerge, confirming the reliability of the derived analytical expressions and the outperformance of ZF over MRT [30]. As N_r increases, a widening gap between the simulation and analytical sum rates is observed, although with marginal differences. For instance, at $N_r = 16$, the gap is approximately 3% for $K = 4$ and approximately 4% for $K = 8$.

Moreover, Figure 3 emphasizes that increasing the number of UAV-relay antennas has a positive impact on the system sum rate, especially in scenarios where channel knowledge is limited and a relatively large number of UEs are involved. For instance, in cases with channel estimation, the system sum rate improves from approximately 21 bps/Hz (at $N_r = 10$) to about 24 bps/Hz (at $N_r = 16$) for $K = 4$, representing a 13% improvement. This highlights the benefits of increasing the number of UAV-relay antennas, as it allows for better spatial multiplexing and enhanced performance. In contrast, for $K = 8$, the gap widens as the number of UAV-relay antennas increases. The system sum rate increases from around 27 bps/Hz (at $N_r = 10$) to about 37 bps/Hz (at $N_r = 16$), indicating an almost 27% improvement. This substantial increase in the sum rate demonstrates the advantages of having a larger number of UAV-relay antennas, particularly when dealing with a higher number of UEs.

C. EFFECTIVENESS OF THE PROPOSED DDPG-BASED ALGORITHM

In Section V-B, both Figures 2 and 3 exhibit a considerable reduction in SE between scenarios with perfect CSI and those with imperfect CSI, demonstrating a substantial penalty due to the lack of channel knowledge. In this section, we evaluate the effectiveness of the proposed algorithm, tailored for the imperfect CSI scenario, in mitigating this penalty. Since DDPG is well-suited for continuous action spaces, leveraging a deterministic policy, the DDPG-based approach can optimize the UAV-relay's position during execution. Here we show the DDPG performance for the proposed system with and without the application of the ZF technique. For conventional precoding transmission, where the ZF technique is not employed, the curve obtained solely by the DDPG algorithm is labeled as "Rate DDPG non-ZF'. Conversely, the curves obtained by the proposed DDPG-based algorithm when integrated with ZF transmission, are labeled as 'Sim DDPG/Opt. UAV pos.' and 'LB DDPG/Opt. UAV pos.' for the simulation and theoretical lower bound results, respectively. It is important to note that the analytical curves labeled as 'Sim ImPerfCSI' and 'LB ImPerfCSI' are obtained under the assumption of a fixed initial position of the UAV. In contrast, the curves representing the DDPG-based algorithm integrated with ZF transmission consider the optimization of the UAV position as part of the overall system performance evaluation.

Figure 4 displays the performance of the sum rate as a function of the number of BS antennas (N), while keeping the number of UAV-relay antennas (N_r) and the number of UEs (K) fixed at $N_r = 16$ and $K = 8$. In this plot, we compare the effectiveness of the proposed DDPG-based algorithm with the theoretical and simulation performance under the assumption of imperfect CSI. Notably, both the theoretical lower bound derivations and simulation results indicate that the proposed DDPG-based algorithm for ZF transmission exhibits a significant improvement over the analytical curves. The sum rate improvement between the DDPG-based algorithm and the theoretical lower bound remains consistently around 7 bps/Hz as the number of BS

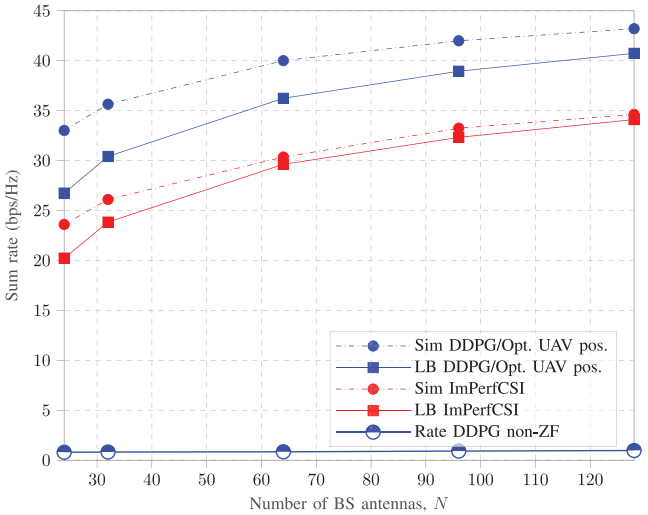

FIGURE 4. Average sum rate as a function of number of BS antennas (N) with fixed $N_r = 16$ and $K = 8$.

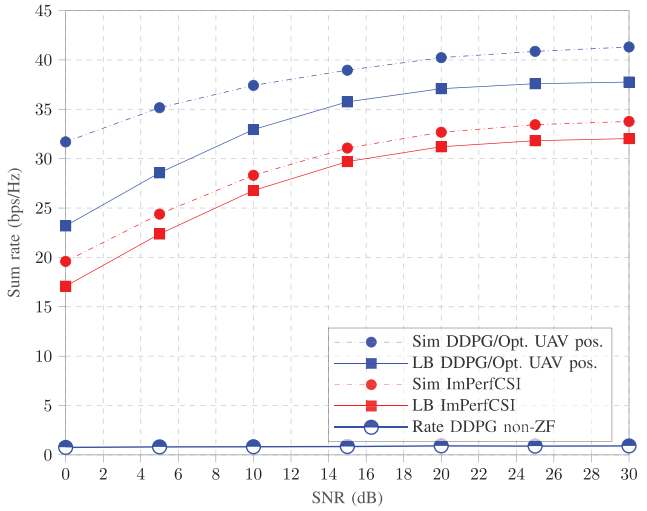

FIGURE 5. Average sum rate as a function of SNR [dB] with fixed $N = 64$, $N_r = 16$ and $K = 8$.

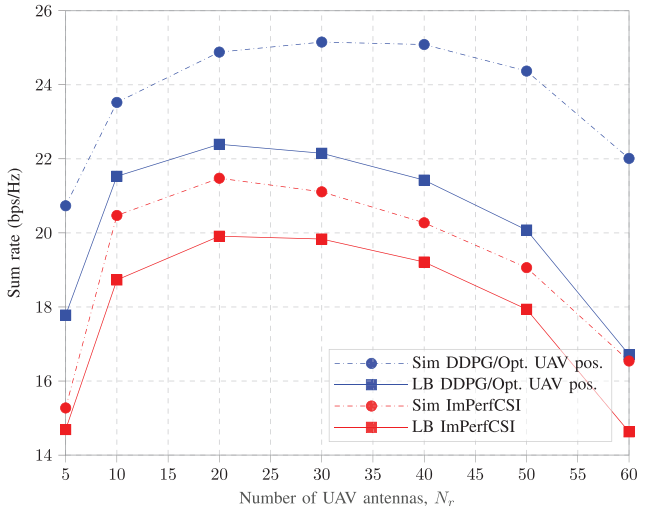

FIGURE 6. Average sum rate as a function of the number of the UAV-relay antennas (N_r) with fixed $N = 64$, $K = 4$, and SNR = 15dB.

antennas (N) increases. Furthermore, when compared to the simulation curve, the proposed algorithm demonstrates a substantial enhancement in the sum rate, increasing from 35 bps/Hz to 43 bps/Hz when $N = 128$ antennas. This notable improvement can be attributed to the DDPG algorithm's ability to effectively balance between exploration and exploitation. By exploring different locations, the algorithm identifies optimal positions that maximize the system sum rate. Nevertheless, the DDPG is effective only when ZF precoding is applied for the system of interest. We show that by involving the cases where the DDPG-based algorithm is employed without using ZF technique. Obviously, the huge gaps observed between the ZF curves and the recently added "Rate DDPG non-ZF" curve can be viewed as evidence of the effectiveness of the powerful ZF processing technique. These gaps highlight the performance disparity between employing ZF and conventional precoding techniques.

Figure 5 exhibits similar trends, focusing on system performance across different SNR levels, where we perform the system performance in terms of SNR, while keeping $N = 64$, $N_r = 16$, and $K = 8$ constant. This further validates the analytical performance and the effectiveness of the proposed algorithm. The DDPG algorithm, based on simulations, consistently outperforms the analytical results, particularly at low SNR levels, showcasing its ability to adapt. These results underscore the efficacy of the proposed DDPG-based algorithm in optimizing system performance, particularly with the integration of ZF, surpassing both theoretical lower bounds and simulations. The algorithm's ability to leverage reinforcement learning techniques and explore the solution space enables it to achieve superior performance in terms of sum rate optimization.

To investigate the impact of the number of UAV-relay antennas (N_r) on the effectiveness of the proposed algorithm, Figure 6 is provided with fixed conditions of $N = 64$, $K = 4$, and SNR = 15dB. The figure illustrates that the performance gap introduced by the simulation-based DDPG approach marginally widens with an increase in N_r. Conversely, the gap for the theoretical-based DDPG approach decreases as N_r grow. Furthermore, both the analytical and DDPG-based results exhibit a similar trend as N_r ranges from 5 to 60 antennas. Within this range, the sum rate curves initially rise proportionally with the growth of N_r, reaching their peaks at $N_r = 20$, after which they gradually decline until $N_r = 60$. This observed behavior aligns with the analytical predictions provided by (21) and (17), indicating that increasing N_r does not invariably lead to performance enhancement. The theoretical analysis suggests the presence of an optimal N_r for a given configuration of N, K, underscoring the necessity of striking a suitable balance between the number of UAV-relay antennas and system performance.

Motivated by the insights gained from Figure 6, we delve deeper into the influence of the number of UEs on system performance. Figure 7 showcases the average sum rate as a function of K, while keeping other parameters fixed at

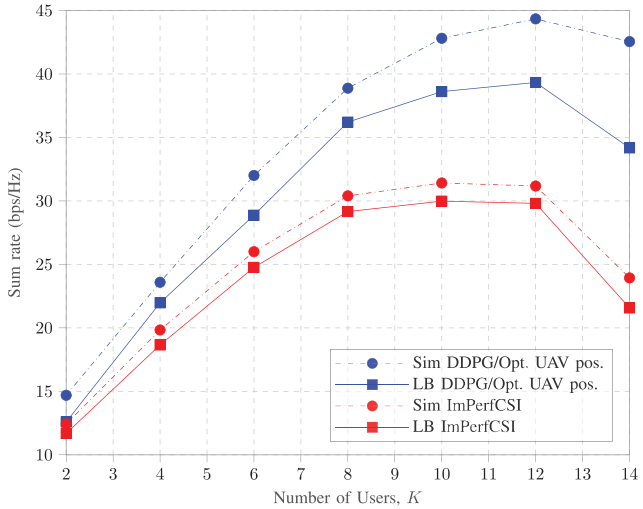

FIGURE 7. Average sum rate as a function of number of UEs (K) with fixed SNR = 15 dB, $N = 64$, and $N_r = 16$.

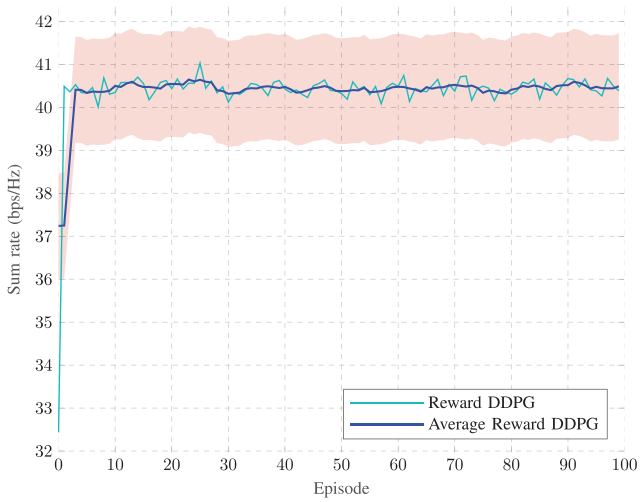

FIGURE 8. Convergence speed of the algorithms.

$N = 64$, $N_r = 16$, and SNR = 15dB. Interestingly, the proposed DDPG-based algorithm demonstrates a significant improvement in the system sum rate, particularly with a larger number of UEs. For example, at $K = 2$, the sum rate enhancement achieved by the LB-based DDPG algorithm over the LB-based imperfect CSI assumption is approximately 1 bps/Hz. Similarly, the improvement between the simulation-based DDPG algorithm and the simulation-based imperfect CSI assumption is about 2 bps/Hz. These gaps become even more pronounced at $K = 12$, reaching 9 bps/Hz and 14 bps/Hz, respectively. Similar to the impact of N_r, the optimal number of UEs can be identified for a specific system configuration based on the theoretical analysis presented in Section II-B. This highlights the importance of considering the number of UEs as a critical parameter in optimizing system performance through the proposed DDPG-based algorithm.

Finally, we investigate the convergence sped of the proposed DDPG-based algorithm over training episodes. Figure 8 showcases the convergence behavior of the system sum rate for two cases: computing reward DDPG and average reward DDPG, with fixed values of $N = 64$, $N_r = 16$, and $K = 8$. Remarkably, the figure demonstrates that the proposed algorithm achieves rapid convergence of the sum rate. After approximately 3 episodes, the algorithm reaches stability and continues to converge towards a steady value. This stable value is observed to be around 40.5 bps/Hz, indicating the optimized system performance achieved by the algorithm.

VI. CONCLUSION

In this study, we conducted an analysis of the theoretical performance of a UAV-relay massive MIMO network under the constraints of imperfect CSI. We utilized the favorable ZF processing technique at the BS due to its high spectral efficiency and low complexity. The absence of direct transmission led us to leverage a UAV as an amplify-and-forward aerial relay to enhance communication from the BS to the target UEs. Our investigation began by deriving a closed-form expression for the lower bound of SE. We employed the MMSE approach for channel estimation and a ZF precoding scheme for downlink data transmission.

Subsequently, we formulated an optimization problem with the primary objective of maximizing the overall sum rate. To tackle this problem, we employed the DDPG approach, which enabled us to optimize the UAV's position and power allocation for the UEs. Our numerical results not only validated the accuracy of the analytical derivations but also demonstrated the effectiveness of the proposed DDPG-based algorithm. The outcomes revealed a significant enhancement in the system sum rate, particularly in scenarios with imperfect CSI. Notably, the proposed DDPG algorithm exhibited superior performance in simulation results, catering to a relatively large number of UEs and ground BS antennas while maintaining a moderate number of UAV-relay antennas.

Overall, our study contributes to the understanding of UAV-relay massive MIMO systems operating under imperfect CSI. We successfully applied analytical derivations and the DDPG algorithm to optimize system performance and achieve significant improvements in the sum rate. These findings have practical implications for scenarios where direct transmission is not feasible, and UAVs can be leveraged as relays to enhance communication in wireless networks.

APPENDIX A
PROOF OF (17)

We begin by introducing the following remark, which will be useful for the subsequent derivation.

Remark 2: The $m \times m$ random matrix $\mathbf{W} = \mathbf{Z}\mathbf{Z}^H$, where $\mathbf{Z} \sim \mathcal{CN}(\mathbf{0}_{m\times n}, \mathbf{I}_m \otimes \mathbf{I}_n)$, is referred to as a central Wishart matrix with n degrees of freedom ($n \geq m$). According to [40, Lemma 2.10], the following property holds:

$$\mathbb{E}\left\{\mathrm{tr}\left(\mathbf{W}^{-1}\right)\right\} = \frac{m}{n-m}. \tag{42}$$

Proof: From (15), it is expressed as

$$\begin{aligned}\mathbf{A} &= \hat{\mathbf{G}}\left(\hat{\mathbf{G}}^H\hat{\mathbf{G}}\right)^{-1}\mathbf{D}_g^{1/2}\\ &= \mathbf{Z}_1\mathbf{Z}_2\mathbf{D}_g^{1/2}\left(\mathbf{D}_g^{1/2}\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\mathbf{D}_g^{1/2}\right)^{-1}\mathbf{D}_g^{1/2}\\ &= \mathbf{Z}_1\mathbf{Z}_2\left(\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\right)^{-1}. \end{aligned} \tag{43}$$

Then

$$\begin{aligned}\mathbf{A}^H\mathbf{A} &= \left(\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\right)^{-1}\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\left(\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\right)^{-1}\\ &= \left(\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2\right)^{-1}. \end{aligned} \tag{44}$$

Following the approach in [37], the precoding matrix is designed as

$$\mathbf{W}_{\mathrm{ZF}} = \sqrt{c}\mathbf{A}\mathbf{D}_p^{1/2}, \tag{45}$$

where the diagonal elements of $\mathbf{D}_p = \mathrm{diag}(p_1, \ldots, p_K)$ represent the signal power of K UEs. To ensure that the total transmitted power remains within the bounds of unity, i.e., $\mathbb{E}\{||\mathbf{W}_{\mathrm{ZF}}\mathbf{x}||^2\} \leq 1$, the scaling factor c is intricately designed as follows:

$$\begin{aligned}\mathbb{E}\left\{||\mathbf{W}_{\mathrm{ZF}}\mathbf{x}||^2\right\} &= \mathbb{E}\left\{\mathrm{tr}\left(\mathbf{W}_{\mathrm{ZF}}\mathbf{x}\mathbf{x}^H\mathbf{W}_{\mathrm{ZF}}^H\right)\right\}\\ &= c\mathbb{E}\left\{\mathrm{tr}\left(\mathbf{A}^H\mathbf{A}\mathbf{D}_p\right)\right\}\\ &= c\mathbb{E}\left\{\mathrm{tr}\left([\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2]^{-1}\mathbf{D}_p\right)\right\}\\ &= c\sum_{k=1}^{K}p_k\mathbb{E}\left\{\mathrm{tr}\left([\mathbf{Z}_2^H\mathbf{Z}_1^H\mathbf{Z}_1\mathbf{Z}_2]_{k,k}^{-1}\right)\right\}\\ &\overset{(a)}{\leq} c\sum_{k=1}^{K}p_k\mathbb{E}\left\{\mathrm{tr}\left([\mathbf{Z}_1^H\mathbf{Z}_1]_{k,k}{}^{-1}\right)\mathrm{tr}\left([\mathbf{Z}_2^H\mathbf{Z}_2]_{k,k}^{-1}\right)\right\}\\ &\overset{(b)}{\leq} c\sum_{k=1}^{K}p_k\frac{N_r}{N-N_r}\frac{1}{N_r-K} \leq 1, \end{aligned} \tag{46}$$

where (a) is based on Jensen inequality [30] and (b) is obtained by using (42) in Remark 2. From (46), it leads to

$$c = \frac{(N-N_r)(N_r-K)}{N_r}. \tag{47}$$

■

REFERENCES

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

[2] C. Pham-Quoc et al., "Robust 3D beamforming for secure UAV communications by DAE," *Mobile Netw. Appl.*, vol. 28, no. 2, pp. 1–9, Apr. 2023.

[3] 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, Sep. 2021, Art. no. 103114.

[4] 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.

[5] H. Nguyen-Kha, H. V. Nguyen, M. T. P. Le, and O.-S. Shin, "Joint UAV placement and IRS phase shift optimization in downlink networks," *IEEE Access*, vol. 10, pp. 111221–111231, 2022.

[6] L. Sanguinetti, E. Björnson, and J. Hoydis, "Toward massive MIMO 2.0: Understanding spatial correlation, interference suppression, and pilot contamination," *IEEE Trans. Commun.*, vol. 68, no. 1, pp. 232–257, Jan. 2020.

[7] M. T. P. Le, L. Sanguinetti, E. Björnson, and M.-G. D. Benedetto, "Code-domain NOMA in massive MIMO: When is it needed?" *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4709–4723, May 2021.

[8] E. Björnson. "A look at an LTE-TDD Massive MIMO product." 2018. Accessed: Nov. 20, 2023. [Online]. Available: http://ma-mimo.ellintech.se/2018/08/27/a-look-at-an-lte-tdd-massivemimo-product/

[9] J. Du, W. Xu, Y. Deng, A. Nallanathan, and L. Vandendorpe, "Energy-saving UAV-assisted multiuser communications with massive MIMO hybrid beamforming," *IEEE Commun. Lett.*, vol. 24, no. 5, pp. 1100–1104, May 2020.

[10] L. Zhu, J. Zhang, Z. Xiao, X. Cao, D. O. Wu, and X.-G. Xia, "3-D beamforming for flexible coverage in millimeter-wave UAV communications," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 837–840, Jun. 2019.

[11] Z. Chen, N. Zhao, D. K. C. So, J. Tang, X. Y. Zhang, and K.-K. Wong, "Joint altitude and hybrid beamspace precoding optimization for UAV-enabled multiuser mmWave MIMO system," *IEEE Trans. Veh. Technol.*, vol. 71, no. 2, pp. 1713–1725, Feb. 2022.

[12] P. Chandhar, D. Danev, and E. G. Larsson, "Massive MIMO for communications with drone swarms," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1604–1629, Mar. 2018.

[13] N. Gao, X. Li, S. Jin, and M. Matthaiou, "3-D deployment of UAV swarm for massive MIMO communications," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 10, pp. 3022–3034, Oct. 2021.

[14] W. Belaoura, K. Ghanem, M. Z. Shakir, and M. O. Hasna, "Performance and user association optimization for UAV relay-assisted mm-wave massive MIMO systems," *IEEE Access*, vol. 10, pp. 49611–49624, 2022.

[15] Y. Wang, M. Giordani, X. Wen, and M. Zorzi, "On the beamforming design of millimeter wave UAV networks: Power vs. capacity trade-offs," *Comput. Netw.*, vol. 205, Mar. 2022, Art. no. 108746.

[16] Y. Wang, X. Chen, Y. Cai, and L. Hanzo, "RIS-aided hybrid massive MIMO systems relying on adaptive-resolution ADCs: Robust beamforming design and resource allocation," *IEEE Trans. Veh. Tech.*, vol. 71, no. 3, pp. 3281–3286, Mar. 2021.

[17] M. Mahmood, A. Koc, and T. Le-Ngoc, "PSO-based joint UAV positioning and hybrid precoding in UAV-assisted massive MIMO systems," in *Proc. IEEE Veh. Technol. Conf.*, 2022, pp. 1–6.

[18] M. Mahmood, M. Ghadaksaz, A. Koc, and T. Le-Ngoc, "Deep learning meets swarm intelligence for UAV-assisted IoT coverage in massive MIMO," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 7679–7696, Mar. 2024.

[19] T. Mir et al., "Relay hybrid precoding in UAV-assisted wideband millimeter-wave massive MIMO system," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7040–7054, Sep. 2022.

[20] D. W. Matolak and R. Sun, "Air—Ground channel characterization for unmanned aircraft systems—Part III: The suburban and near-urban environments," *IEEE Trans. Veh. Technol.*, vol. 66, no. 8, pp. 6607–6618, Aug. 2017.

[21] X. Zhu, F. Qi, and Y. Feng, "Deep-learning-based multiple beamforming for 5G UAV IoT networks," *IEEE Netw.*, vol. 34, no. 5, pp. 32–38, Sep./Oct. 2020.

[22] 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 Tut.*, vol. 25, no. 4, pp. 3038–3067, 4th Quart., 2023.

[23] B. Omoniwa, B. Galkin, and I. Dusparic, "Optimizing energy efficiency in UAV-assisted networks using deep reinforcement learning," *IEEE Wireless Commun. Lett.*, vol. 11, no. 8, pp. 1590–1594, Aug. 2022.

[24] 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.

[25] E. Fonseca, B. Galkin, R. Amer, L. A. DaSilva, and I. Dusparic, "Adaptive height optimization for cellular-connected UAVs: A deep reinforcement learning approach," *IEEE Access*, vol. 11, pp. 5966–5980, 2023.

[26] H. Huang, Y. Yang, H. Wang, Z. Ding, H. Sari, and F. Adachi, "Deep reinforcement learning for UAV navigation through massive MIMO technique," *IEEE Trans. Veh. Technol.*, vol. 69, no. 1, pp. 1117–1121, Jan. 2020.
[27] K. Zhi, C. Pan, G. Zhou, H. Ren, M. Elkashlan, and R. Schober, "Is RIS-aided massive MIMO promising with ZF detectors and imperfect CSI?" *IEEE J. Sel. Areas Commun.*, vol. 40, no. 10, pp. 3010–3026, Oct. 2022.
[28] X. Li, M. Zhang, H. Chen, C. Han, L. Li, D.-T. Do, S. Mumtaz, and A. Nallanathan, "UAV-enabled Multi-pair Massive MIMO-NOMA Relay Systems with Low-Resolution ADCs/DACs," *IEEE Trans. Veh. Tech.*, vol. Early Access, pp. 1–14, 2023.
[29] M. T. P. Le, H. V. Nguyen, V. Nguyen-Duy-Nhat, H. N. Tany, and H. Nguyen-Le, "On the spectral efficiency analysis and optimization for UAV-relay massive MIMO network," in *Proc. 9th IEEE Int. Conf. Commun. Electron. (ICCE)*, 2022, pp. 118–123.
[30] M. T. P. Le, H. V. Nguyen, V. Nguyen-Duy-Nhat, and L. Sanguinetti, "QoE-aware power allocation for aerial-relay massive MIMO networks," *IEEE Trans. Netw. Service Manag.*, vol. 21, no. 1, pp. 477–489, Feb. 2024.
[31] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Trajectory design and power control for multi-UAV assisted wireless networks: A machine learning approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7957–7969, Aug. 2019.
[32] Q. Feng, J. McGeehan, E. K. Tameh, and A. R. Nix, "Path loss models for air-to-ground radio channels in urban environments," in *Proc. IEEE 63rd Veh. Technol. Conf.*, vol. 6, 2006, pp. 2901–2905.
[33] T. Q. Duong, L. D. Nguyen, H. D. Tuan, and L. Hanzo, "Learning-aided realtime performance optimisation of cognitive UAV-assisted disaster communication," in *Proc. IEEE Glob. Commun. Conf.(GLOBECOM)*, 2019, pp. 1–6.
[34] A. Al-Hourani and K. Gomez, "Modeling cellular-to-UAV path-loss for suburban environments," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 82–85, Feb. 2018.
[35] 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.
[36] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Hoboken, NJ, USA: Prentice-Hall, Inc., 1993.
[37] T. L. Marzetta and H. Yang, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[38] T. P. Lillicrap et al. "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.
[39] M. Alzenad, A. El-Keyi, and H. Yanikomeroglu, "3-D placement of an unmanned aerial vehicle base station for maximum coverage of users with different QoS requirements," *IEEE Wireless Commun. Let.*, vol. 7, no. 1, pp. 38–41, Feb. 2018.
[40] A. M. Tulino et al., "Random matrix theory and wireless communications," *Foundations and Trends® Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.

MAI T. P. LE received the Ph.D. degree from the Sapienza University of Rome, Rome, Italy, in February 2019.

Since 2011, she has been with the Faculty of Electronics and Telecommunication Engineering, University of Science and Technology - The University of Danang, Da Nang, Vietnam, where she is currently a Lecturer. From 2015 to 2020, she was a Ph.D. Student and the Postdoctoral Researcher with the Department of Information Engineering, Electronics and Telecommunications, Sapienza University of Rome. In 2016, she was a Visiting Researcher with the Singapore University of Technology and Design, Singapore, in 2016, and the Arizona State University, Tempe, AZ, USA, in 2012. Her main research interests include information theory, mathematical theories, and machine learning and their application in wireless communications. Her current research focuses on physical-layer techniques for the next generation network.

VIEN NGUYEN-DUY-NHAT received the B.Eng. degree in electronic engineering from the University of Science and Technology - The University of Danang (DUT-UD), Vietnam, in July 1997, the M.Eng. degree in electrical engineering from the Ho Chi Minh City University of Technology, Vietnam, in 2003, and the Ph.D. degree from DUT-UD in 2017, where he joined the Faculty of electronics and telecommunication engineering in September 1997. His current areas of interests include Internet of Things, signal processing, optimization, wireless communications, and machine learning.

HIEU V. NGUYEN received the B.E. degree in electronics and telecommunications from the University of Science and Technology - The University of Danang (DUT-UD), Vietnam, in 2011 and the M.E. and Ph.D. degrees in electronic engineering from Soongsil University, Seoul, South Korea, in 2016 and 2020, respectively.

Since 2011, he has been with DUT-UD, where he is currently a Lecturer. From 2014 to 2021, he was a Research Assistant and the Postdoctoral Researcher with the Wireless Communications Laboratory, Soongsil University. From 2021 to 2022, he was a Research Associate with the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. His research interests include wireless communications, with a particular focus on optimization techniques and machine learning for wireless communications, such as UAV/drone communications, device-to-device communications, full-duplex radios, green communication systems, Internet of Things, and satellite networks.

OH-SOON SHIN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 1998, 2000, and 2004, respectively.

From 2004 to 2005, he was a Postdoctoral Fellow with the Division of Engineering and Applied Sciences, Harvard University, MA, USA. From 2006 to 2007, he was a Senior Engineer with Samsung Electronics, Suwon, South Korea. In September 2007, he joined with the School of Electronic Engineering, Soongsil University, Seoul, where he is currently a Professor. His research interests include communication theory, wireless communication systems, and signal processing for communications.