drones

MDPI

Article

NOMA-Based Rate Optimization for Multi-UAV-Assisted D2D Communication Networks

Guowei Wu, Guifen Chen * and Xinglong Gu

College of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130022, China; 2022200108@mails.cust.edu.cn (G.W.); 2020100674@mails.cust.edu.cn (X.G.)
* Correspondence: 2022200109@mails.cust.edu.cn

Abstract: With the proliferation of smart devices and the emergence of high-bandwidth applications, Unmanned Aerial Vehicle (UAV)-assisted Device-to-Device (D2D) communications and Non-Orthogonal Multiple Access (NOMA) technologies are increasingly becoming important means of coping with the scarcity of the spectrum and with high data demand in future wireless networks. However, the efficient coordination of these techniques in complex and changing 3D environments still faces many challenges. To this end, this paper proposes a NOMA-based multi-UAV-assisted D2D communication model in which multiple UAVs are deployed in 3D space to act as airborne base stations to serve ground-based cellular users with D2D clusters. In order to maximize the system throughput, this study constructs an optimization problem of joint channel assignment, trajectory design, and power control, and on the basis of these points, this study proposes a joint dynamic hypergraph Multi-Agent Deep Q Network (DH-MDQN) algorithm. The dynamic hypergraph method is first used to construct dynamic simple edges and hyperedges and to transform them into directed graphs for efficient dynamic coloring to optimize the channel allocation process; subsequently, in terms of trajectory design and power control, the problem is modeled as a multi-agent Markov Decision Process (MDP), and the Multi-Agent Deep Q Network (MDQN) algorithm is used to collaboratively determine the trajectory design and power control of the UAVs. Simulation results show the following: (1) the proposed algorithm can achieve higher system throughput than several other benchmark algorithms with different numbers of D2D clusters, different D2D cluster communication spacing, and different UAV sizes; (2) the proposed algorithm designs UAV trajectory optimization with a 27% improvement in system throughput compared to the 2D trajectory; and (3) in the NOMA scenario, compared to the case of no decoding order constraints, the system throughput shows on average a 34% improvement.

Keywords: unmanned aerial vehicles; non-orthogonal multiple access; D2D; dynamic hypergraph; multi-agent reinforcement learning

check for updates

Academic Editor: Emmanouel T. Michailidis

Received: 13 December 2024
Revised: 5 January 2025
Accepted: 13 January 2025
Published: 16 January 2025

Citation: Wu, G.; Chen, G.; Gu, X. NOMA-Based Rate Optimization for Multi-UAV-Assisted D2D Communication Networks. *Drones* **2025**, *9*, 62. https://doi.org/10.3390/drones9010062

1. Introduction

With the emergence of new devices and innovative applications, the global demand for mobile data is growing at an accelerated pace. As a result, the need for high-quality services has become increasingly urgent worldwide. Furthermore, the rapid growth of the Internet of Things (IoT) requires ubiquitous connectivity [1]. To satisfy users' high data throughput, mobile networks must significantly increase their capacity. However, because radio spectrum resources are scarce, there is an urgent need for new spectrum management and power optimization approaches to fulfill the rapidly expanding demand for wireless connectivity.

Driven by the above challenges, Non-Orthogonal Multiple Access (NOMA) has received significant attention [2,3] and is recognized by both industry and academia. Unlike traditional multiple-access techniques, which require independent or orthogonal band protection, NOMA allows a single resource block to be multiplexed by multiple users who have different allocated power factors through flexible user scheduling and bandwidth allocation strategies. Moreover, the use of overlay coding techniques [4] and serial interference cancellation techniques [5] not only effectively mitigates the rapidly growing capacity problem but also significantly improves network performance in terms of fairness and security.

In addition, as one of the key technologies of 5G, Device-to-Device (D2D) communication refers to the direct exchange of information between devices without requiring transmission through a base station (BS) [6,7]. In traditional cellular networks, D2D users are able to communicate by multiplexing the communication links of cellular users, which improves the spectrum utilization of the system, which in turn improves the network throughput. However, the introduction of D2D technology in cellular networks increases the inter-user cochannel interference in the network, which affects the quality of user communication; furthermore, the quality of communication is ensured by introducing NOMA technology [8], which utilizes a decoding mechanism to eliminate the inter-user interference caused by D2D communication [9,10].

Unmanned Aerial Vehicles (UAVs) have been gradually applied to assisted communication scenarios due to their flexible maneuverability and high-altitude-coverage capability [11,12]. A UAV can be used as an airborne base station or relay device to expand network coverage and enhance communication quality. When UAV technology is combined with D2D and NOMA technologies, wireless resource allocation can be further optimized, especially in high-density device distribution and complex channel environments, to effectively enhance system performance. Specifically, the dynamic deployment capability of UAVs can achieve accurate coverage of D2D user clusters while combining the spectrum multiplexing characteristics of NOMA to maximize spectrum efficiency.

Although the combination of NOMA, D2D technology, and UAVs demonstrates significant performance improvement potential, its converged application still faces a number of challenges. First, in dynamic environments, it is a complex issue to realize UAV trajectory design to cover multiple D2D user clusters while ensuring communication stability and efficiency [13]. Second, the introduction of NOMA technology further increases the complexity of channel assignment and power control. In multi-user coexistence scenarios, how to achieve fairness among users under dynamic channel conditions and effectively suppress interference has become a core problem that needs to be solved [14]. Moreover, in UAV-assisted D2D-NOMA scenarios, there is strong coupling among trajectory design, channel assignment, and power control. Designing joint optimization algorithms to maximize system performance under limited resources while reducing the computational complexity is the focus and challenge of current research.

In view of the above challenges, this paper is devoted to the study of a UAV-assisted D2D-NOMA communication model focusing on the joint optimization of channel allocation, UAV trajectory design, and power control in order to achieve the dual enhancement of spectral efficiency and system capacity. Specifically, this paper addresses the resource allocation and optimization problem of multi-UAV cooperative work in complex 3D environments, including efficient channel allocation to avoid interference, optimization of UAV trajectories, and dynamic adjustment of power control to improve the overall system throughput.

In disaster recovery scenarios, the NOMA-based multi-UAV-assisted D2D communication model proposed in this paper can rapidly deploy UAVs as temporary communication base stations to ensure that the communication needs in the affected area are met in a timely manner after a natural disaster or an emergency, thus enhancing the efficiency and coordination of rescue operations. In a smart city environment, the model supports the communication needs of high-density IoT devices by optimizing the resource allocation and trajectory planning of UAVs, enhancing the communication coverage and system throughput of the city, and facilitating the construction and operation of smart cities. Through this study, we aim to provide an efficient resource optimization scheme for next-generation wireless networks as well as theoretical support and technical guidance for the deployment of practical systems.

1.1. Related Work

Existing work related to this paper covers three different areas, UAV-assisted D2D communication, NOMA in UAV communication, and UAV-assisted wireless communication in reinforcement learning, as follows:

(1) UAV-assisted D2D communication: Currently, UAV-assisted D2D communication research mainly focuses on key issues such as resource scheduling, trajectory design, and power control. The literature [15] constructs a multi-objective optimization model for resource scheduling by comprehensively considering factors such as the number of UAVs and their positions, flight speeds, transmit power, and communication channel allocation, and the study adopts the Non-dominated Sorting Genetic Algorithm-III (NSGA-III) combined with a flexible solution dimensionality mechanism, a discrete portion generation mechanism, and a UAV numbering adjustment mechanism to achieve an efficient solution to the resource scheduling problem. In UAV networks supporting D2D communication, the literature [14] explores how to maximize the total system rate with seamless coverage with a minimum number of UAVs. The study formulates the problem as a non-linear and non-convex optimization model and proposes the maximum rate minimum number (MRMN) scheme. In addition, by applying coalitional game theory, a collaboration strategy between UAV clusters and ground devices is designed and a coalition formation algorithm is developed to optimize the scheduling of communication resources while satisfying the seamless coverage constraints, which significantly improves the system throughput. To address the challenges of resource optimization and latency reduction in heterogeneous Mobile Edge Computing (MEC) environments, the literature [16] proposes a framework for integrating UAV and D2D communications. The study constructs a joint non-convex optimization model containing UAV trajectory design, resource allocation, and task offloading, and it solves the optimal solution by an algorithm combining block coordinate descent (BCD) and potential game. Experimental results show that the method can reduce the system delay by about 20% and exhibits excellent performance in dynamic network scenarios. In addition, the literature [17] proposes a UAV positioning method based on a Gauss–Markov stochastic motion model for the problem of coverage establishment and user rate optimization when UAVs are used as flying base stations. Simulation results show that the method is able to cover 95% of the users and achieves an average available rate of 0.15 Gbps for downlink users, demonstrating its effectiveness in optimizing user coverage and rate.

(2) NOMA in UAV communication: The application of NOMA technology in UAV communication has received much attention. The literature [18] explores the application of a reconfigurable intelligent surface (RIS) in UAV-assisted NOMA networks. The study aims to minimize the system power consumption by optimizing the UAV position, RIS reflection coefficient, transmit power, and decoding order while satisfying the constraints of user rate and UAV spacing. The study decomposes the optimization problem into four subproblems, which are solved iteratively using successive convex approximation (SCA), Gaussian randomization, and standard convex optimization, respectively, which significantly reduces

the total power consumption and validates the advantages of combining RIS with multi-UAV-assisted NOMA networks. To address the need for UAVs to help survivors in post-disaster scenarios, the literature [19] proposes a system that combines UAVs with NOMA by developing effective power allocation (PA) and trajectory planning algorithms (TPAs). The study formulates the problem as a budgeted multi-armed slot machine (BMAB) problem to optimize UAV trajectories and minimize battery consumption, where the UAVs act as "bandit" players and clusters of disaster zones act as "weapons". The problem is solved by two upper confidence bound (UCB) schemes. Compared with the traditional UAV-OMA system, this method increases the total number of assisted survivors by 60%, improves the convergence speed by 80%, and saves energy effectively. The UAV-assisted NOMA downlink scenario is studied in the literature [20]. By jointly optimizing the beamforming and position of the ground base station and the UAV, this study shows that integrating airborne relays in NOMA networks can significantly improve the sum rate of the system under a power budget, successive interference cancellation (SIC) constraints, and quality of service requirements of the ground equipment. Furthermore, the literature [21] proposes a secure NOMA system based on UAV relaying. In this system, a UAV acts as an airborne relay station to assist in the transmission from the source to the two users, taking into account the presence of ground eavesdroppers and friendly jamming UAVs. The results show that the proposed NOMA system significantly improves the secrecy rate of the system and enhances the overall system security compared to orthogonal multiple access (OMA).

(3) UAV-assisted wireless communication in reinforcement learning: The application of reinforcement learning techniques in UAV-assisted wireless communication is significantly improving system performance and network efficiency. The literature [22] explores the combination of UAV and NOMA techniques in IoT, aiming to address the challenges of remote terminal access and low-energy communication. To optimize the system performance, the study proposes a Multi-Agent Federated Reinforcement Learning (MAFAL) algorithm for joint optimization of 3D trajectory design and time allocation of UAVs, with the goal of maximizing energy efficiency (EE) while guaranteeing quality of service (QoS), fairness, and low energy consumption. Another study [23] constructs a NOMA-based UAV-assisted cellular offloading (UACO) framework, focusing on analyzing the coupling between UAV path selection and resource offloading. The study considers the autonomous obstacle avoidance ability of UAVs in complex 3D environments and the influence of obstacles on the channel model, and it significantly improves the spectrum utilization efficiency and communication throughput of the system through the proposed deep reinforcement learning path selection and resource offloading algorithm (UPRA). The study also explores the effect of the reward function on the training convergence and demonstrates the excellent adaptability of the proposed algorithm to random user deployment and maximum mobile speed variation in dynamic networks and complex environments. The literature [24], on the other hand, develops an intelligent UAV navigation solution via federated deep reinforcement learning (DRL) by utilizing flying small base stations (UAV-BS) as relays to enable multiple UAV-BS to fly autonomously and serve ground devices (GDs) in a cost-effective manner. Compared with the conventional navigation methods based on greedy policy (GP) and the traveling salesman problem (TSP), the proposed solution exhibits significant advantages in system performance metrics such as coverage time, coverage score (CS), and channel noise ratio (CNR). The paper [25] proposes a NOMA network framework based on collaborative UAV caching, which aims to minimize the system content retrieval latency by jointly optimizing the caching decision, 3D trajectory design, and spectrum resource allocation. The study addresses the joint optimization problem of UAV trajectory design, power allocation, and channel multiplexing, which is solved using a deep reinforcement learning (DRL) algorithm. Simulation results show that the proposed method significantly outperforms other

benchmark algorithms in key performance metrics such as content hit rate, retrieval delay, and system throughput, and it demonstrates fast convergence capability.

Although there have been studies that have made significant progress in UAV-assisted D2D communication, NOMA in UAV communication, and the application of reinforcement learning in UAV-assisted wireless communication, most of these studies have only considered UAV motion in a 2D plane, failing to take full advantage of UAVs' 3D mobility and failing to incorporate D2D communication, NOMA, and NOMA in 3D scenarios where multiple UAVs collaborate. Therefore, the question of how to jointly optimize channel allocation, UAV trajectory design, and power control based on NOMA's multi-UAV-assisted D2D communication model in a multi-UAV three-dimensional dynamic environment remains a challenging and under-studied problem.

Based on the above research status and its shortcomings, this paper is devoted to the study of a NOMA-based multi-UAV-assisted D2D communication model, focusing on the joint optimization of channel allocation, trajectory design, and power control. By combining advanced dynamic hypergraphs and reinforcement learning algorithms, this paper proposes a multi-intelligent deep Q network algorithm for joint dynamic hypergraphs in multi-UAV 3D scenarios, aiming to cope with complex and dynamic network environments and maximize system throughput. This study not only fills the research gap of joint optimization of D2D-NOMA communication models in multi-UAV 3D scenarios but also provides new ideas for UAV communication in future three-layer heterogeneous air-to-ground networks.

1.2. Motivation and Contribution

The main contributions of this paper are as follows:

(1) This study proposes a NOMA-based multi-UAV-assisted D2D communication model. In this model, multi-UAVs are deployed in 3D space as airborne base stations to provide services to ground cellular users and D2D clusters in which NOMA techniques are applied to D2D communications to improve spectrum utilization efficiency and system throughput. Based on the system model, a problem to maximize the system throughput is proposed: firstly, an in-depth study is carried out for the channel allocation problem, and subsequently, this study jointly optimizes the trajectory design and power control of the UAVs.

(2) For the above channel allocation, trajectory design, and power control problems, this study proposes a joint optimization strategy aimed at maximizing the system throughput. To effectively solve this complex multivariate optimization problem, the overall problem is decoupled into two subproblems. First, for channel allocation, this study constructs dynamic simple edges and hyperedges using the dynamic hypergraph method and achieves efficient dynamic coloring by transforming the dynamic hypergraph into a directed graph, thus optimizing the allocation of channel resources. Secondly, for the trajectory design and power control problem, we propose the MDQN algorithm to intelligently optimize the flight trajectory and power control of the UAVs. Finally, this study proposes a joint dynamic hypergraph Multi-Agent Deep Q Network algorithm, which can dynamically optimize channel allocation, trajectory design, and power control in real-time changing 3D environments, thus maximizing the system throughput.

(3) This study conducts extensive simulations to compare the proposed algorithm with several benchmark algorithms. The results verify that our proposed joint dynamic hypergraph Multi-Agent Deep Q Network algorithm significantly improves the system throughput under different numbers of D2D clusters and different D2D cluster communication spacing schemes, UAV sizes, trajectory designs, and NOMA decoding orders. In addition, the trajectories of 3D UAVs are more reasonable when considering real scenarios.

The rest of this paper is organized as follows. Section 2 introduces the system model. Section 5 establishes a formulation of the maximizing system throughput problem. Section 4 presents the solution to each subproblem. Section 5 discusses the simulation results to demonstrate the advantages of our proposed algorithm, and Section 6 concludes this paper.

2. System Model

In this section, this study first proposes a 3D dynamic UAV communication network model based on NOMA. Then the network model, propagation model, and communication model of the system are proposed.

2.1. Network Model

Consider the NOMA-based UAV-assisted D2D communication model shown in Figure 1, where U UAVs exist in the air to act as airborne base stations and provide communication services to nearby ground users using Orthogonal Frequency Division Multiple Access (OFDMA) technology. Specifically, the set of airborne base stations is denoted as $\mathcal{U} = \{1, 2, \cdots, u, \cdots, U\}$, and $\{\mathcal{S}_1, \mathcal{S}_2, \cdots, \mathcal{S}_u, \cdots, \mathcal{S}_U\}$ is the set of available channels for the U airborne base stations. Meanwhile, we consider the powerful maneuverability of a UAV, assume that the mission period of the UAV's flight in the case of dynamic changes in altitude is $T_{\max}$, partition $T_{\max}$ into L sufficiently small and equal time slots of length $T = T_{\max}/L$, and denote the set of time slot serial numbers as $\mathcal{L} = \{1, 2, \cdots, t, \cdots, L\}$; therefore, the position of the UAV is nearly constant in each time slot, and the UAVs have a position change in the neighboring time slots. In each UAV base station, cellular users occupy a subchannel resource block alone, and these subchannels are independently orthogonal to each other. In order to enhance the spectrum utilization and improve the network transmission efficiency, the NOMA technique is applied to D2D to form D2D-NOMA clusters. Therefore, a large number of cellular users as well as D2D-NOMA clusters are randomly generated in the ground presence, and the number of cellular users as well as the number of D2D-NOMA clusters are denoted as M and N, respectively, with the cellular users denoted by the set $\mathcal{C} = \left\{C_1, C_2, \cdots, C_i, \cdots, C_M\right\}$, the set of M orthogonal channels pre-allocated to the cellular users denoted by the set $\mathcal{S} = \{S_1, S_2, \cdots, S_i, \cdots, S_M\}$, and the D2D-NOMA clusters denoted by the set $\mathcal{D} = \left\{D_1, D_2, \cdots, D_j, \cdots, D_N\right\}$. For convenience, let $S_i \in \mathcal{S}_u$ be the channels pre-allocated to users $C_i \in \mathcal{C}_u$ and $C_i \in \mathcal{C}_u$ be the subset of cellular users associated with UAV base stations u. The D2D-NOMA cluster will reuse the uplink subchannel resources of the cellular users, considering that the user terminal equipment has less processing power and limited battery power consumption compared to the UAV base stations. Meanwhile, in order to keep the signal processing delay as well as the hardware decoding complexity low when transmitting using the D2D-NOMA technique, each D2D cluster D_j contains one transmitter(DT_j) and two receivers ($DR_{j,1}$ and $DR_{j,2}$). More importantly, it is assumed that a D2D cluster D_j can multiplex the channel resources of at most one cellular user, while the channel resources of one cellular user can be multiplexed by multiple D2D clusters. Furthermore, since some D2D clusters are located in the overlapping area of multiple UAV services, when the UAV base station is $u \in \mathcal{U}$, let ξ_u be the maximum number of D2D clusters associated with the UAV base station.

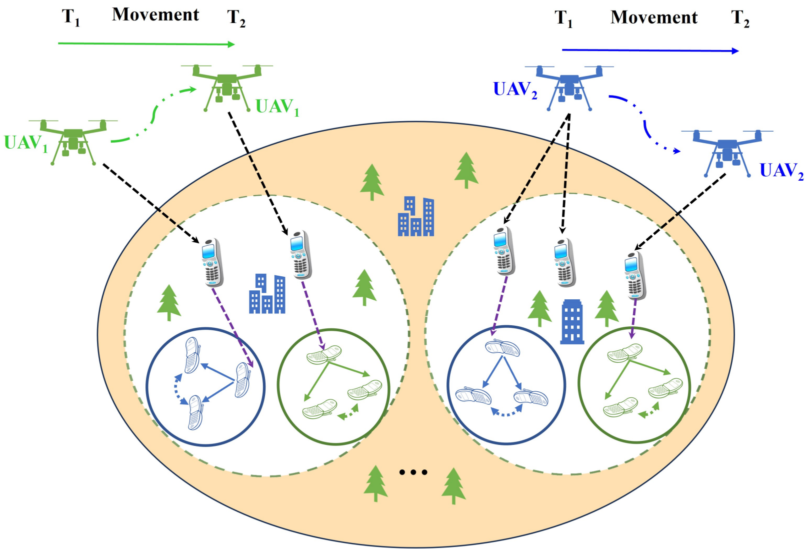


Figure 1. Model diagram of the UAV-assisted D2D-NOMA system.

2.2. Propagation Model

The path loss depends on the LoS and $NLoS$ link states, where the probabilities under the LoS and $NLoS$ conditions are denoted as P_{LoS} and P_{NLoS}, respectively, and $P_{LoS} + P_{NLoS} = 1$ is satisfied. The path losses under conditions LoS and $NLoS$ are defined as L_{LoS} and L_{NLoS}, respectively. Therefore, the average path loss $L_j^i(t)$ between any two points i and j in the model between time slot t is denoted as

$$L_j^i(t) = P_{\mathrm{LoS}} \cdot L_{\mathrm{LoS}} + P_{\mathrm{NLoS}} \cdot L_{\mathrm{NLoS}} \tag{1}$$

The channels between the cellular user and the UAV and between the D2D transmitter and the UAV are modeled as air-to-ground channels. The air-to-ground channel model adopted in this paper is provided by the 3GPP-36.777 standard [26]. For the UAV movement model, $\{x_u(t), y_u(t), h_u(t)\}$ denotes the position of the UAV in time slot t, $v(t)$ denotes the UAV flight speed, and the 3D distance between UAV u and cellular user C_i at the moment t is denoted as $d_i^u\left(t\right)$.

$$d_i^u(t) = \sqrt{h_u^2(t) + [x_u(t) - x_i(t)]^2 + [y_u(t) - y_i(t)]^2} \tag{2}$$

where $\{x_i(t), y_i(t)\}$ denotes the location of the cellular user of the UAV service in time slot t.

Let f_c be the carrier frequency. The path loss $L_i^u(t)$ and the corresponding conditional probability $P_{\mathrm{LoS}}(t)$ between cellular user C_i and UAV u can be expressed by Equations (3) and (4).

$$L_i^u(t) = \begin{cases} 30.9 + (22.25 - 0.5\log_{10} h_u(t)) \log_{10} d_i^u(t)(t) + 20\log_{10} f_c, \text{if LoS link} \\ \\ \max\{L_{\mathrm{Los}}, 32.4 + (43.2 - 7.6\log_{10} h_u(t)) \log_{10} d_i^u(t) + 20\log_{10} f_c\}, \text{if NLoS link} \end{cases} \tag{3}$$

$$P_{\mathrm{LoS}}(t) = \begin{cases} 1, & \text{if } \sqrt{\left(d_i^u(t)\right)^2 - (h_u(t))^2} \leqslant d_0 \\ \frac{d_0}{\sqrt{\left(d_i^u(t)\right)^2 - (h_u(t))^2}} + \exp\left\{\frac{-\sqrt{\left(d_i^u(t)\right)^2 - (h_u(t))^2}}{p_1} + \frac{d_0}{p_1}\right\}, & \text{if } \sqrt{\left(d_i^u(t)\right)^2 - (h_u(t))^2} > d_0 \end{cases} \tag{4}$$

where $d_0 = \max[294.05 \cdot \log_{10} h_u(t) - 432.94, 18]$, $p_1 = 233.98 \cdot \log_{10} h_u(t) - 0.95$.

Considering the small-scale fading, the channel gain from UAV u to cellular user C_i at t time slot can be calculated as in Equation (5):

$$g_i^u(t) = H_i^u(t) \cdot 10^{-L_i^u(t)/10} \tag{5}$$

where $H_i^u(t)$ denotes the attenuation coefficient for UAV u and cellular user C_i at time slot t.

Similarly, the channel gain of D2D transmitter D_j and UAV u in a D2D cluster at time slot t can be calculated as in Equation (6):

$$g_{DT_j}^u(t) = H_{DT_j}^u(t) \cdot 10^{-L_{DT_j}^u(t)/10} \tag{6}$$

where $= H_{DT_j}^u(t)$ denotes the attenuation coefficient of UAV u and D2D transmitter DT_j at t time slot.

The channels between the cellular user and the D2D transmitter and between the D2D receiver and the D2D transmitter are modeled as ground-to-ground channels. The distance between cellular user C_i and D2D transmitter DT_j at the moment t is denoted as $d_{i,DT_j}^u(t)$.

$$d_{i,DT_j}^u(t) = \sqrt{\left[x_i(t) - x_{DT_j}(t)\right]^2 + \left[y_i(t) - y_{DT_j}(t)\right]^2} \tag{7}$$

where $\{x_{DT_j}(t), y_{DT_j}(t)\}$ denotes the position of D2D transmitter DT_j in time slot t.

Thus, the channel gain $g_{i,DT_j}^u(t)$ between the cellular user and the D2D transmitter is [27]

$$g_{i,DT_j}^u(t) = G_{i,DT_j}^u(d_{i,DT_j}^u)^{-\xi}\zeta_{i,DT_j}^u \tag{8}$$

where ξ is the path loss exponent and ζ_{i,DT_j}^u represents the channel gain corresponding to the channel shadow fading between cellular user C_i and D2D transmitter DT_j, obeying a lognormal distribution.

Similarly, the channel gain $g_{i,DT_j,DR_{j,n}}^u(t)$ between the D2D transmitter DT_j and the D2D receiver $DR_{j,n}(n = 1, 2)$ is

$$g_{i,DT_j,DR_{j,n}}^u(t) = G_{i,DT_j,DR_{j,n}}^u(d_{i,DT_j,DR_{j,n}}^u)^{-\xi}\zeta_{i,DT_j,DR_{j,n}}^u \tag{9}$$

where $d_{i,DT_j,DR_{j,n}}^u$ is the distance between D2D transmitter DT_j and D2D receiver $DR_{j,n}$ at time t, and $\zeta_{i,DT_j,DR_{j,n}}^u$ represents the channel gain corresponding to the channel shadow fading between D2D transmitter DT_j and D2D receiver $DR_{j,n}$, obeying a lognormal distribution.

The channel gain $g_{i,DR_{j,n}}^u(t)$ between the cellular user and the D2D receiver is

$$g_{i,DR_{j,n}}^u(t) = G_{i,DR_{j,n}}^u(d_{i,DR_{j,n}}^u)^{-\xi}\zeta_{i,DR_{j,n}}^u \tag{10}$$

where $d_{i,DR_{j,n}}^u$ is the distance between the cellular user and the D2D receiver $DR_{j,n}$ at the moment t, and $\zeta_{i,DR_{j,n}}^u$ represents the channel gain corresponding to the channel shadow fading between the cellular user and the D2D receiver $DR_{j,n}$, obeying a lognormal distribution.

2.3. Communication Model

As analyzed in Figure 1, the following interferences exist under the UAV-assisted D2D-NOMA system: (1) inter-group interference: interference from D2D transmitters in D2D groups that reuse the same subchannels; (2) intra-group interference: interference from superimposed signals to another receiver in the same D2D group; and (3) cellular interference: interference from cellular users that reuse the same subchannels.

Therefore, UAV u receives a signal on subchannel $S_i \in \mathcal{S}_u$ denoted as

$$y_i^u = \sqrt{P_i^u} g_i^u x_i^u + \sum_{D_j \in \mathcal{D}_u} I_{i,DT_j}^u \sqrt{P_{i,D_j}^u} x_{i,DT_j}^u g_{i,DT_j}^u + \eta \tag{11}$$

where P_i^u and P_{i,D_j}^u are denoted as the transmit power of cellular user C_i and D2D cluster D_j, respectively, $\mathcal{D}_u \in \mathcal{D}$ denotes the set of D2D clusters in the coverage area of UAV u, and η is an additive Gaussian white noise with mean 0 and variance σ^2. x_i^u denotes the signal of the cellular user, and x_{i,DT_j}^u is the superimposed signal of DT_j, which can be represented by Equation (12):

$$x_{i,DT_j}^u = \sqrt{a_{i,DR_{j,1}}^u} x_{i,DR_{j,1}}^u + \sqrt{a_{i,DR_{j,2}}^u} x_{i,DR_{j,2}}^u \tag{12}$$

where $a_{i,DR_{j,1}}^u$ and $a_{i,DR_{j,2}}^u$ denote the power allocation coefficients of the D2D receivers $DR_{j,1}$ and $DR_{j,2}$, respectively, and the relationship between them is $a_{i,DR_{j,1}}^u + a_{i,DR_{j,2}}^u \leq 1$. $x_{i,DR_{j,1}}^u$ and $x_{i,DR_{j,2}}^u$ denote the signals of $DR_{j,1}$ and $DR_{j,2}$, respectively. In addition, I_{i,DT_j}^u is a binary decision variable:

$$I_{i,DT_j}^u = \begin{cases} 1, & D_j \in \mathcal{D}_u \text{ occupies } C_i \in \mathcal{C}_u \text{ channel resources,} \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

A D2D cluster can only occupy at most one cellular user channel resource among all UAV BTSs, i.e., there is $\sum_{u \in \mathcal{U}} \sum_{C_i \in \mathcal{C}_u} I_{i,DT_j}^u \leq 1$. In addition, the maximum number of D2D clusters associated with UAV u must satisfy $\sum_{C_i \in \mathcal{C}_u} \sum_{D_j \in \mathcal{D}_u} I_{i,DT_j}^u \leq \xi_u$. Finally, the total transmit power of each channel $S_i \in \mathcal{S}_u$ of each UAV u must not exceed P and must satisfy the total power constraint on each channel, andthere is $P_i^u + \sum_{D_i \in \mathcal{D}_u} I_{i,DT_j}^u P_{i,D_j}^u \leq P$.

Based on Equation (11), the signal-to-interference-plus-noise ratio (SINR) at UAV u corresponding to the received signal at C_i is

$$r_i^u = \frac{P_i^u |g_i^u|^2}{\sum_{D_j \in \mathcal{D}_u} I_{i,DT_j}^u P_{i,D_j}^u \left| g_{i,DT_j}^u \right|^2 + \sigma^2} \tag{14}$$

D2D receiver $DR_{j,n} (n = 1, 2)$ receives the signal on subchannel $S_i \in \mathcal{S}_u$, denoted as

$$y_{i,DR_{j,n}}^u = \sqrt{P_{i,D_j}^u} x_{i,DT_j}^u g_{i,DT_j,DR_{j,n}}^u + \sqrt{P_i^u} x_i^u g_{i,DR_{j,n}}^u + \eta \tag{15}$$

Based on the principle of NOMA, the intra-group interference will be eliminated at the receiver side according to the SIC technique, where the strong users are assigned low power and the weak users are assigned high power in the superposition coding so that the weak users are less interfered with by the strong users and can demodulate their own signals on their own, whereas the strong users need to remove the signals of the weak users and then demodulate their own signals through the SIC technique. To simplify the illustration, assume that receiver $DR_{j,2}$ has worse channel conditions than receiver $DR_{j,1}$, i.e., receiver $DR_{j,1}$ is a weak user and can demodulate its own signal, and in order for $DR_{j,1}$ to remove $DR_{j,2}$ signal to decode its own signal, the following constraints must be satisfied:

$$\frac{\left| g_{i,DT_j,DR_{j,1}}^u \right|^2 P_{i,D_j}^u}{P_i^u \left| g_{i,DR_{j,1}}^u \right|^2 + \sigma^2} \geq \frac{\left| g_{i,DT_j,DR_{j,2}}^u \right|^2 P_{i,D_j}^u}{P_i^u \left| g_{i,DR_{j,2}}^u \right|^2 + \sigma^2} \tag{16}$$

After simplification, the inequality can be transformed into

$$S^u_{i,D_j} = \left|g^u_{i,DT_j,DR_{j,1}}\right|^2 \left(P^u_i \left|g^u_{i,DR_{j,2}}\right|^2 + \sigma^2\right) - \left|g^u_{i,DT_j,DR_{j,2}}\right|^2 \left(P^u_i \left|g^u_{i,DR_{j,1}}\right|^2 + \sigma^2\right) \geq 0 \tag{17}$$

Thus, based on Equation (15) and the NOMA principle, the SINR of receiver $DR_{j,2}$ in a D2D cluster can be expressed as

$$r^u_{i,DR_{j,2}} = \frac{\left|g^u_{i,DT_j,DR_{j,2}}\right|^2 P^u_{i,D_j} a^u_{i,DR_{j,2}}}{\left|g^u_{i,DT_j,DR_{j,2}}\right|^2 P^u_{i,D_j} a^u_{i,DR_{j,1}} + P^u_i \left|g^u_{i,DR_{j,2}}\right|^2 + \sigma^2} \tag{18}$$

The receiver $DR_{j,1}$ in the D2D cluster can demodulate its own signal, so the SINR of $DR_{j,1}$ can be expressed as

$$r^u_{i,DR_{j,1}} = \frac{\left|g^u_{i,DT_j,DR_{j,1}}\right|^2 P^u_{i,D_j} a^u_{i,DR_{j,1}}}{P^u_i \left|g^u_{i,DR_{j,1}}\right|^2 + \sigma^2} \tag{19}$$

According to Shannon's theorem, the achievable rate for cellular user C_i using sub-channel $S_i \in \mathcal{S}_u$ is

$$R^u_i = B \log_2(1 + r^u_i) \tag{20}$$

where B denotes the bandwidth. Similarly, the achievable rates of receivers $DR_{j,1}$ and $DR_{j,2}$ in D2D cluster D_j can be expressed as, respectively,

$$R^u_{i,DR_{j,1}} = B \log_2\left(1 + r^u_{i,DR_{j,1}}\right) \tag{21}$$

$$R^u_{i,DR_{j,2}} = B \log_2\left(1 + r^u_{i,DR_{j,2}}\right) \tag{22}$$

According to Equations (21) and (22), the sum rate expression in D2D cluster D_j is

$$\begin{aligned} R^u_{i,D_j} = B \Bigg[& \log_2\left(1 + \frac{\left|g^u_{i,DT_j,DR_{j,1}}\right|^u_{i,D_j} a^u_{i,DR_{j,1}}}{P^u_i \left|g^u_{i,DR_{j,1}}\right|^2 + \sigma^2}\right) + \\ & \log_2\left(1 + \frac{\left|g^u_{i,DT_j,DR_{j,2}}\right|^2 P^u_{i,D_j} a^u_{i,DR_{j,2}}}{\left|g^u_{i,DT_j,DR_{j,2}}\right|^2 P^u_{i,D_j} a^u_{i,DR_{j,1}} + P^u_i \left|g^u_{i,DR_{j,2}}\right|^2 + \sigma^2}\right) \Bigg] \end{aligned} \tag{23}$$

Thus, the whole system throughput can be expressed as

$$R = \sum_{u=1}^{U} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(R^u_i + R^u_{i,D_j}\right) \tag{24}$$

3. Problem Formation

To maximize the system throughput, this study optimizes the channel allocation and trajectory design with power control under spatial constraints, power constraints, channel constraints, and minimum rate constraints. $H = \{x_u(t), y_u(t), h_u(t), u \in \mathcal{U}\}$ denotes the position of the UAV during the service time of $0 \leq t \leq T$. The channel allocation factor is denoted by $I = \{I^u_{i,DT_j}, u \in \mathcal{U}, C_i \in \mathcal{C}_u, D_j \in \mathcal{D}_u\}$. The problem is expressed as: (25a)

$$\max_{H,I,A} R \tag{25a}$$

$$s.t. h_{\min} \leq h_u(t) \leq h_{\max} \tag{25b}$$

$$x_{\min} \leq x_u(t) \leq x_{\max} \tag{25c}$$

$$y_{\min} \leq y_u(t) \leq y_{\max} \tag{25d}$$

$$P^u_i \geq 0, P^u_{i,D_j} \geq 0 \tag{25e}$$

$$P^u_i + \sum_{D_j \in D_u} I^u_{i,DT_j} P^u_{i,D_j} \leq P \tag{25f}$$

$$I^u_{i,DT_j} \in \left\{0,1\right\} \tag{25g}$$

$$\sum_{C_i \in \mathcal{C}_u} \sum_{D_j \in \mathcal{D}_u} I^u_{i,DT_j} \leq \xi_u \tag{25h}$$

$$\sum_{u \in \mathcal{U}} \sum_{C_i \in \mathcal{C}_u} I^u_{i,DT_j} \leq 1 \tag{25i}$$

$$S^u_{i,D_j} \geq 0 \tag{25j}$$

$$r^u_i \geq r^{u,thr}_i \tag{25k}$$

$$r^u_{i,DR_{j,n}} \geq r^{u,thr}_{i,DR_{j,n}} \tag{25l}$$

$$a^u_{i,DR_{j,n}} \geq 0 \tag{25m}$$

$$a^u_{i,DR_{j,1}} + a^u_{i,DR_{j,2}} \leq 1 \tag{25n}$$

where constraints (25b)–(25d) denote constraints on the 3D position of the UAV, i.e., the UAV must be located in the service area before the airspace within the altitude range can be realized in order to avoid collisions between UAVs. Constraints (25e) and (25f) are the transmit power limitations, where constraint (25f) denotes the total power constraints, i.e., the total power of each channel under UAV u must not exceed P. Constraints (25g)–(25i) denote the channel multiplexing constraints. Specifically, constraint (25g) is a binary constraint, which denotes a binary variable indicating whether $D_j \in \mathcal{D}_u$ occupies the $C_i \in \mathcal{C}_u$ channel resource or not; constraint (25h) denotes the maximum number of D2D clusters associated with UAV u's constraint; and constraint (25i) denotes that a D2D cluster is allocated at most one channel resource. Constraint (25j) denotes the SIC successful demodulation constraint. Constraints (25k) and (25l) denote the minimum rate constraints, i.e., cellular users and D2D cluster users should satisfy the minimum transmission rates $r^{u,thr}_i$ and $r^{u,thr}_{i,DR_{j,n}}$, respectively, in order to ensure QoS guarantees for both cellular users' links and D2D clusters' links. Constraints (25m) and (25n) denote the power allocation factor constraints.

From the above optimization objective function, it can be seen that the optimization of this objective function mainly consists of three parts, the channel multiplexing indicator I, the position variable of the UAV H, and the power allocation coefficient A. Here, the optimization problem is a mixed-integer programming problem and the objective function is non-convex, so it is difficult to solve I, H, and A at the same time. So in this paper, this study adopts the joint dynamic hypergraph Multi-Agent Deep Q Network (DH-MDQN) to solve it. The problem is decoupled into two subproblems, solving the channel assignment problem by dynamic hypergraph coloring method and then using the MDQN algorithm to solve the trajectory design and power control problem.

4. Problem Solution

4.1. Subchannel Assignment Based on Dynamic Hypergraph Coloring

Given UAV trajectory design and power control, problem (25a) transforms into

$$\max_{I} \sum_{u=1}^{U} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(R_i^u + R_{i,D_j}^u \right) \tag{26a}$$

$$P_i^u + \sum_{D_j \in D_u} I_{i,DT_j}^u P_{i,D_j}^u \le P \tag{26b}$$

$$I_{i,DT_j}^u \in \left\{ 0, 1 \right\} \tag{26c}$$

$$\sum_{C_i \in \mathcal{C}_u} \sum_{D_j \in \mathcal{D}_u} I_{i,DT_j}^u \le \xi_u \tag{26d}$$

$$\sum_{u \in \mathcal{U}} \sum_{C_i \in \mathcal{C}_u} I_{i,DT_j}^u \le 1 \tag{26e}$$

It can be seen that the above problem still has integer constraints and problem (26a) is still a non-convex problem. In this section, the problem is solved using a dynamic-hypergraph-based coloring method.

In general, edges in a traditional graph can only connect two vertices, and when a cellular user and a D2D user are used as vertices to construct an interference graph A using a traditional graph, only strong interference between communication links can be modeled. Since multiple D2D links are allowed to multiplex the channel resources of a cellular user simultaneously, there may be cumulative interference that affects the link quality, and the conventional graph cannot model such interference. In order to consider both independent and cumulative interference in the system, the literature [28] introduces the hypergraph model, which avoids interference by constructing a hypergraph structure containing nodes and hyperedges (representing interference relationships between multiple nodes), and it colors the nodes to ensure that neighboring nodes use different resources (e.g., subchannels).

A hypergraph $H_g(t)$ is characterized by a bipartite group $\left(V_g(t), E_g(t)\right)$, where $V_g(t) = \left\{ v_1(t), v_2(t), \cdots, v_n(t) \right\}$ is a finite set of vertices and $E_g(t) = \left\{ e_1(t), e_2(t), \cdots, e_m(t) \right\}$ is the set of edges of the hypergraph $H_g(t)$. Each edge $e_i(t)$ is determined by a finite element in $V_g(t)$ and satisfies $e_i(t) \neq \varnothing (i = 1, 2, \cdots, m), 2 \le |e_i(t)| \le n$. Where the edges of $|e_i(t)| = 2$ are called ordinary edges, the edges of $|e_i(t)| > 2$ are collectively called hyperedges, the edges of $|e_i(t)| = Q$ are called Q–element edges, and the hypergraph $H_g(t)$ is also called a $\max |e_i(t)| (i = 1, \cdots, m)$–element hypergraph.

The hypergraph theory is used to solve the subchannel assignment problem in two steps. The first step is to construct a dynamic hypergraph, and the second step is to color the dynamic hypergraph.

4.1.1. Construct the Dynamic Hypergraph

In dynamic hypergraphs, interference relations are categorized into simple and multilateral edges. Simple edges represent two vertices connected by a strong interference source and are suitable for the case of independent interference, where users connected to the simple edges are assigned different subchannels by the dynamic hypergraph coloring process to ensure that the interfering pairs of communications can be clearly identified and separated at any moment; in contrast, polygons represent the cumulative interference among multiple users, where multiple vertices are connected to the same hyperedge to characterize the interference superposition effect among them. The dynamic adjustment of the hyperedges allows the model to respond flexibly to changes in network topology and interference conditions, thus optimizing resource allocation and communication quality.

(1) Dynamic simple edge construction.

When cellular user C_i and D2D cluster D_j use the same resources such that the received signal-to-interference-noise ratio (SIR) obtained by the cellular link is below a certain threshold, a certain interference relationship is considered to exist between them, and a simplex edge will be established between C_i and D_j at this point. Specifically, under time slot t, if the signal-to-interference ratios of C_i and D_j are below their respective thresholds (λ_c and λ_d), i.e., when the cellular link satisfies Equation (27) or the D2D link satisfies Equation (28), it is indicated that there is an interfering relationship between C_i and D_j under the current time slot, which is not suitable for direct resource sharing.

$$\frac{P_i^u \left|g_i^u\right|^2}{P_{i,D_j}^u \left|g_{i,DT_j}^u\right|^2} < \lambda_c \tag{27}$$

$$\frac{P_{i,D_j}^u \left|g_{i,DT_j,DR_{j,n}}^u\right|^2}{P_i^u \left|g_{i,DR_{j,n}}^u\right|^2} < \lambda_d \tag{28}$$

Similarly, if two D2D clusters D_j and D_j' satisfy Equations (29) and (30), simple edges will be formed between them. If there are simple edges between D_j and D_j', then these two D2D clusters will not be able to share channel resources. In addition, since cellular users can only use orthogonal channel resources, cellular links are constructed as simple edges between them to ensure that different resources are allocated.

$$\frac{a_{i,DR_{j,n}}^u \left|g_{i,DT_j,DR_{j,n}}^u\right|^2}{\left|g_{i,DT_j',DR_{j,n}}^u\right|^2} < \lambda_d \tag{29}$$

$$\frac{a_{i,DR_{j,n}'}^u \left|g_{i,DT_j',DR_{j,n}'}^u\right|^2}{\left|g_{i,DT_j,DR_{j,n}'}^u\right|^2} < \lambda_d \tag{30}$$

(2) Dynamic multilateral construction.

A multilateral connects more than two vertices to represent the cumulative interference to the user, considering weaker interference sources other than those that have strong independent interference to the user. From these, a group of interference sources is selected, and the ratio of the signal strength to the cumulative interference strength brought about by their simultaneous multiplexing of the resource is compared with the interference threshold. If it is lower than the threshold, the communication link is connected to this group of interference links as a hyperedge. Under time slot t, for cellular user C_i, if the ratio of received signal strength to cumulative interference is less than the threshold λ_c, as in Equation (31), a multilateral between the cellular user and the D2D cluster is established. Correspondingly, for D2D cluster D_j, if the ratio of received useful signal to cumulative interference for any of the D2D links it contains is less than the threshold λ_d, as in Equation (32), then all the polygons connecting this D2D cluster are created.

$$\frac{P_i^u \left|g_i^u\right|^2}{\sum_{j=1}^{S} P_{i,D_j}^u \left|g_{i,DT_j}^u\right|^2} < \lambda_c \tag{31}$$

$$\frac{P^{u}_{i,D_j} a^{u}_{i,DR_{j,n}} \left| g^{u}_{i,DT_j,DR_{j,n}} \right|^2}{P^{u}_{i} \left| g^{u}_{i,DR_{j,n}} \right|^2 + \sum_{j' \neq j, j'=1}^{S} P^{u}_{i,D_j} \left| g^{u}_{i,DT'_j,DR_{j,n}} \right|^2} < \lambda_d \tag{32}$$

where S denotes the total number of D2D clusters transmitted over the shared subchannel.

4.1.2. Dynamic Hypergraph Coloring

Hypergraph coloring is an NP-hard problem, and traditional heuristic algorithms have high time complexity. In order to solve the hypergraph coloring problem, inspired by refs. [29,30], this study firstly transforms hypergraphs into directed graphs to achieve efficient dynamic coloring, as shown in Definition 1. Specifically, vertex-based "hyperdegree" (mdeg) and unique ID (id) properties are adopted, where $\mathrm{mdeg(w)}$ is defined as the number of hyperedges connected to vertex w to measure the connection strength of vertices in the hypergraph, and $\mathrm{id(w)}$ is the unique vertex ID in the hypergraph $H_g(t)$.

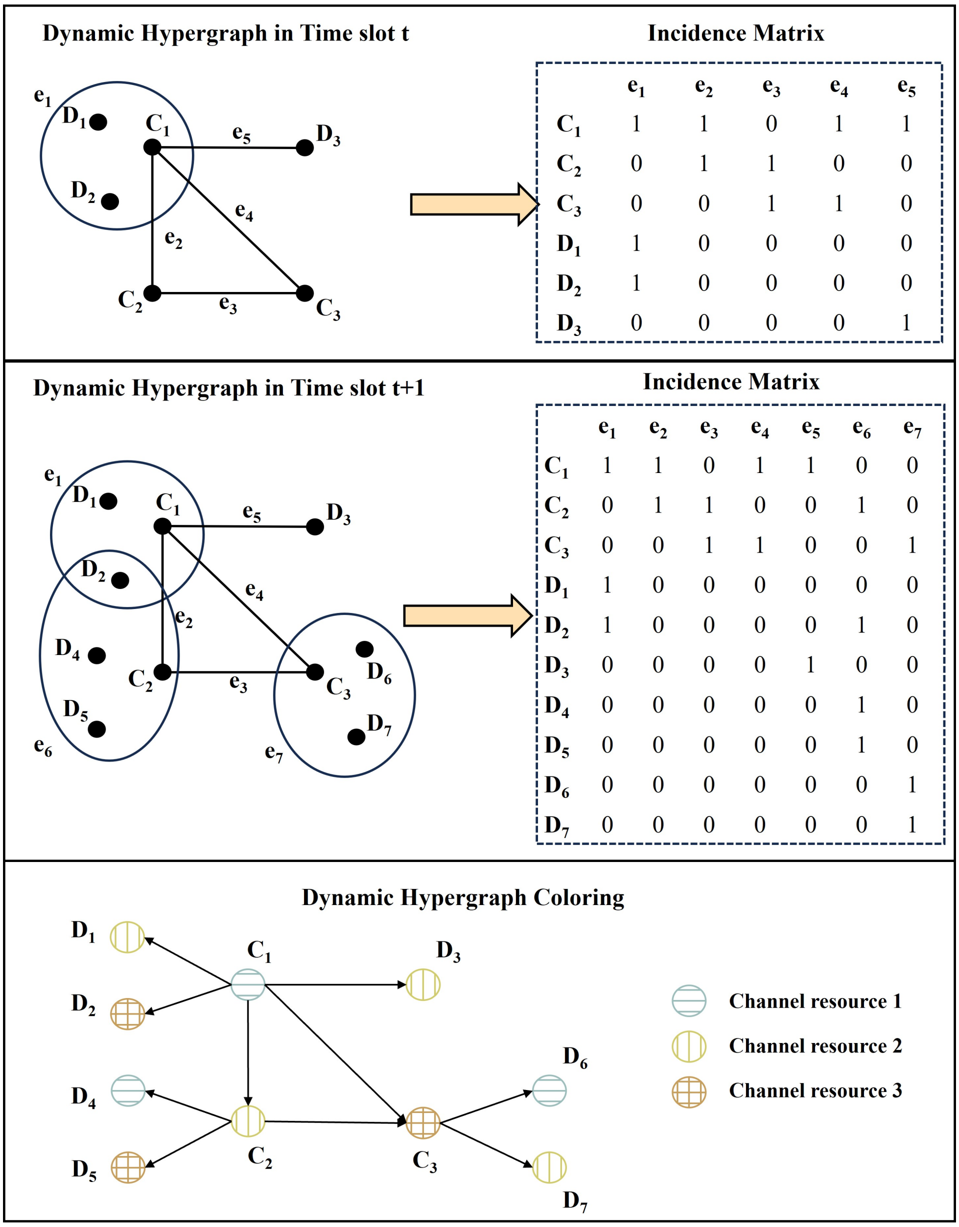
Definition 1. *For a dynamic hypergraph $H_g(t)$ and two vertices $v_i(t)$ and $v_j(t)$, we define $v_i(t) \lhd v_j(t)$, i.e., satisfying (1) $mdeg(v_i) \geq mdeg(v_j)$; (2) $mdeg(v_i) = mdeg(v_j)$, $\mathrm{id}(v_i) < \mathrm{id}(v_j)$.*

In previous studies, researchers used an algorithm based on greedy hypergraph coloring [28] for optimizing subchannel allocation to users in cellular networks. However, this method requires recoloring all vertices each time the hypergraph structure changes, resulting in high computational overhead and insufficient stability. To solve this problem, this study proposes a dynamic hypergraph coloring algorithm, which combines Definitions 2 and 3 to improve the efficiency and robustness of subchannel allocation by dynamically adjusting the coloring rules to effectively respond to dynamic changes in the hypergraph structure.

Definition 2 (hypergraph-based oriented coloring graph (OGC) [30])**.** *For a dynamic hypergraph $H_g(t) = \left(V_g(t), E_g(t)\right)$, its oriented coloring graph $H_o(t) = (V^*(t), E^*(t))$ is a directed acyclic graph. In this graph, for any two vertices $v_i(t)$ and $v_j(t)$, a directed edge pointing from $v_i(t)$ and $v_j(t)$, denoted $\langle v_i(t), v_j(t) \rangle$, is drawn in if $v_i(t) \lhd v_j(t)$ is satisfied.*

Definition 3 (OGC coloring [29])**.** *For a directed acyclic graph $H_o(t) = (V^*(t), E^*(t))$, the coloring function f_o requires that for any directed edge $\langle v_i(t), v_j(t) \rangle$, the associated two vertices $v_i(t)$ and $v_j(t)$ must have different colors.*

Examples of dynamic hypergraphs and correlation matrices as well as dynamic graph coloring for different time slots are given in Figure 2. At time slot t, we have three cellular users and three D2D clusters denoted as $\{C_1, C_2, C_3\}$ and $\{D_1, D_2, D_3\}$. Simple edges $\{e_2, e_3, e_4\}$ are used between the cellular users to ensure that they transmit on different subchannels, and simple edge e_5 connects cellular user C_1 and D2D cluster D_3. In addition, cellular user C_1 and D2D cluster D_1 form a multilateral with D_2. Due to the motion of the UAV, four new D2D clusters and two new polygons are added at the time slot $t + 1$, and their corresponding association matrices are shown in the right part of Figure 2. Finally the results of the steps of dynamic hypergraph coloring are shown according to Definitions 1–3, where first the hypergraph at the time slot is converted into a directed graph according to Definition 1, e.g., the hyperedge e_5 is a directed edge from C_1 to D_3 because of $mdeg(C_1) \geq mdeg(D_3)$, and so on. After the directed graph conversion is completed, the graph is colored according to Definitions 2 and 3.



	e_1	e_2	e_3	e_4	e_5
C_1	1	1	0	1	1
C_2	0	1	1	0	0
C_3	0	0	1	1	0
D_1	1	0	0	0	0
D_2	1	0	0	0	0
D_3	0	0	0	0	1

	e_1	e_2	e_3	e_4	e_5	e_6	e_7
C_1	1	1	0	1	1	0	0
C_2	0	1	1	0	0	1	0
C_3	0	0	1	1	0	0	1
D_1	1	0	0	0	0	0	0
D_2	1	0	0	0	0	1	0
D_3	0	0	0	0	1	0	0
D_4	0	0	0	0	0	1	0
D_5	0	0	0	0	0	1	0
D_6	0	0	0	0	0	0	1
D_7	0	0	0	0	0	0	1

Figure 2. Schematic of subchannel assignment using dynamic hypergraph coloring.

Based on the above analysis, the subchannel assignment algorithm based on dynamic hypergraph coloring is shown in Algorithm 1.

Algorithm 1 Subchannel assignment based on dynamic hypergraph coloring.

```
Initialize the UAV, cellular users, and D2D clusters with position information, power allocation factors, and set all vertex candidate color sets to color full sets
**Step 1: Dynamic hypergraph construction**
Construct simple edges $e_m(t)$ between cellular users under time slot $t$
**repeat**
**if** $SIR < \lambda_c$ or $SIR < \lambda_d$ **then**
Under time slot $t$, if Equations (27) and (28) are satisfied, a simple edge $e_m(t)$ is constructed between cellular user $C_i$ and D2D cluster $D_j$
Under time slot $t$, if Equations (29) and (30) are satisfied, a simple edge $e_m(t)$ is constructed between D2D clusters $D_j$ and $D'_j$
**end if**
```

Algorithm 1 *Cont.*

```
until Output the simple edge $e_m(t)$ between $C_i$, $D_j$ and $D_j'$
repeat
if $SIR < \lambda_c$ or $SIR < \lambda_d$ then
Under time slot $t$, if Equation (31) is satisfied, a multilateral $e_m(t)$ is constructed between cellular user $C_i$ and D2D cluster $D_j$
Under time slot $t$, if Equation (32) is satisfied, the construction of a polygon $e_m(t)$ between D2D clusters $D_j$ and $D_j'$ occurs
until Output the multilateral $e_m(t)$ between $C_i$, $D_j$ and $D_j'$
Step 2: Dynamic Hypergraph Coloring
Convert the hypergraph to a directed graph under time slot $t$ using Definition 1
repeat
for all cellular users and D2D clusters in a directed graph do
Coloring cellular users and D2D cluster users using Definition 2 and Definition 3
end for
until All cellular users and D2D cluster users are colored
```

4.2. MDQN-Based Trajectory Design and Power Control

After fixing the subchannel assignment, problem (25a) transforms into

$$\max_{H,I,A} R \tag{33a}$$

$$s.t. h_{\min} \leq h_u(t) \leq h_{\max} \tag{33b}$$

$$x_{\min} \leq x_u(t) \leq x_{\max} \tag{33c}$$

$$y_{\min} \leq y_u(t) \leq y_{\max} \tag{33d}$$

$$P_i^u \geq 0, P_{i,D_j}^u \geq 0 \tag{33e}$$

$$P_i^u + \sum_{D_j \in D_u} I_{i,DT_j}^u P_{i,D_j}^u \leq P \tag{33f}$$

$$S_{i,D_j}^u \geq 0 \tag{33g}$$

$$r_i^u \geq r_i^{u,thr} \tag{33h}$$

$$r_{i,DR_{j,n}}^u \geq r_{i,DR_{j,n}}^{u,thr} \tag{33i}$$

$$a_{i,DR_{j,n}}^u \geq 0 \tag{33j}$$

$$a_{i,DR_{j,1}}^u + a_{i,DR_{j,2}}^u \leq 1 \tag{33k}$$

Similarly, it can be seen that problem (33a) is still a non-convex problem, and traditional optimization algorithms such as exhaustive search, the branch-and-pricing method, and linear programming may be affected by dimensional catastrophe when dealing with optimization problems, leading to problems such as high computational cost and huge search space. Furthermore, intelligent algorithms such as genetic algorithms, particle swarm algorithms, and annealing simulation algorithms, despite being suitable for solving complex optimization problems, suffer from the problem of falling into local optima. Therefore, in order to be able to obtain the optimal solution of the optimization problem, deep reinforcement learning is used in this section to solve the problem. Compared with traditional reinforcement learning algorithms, the DQN is able to deal with high-dimensional and continuous-state-space situations by using a deep neural network to approximate the Q-value function. The DQN constructs an experience playback pool by continuously interacting with the dynamic communication environment, which contains information related to the states, actions, rewards, and the next state obtained by the intelligent body's interaction with the environment. Through the approximation ability of neural networks, the DQN is able to learn more complex state–action mapping relationships and gradually improve the strategy to maximize long-term rewards during the training process. In order to solve the non-linear non-convex optimization problem of trajectory design and power control, a Multi-Agent Deep Q Network algorithm is designed in this section, which is divided into two steps: (1) MDP model and (2) MDQN algorithm.

4.2.1. MDP Model

In the NOMA-based UAV-assisted D2D communication model, each UAV is considered as an intelligent body for maximum system performance. The three key elements of the agent u are constructed in detail as follows: (1) State space: The state space should be composed of the 3D spatial position of the UAV and the position information of the cellular users and D2D cluster users. The 3D position $h_u(t)$ of the UAV is an important feature of the current state because the position information of the cellular users and D2D cluster users is difficult to obtain in real time, so according to Equations (5), (6), and (8)–(10), this study adopts the known channel gain among the UAV, the cellular users, and the D2D cluster users to characterize the state of the cellular users and the D2D cluster users. The known channel gains between the three are used to characterize the states of cellular users and D2D cluster users, and the input array $s_u(t)$ of the final state space is expressed as

$$s_u(t) = \{h_u(t), g_i^u(t), g_{DT_j}^u(t), g_{i,DT_j}^u(t), g_{i,DR_{j,n}}^u(t), g_{i,DT_j,DR_{j,n}}^u(t)\} \tag{34}$$

In order for multiple UAVs to share the same neural network, the data in the state space need to be scalarized and normalized. Specifically, there are UAV 3D coordinates $h_u(t)$ which are decomposed into three scalars that serve as independent inputs to the neural network. When a drone is connected to the neural network, its position information is input into the first neuron; when another UAV is accessed, the position information of that drone must also be input into the same neuron, as shown in Figures 3 and 4. Compared with the way that drones are trained separately and independently, the MDQN is able to significantly improve the convergence speed compared with the DQN.

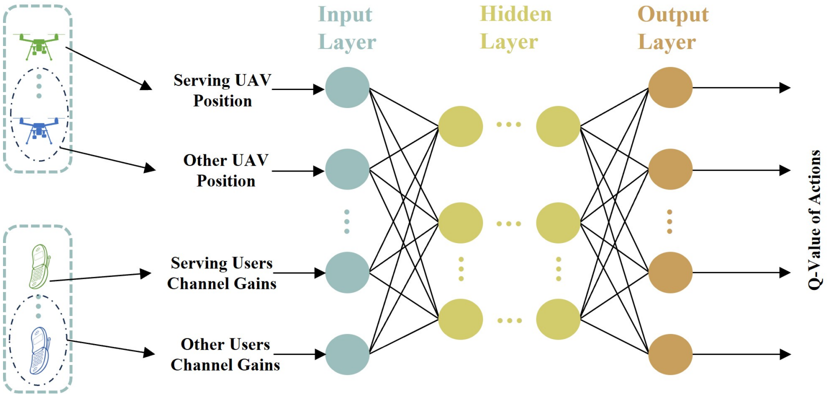


Figure 3. MDQN neural network connection.

(2) Action space: The action space in this section consists of two subsets, the direction of motion of the UAV and the power allocation factor, which are discretized due to the infinite action complexity of the continuous action space and in order to adapt the discrete action output of the MDQN model.

In the direction of UAV movement, the UAV can perform horizontal back-and-forth, vertical up-and-down, and stationary-in-flight operations; in terms of the power factor, the transmission power is preset to a number of fixed slots; and at each movement execution, the UAV selects and maintains a power for its associated D2D cluster until the next movement update.

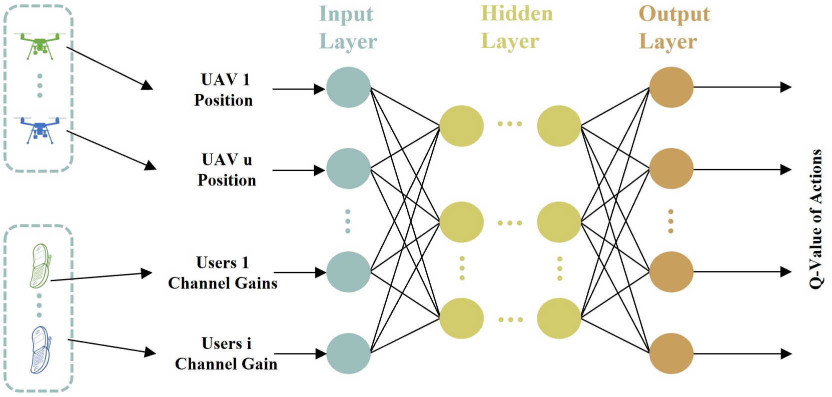


Figure 4. DQN neural network connection.

(3) Reward: As stated in Equation (25a), the objective function of this paper is to maximize the system throughput, so the reward function is set as

$$r_u(t) = R \tag{35}$$

4.2.2. MDQN Algorithm

In this section, the MDQN-based trajectory design and power control allocation algorithm is designed with the training and operation modes shown in Figure 5.The MDQN algorithm uses two neural networks for Q-value estimation and updating: the evaluation network and the target network. It accepts the current state as input and then outputs the Q-value for each action. The parameters of the Q network are updated as training progresses to approximate the true Q-value function. The target-value network is also a deep neural network that is used to estimate the target Q-value, which is the expected maximum cumulative reward. The parameters of the target-value network are copied from the Q network by periodic replication and remain constant during training. After the estimated value network obtains all the Q-values in that state, the target network selects the maximum among the obtained Q-values, whose Q-value is the output of the target network plus the rewards of the samples, and its updated expression is shown in Equation (36).

$$Q(S,A) \leftarrow Q(S,A) + \alpha[R + \beta \max Q(S',A') - Q(S,A)] \tag{36}$$

In order to determine the optimal action for the current state, the UAV inputs the current state information into the evaluation network through the neural network and calculates the reward R by executing and completing the action A. The state information S' is obtained at the next moment, and the system is updated to the next state. Meanwhile, this paper adopts a hybrid selection strategy based on randomness and experience accumulation to enhance the neural network's ability to explore the action selection; specifically, the neural network effectively reduces the sampling correlation of deep reinforcement learning by storing and playing back randomly sampled data samples. In the early stage of training, the UAV relies more on randomly selected actions and continuously enriches the experience pool through interaction with the environment ε_{total}. In order to gradually improve the quality of the strategy in the later stages of training, the actions are selected using a parameter $\delta(t)$ that gradually converges to determinism, and the values of the probability of exploration and the corresponding action selection expressions are shown in (37) and (38) when the action is selected for the N_t-th time.

$$\delta(t) = \delta_{\text{final}} + (\delta_{\text{start}} - \delta_{\text{final}})e^{-N_t/\delta_{\text{decay}}} \tag{37}$$

$$A = \begin{cases} \text{random action}, & \delta(t) \\ \text{argmax}_A\, Q(S, A, w_e), & 1 - \delta(t) \end{cases} \tag{38}$$

where δ_{start} denotes the initial value of the exploration probability, δ_{final} denotes the minimum value of the exploration probability, and δ_{decay} denotes the decay rate of the exploration probability.

The neural network parameters are optimized using the mean square error (MSE) as a loss function. The update formula for the target network and the loss function definition are shown in Equations (39) and (40):

$$q_{target} = R + \beta \max_{A'} Q(S', A', \omega_t) \tag{39}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(q_{target} - Q(S, A, \omega_e)\right)^2 \tag{40}$$

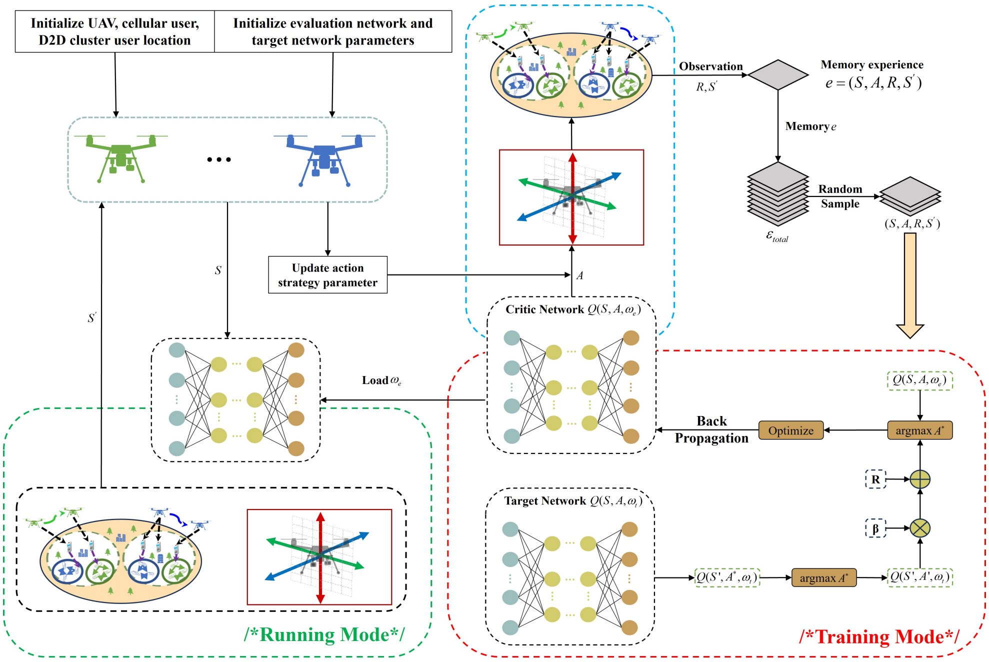


Figure 5. Diagram of MDQN-based trajectory design and power control algorithm.

Based on the above analysis, the MDQN-based trajectory design and power control algorithm is shown in Algorithm 2.

Algorithm 2 MDQN-based trajectory design and power control algorithm.

```
Step 1: Model Training
Initialize the evaluation network parameter $\omega_e$ and the target network parameter $\omega_t$ and make $\omega_t = \omega_e$
Initialize the experience replay pool $e$
for each episode do
    Initialize UAV, cellular users, and D2D cluster locations
    for each step $t$ do
        Update the action strategy parameters $\delta(t)$ according to Equation (37)
        for each UAV do
            Generate the state space $S$ according to Equation (34)
            Determine the action $A$ according to Equations (37) and (38)
```

Algorithm 2 *Cont.*

```
Perform action $A$, observe reward $R$ and next state $S'$
Store experience $e = (S, A, R, S^{'})$ to experience pool $\mathcal{E}_{total}$ according to Equations (34) and (35)
Random sampling from the experience pool $\mathcal{E}_{total}$
Calculation of target $q_{target}$ according to Equation (39)
Perform the optimization step according to Equation (40)
if update $\omega_e$ do
$\omega_t \leftarrow \omega_e$
end if
$S \leftarrow S'$
end for
end for
end for
Step 2: Model Run
Load the trained network parameters $\omega_e$
for each episode do
for each UAV do
Generate the state space $S$ according to Equation (34)
Determine action $A$ from equation $A = \max_A Q(S, A, \omega_e)$
Executing action $A$, the UAV and the user move and interact with the environment, observing state $S'$ according to Equation (34)
$S \leftarrow S'$
end for
end for
```

4.3. Joint Algorithm Design

Based on the above analysis and solution, the optimization algorithm for joint channel assignment, trajectory design, and power control is designed in this section to solve the problem (25a) in an iterative manner. The NOMA-based rate optimization algorithm for multi-UAV-assisted D2D communication networks can be summarized as Algorithm 3.

Algorithm 3 Joint dynamic hypergraph Multi-Agent Deep Q Network algorithm.

```
Initialize the maximum tolerance error $\xi$, the maximum number of iterations $J$, the subchannel assignment vector ${I_{i,DT_j}^{u}}^{(0)}$, the UAV position vector $\left(x_u^{(0)}, y_u^{(0)}, h_u^{(0)}\right)$, and the power allocation factor ${a_{i,DR_{j,n}}^{u}}^{(0)}$
Initialize the number of iterations $j = 0$
Calculate the objective value $R^{(0)}$ of problem (25a) through $\left\{ {I_{i,DT_j}^{u}}^{(0)}, {x_u}^{(0)}, {y_u}^{(0)}, {h_u}^{(0)}, {a_{i,DR_{j,n}}^{u}}^{(0)} \right\}$
while $\left| R^{(j)} - R^{(j-1)} \right| > \xi$ and $j < J$ do
$j = j + 1$
Algorithm 1 is used to solve problem (26a) through $\left\{ x_u^{(j-1)}, y_u^{(j-1)}, h_u^{(j-1)}, {a_{i,DR_{j,n}}^{u}}^{(j-1)} \right\}$, obtaining the optimal solution $\left\{ {I_{i,DT_j}^{u}}^{(j)} \right\}$
Algorithm 2 is used to solve problem (33a) through $\left\{ {I_{i,DT_j}^{u}}^{(j-1)} \right\}$ to obtain the optimal solution $\left\{ x_u^{(j)}, y_u^{(j)}, h_u^{(j)}, {a_{i,DR_{j,n}}^{u}}^{(j)} \right\}$
Calculate the objective value $R^{(j)}$ of problem (25a) through $\left\{ {I_{i,DT_j}^{u}}^{(j)}, {x_u}^{(j)}, {y_u}^{(j)}, {h_u}^{(j)}, {a_{i,DR_{j,n}}^{u}}^{(j)} \right\}$
end while
```

5. Simulation Experiment and Result Analysis

In order to verify the effectiveness of the Multi-Agent Deep Q Network algorithm for joint dynamic hypergraphs proposed in this paper, a series of simulation experiments are conducted. First, the main simulation parameter settings used in the experiments of this paper are listed in detail. Second, the advantages of the proposed DH-MDQN algorithm in improving the system throughput are verified by comparing it with other benchmark algorithms.

5.1. Simulation Experiment Parameter Setting

Cellular users and D2D cluster users are randomly distributed within the service area, the UAV is initially deployed in the vicinity of the cellular users, and the initial flight altitude is set to 100 m. The neural network architecture contains 3 hidden layers, and each hidden layer contains 256 neurons; the activation function is selected as ReLU; and the loss function is MSE. The training process is optimized by using the Adam optimizer, and the value of δ in the greedy strategy decreases linearly from 0.9 to 0. The simulation parameters are shown in Table 1.

Table 1. Simulation parameter setting.

Simulation Parameters	**Value**
Plane area boundaries	$x_{\min} = y_{\min} = 0$ m, $x_{\max} = y_{\max} = 500$ m
UAV altitude range	$h_{\min} = 20m,\ h_{\max} = 150$ m
Number of UAVs	$U = 3$
Maximum UAV flight speed	$V = 5$ m/s
UAV maximum transmit power	$P = 29$ dBm
Maximum cellular user transmit power	$P^{u}_{i\,\max} = 23$ dBm
Number of cellular users	$M = 15$
D2D cluster maximum transmit power	$P^{u}_{i,D_j\,\max} = 20$ dBm
Maximum spacing of D2D clusters	$d^{u}_{i,DT_j,DR_{j,n}\,\max} = 50$ m
Number of D2D clusters	$N = 30$
Maximum number of D2D clusters associated with a UAV	$\xi_u = 10$
Carrier frequency	$f_c = 2$ GHz
Bandwidth	$B = 15$ kHz
AWGN power	$\sigma = -100$ dBm/Hz
Path loss coefficient	$\xi = 2$
Threshold	$\lambda_c = \lambda_d = 18$ dBm
Learning rate	0.001
Discount factor	1
Experience replay pool	10,000 samples
Batch size	128 samples
Optimizer	Adam
Greed coefficient	0–0.9

5.2. Analysis of Simulation Results

In order to comprehensively evaluate the performance of the DH-MDQN algorithm proposed in this paper, the following four different benchmark algorithms are selected for comparison experiments:

(1) Joint dynamic hypergraph Deep Q Network algorithm (DH-DQN): this algorithm uses a dynamic hypergraph to solve the subchannel assignment problem and a DQN algorithm to solve the trajectory planning and power control problem.

(2) Joint graph–theoretic Multi-Agent Deep Q Network algorithm (G-MDQN): this algorithm employs graph theory to solve the subchannel assignment problem and the MDQN algorithm to solve the trajectory planning and power control problem.

(3) Multi-Agent Deep Q Network algorithm (MDQN): this algorithm optimizes trajectory planning and power control using only the MDQN, while the subchannel allocation is based on a random assignment strategy.

(4) Dynamic hypergraph algorithm (DH): the algorithm is based on the dynamic hypergraph to realize the optimization of subchannel allocation, and under the constraint of guaranteeing the basic performance of users, the power control is completed by using a fixed trajectory with random generation.

Next in this study, the proposed DH-MDQN algorithm is simulated and validated in several dimensions.

(1) Validation of the effectiveness of the DH-MDQN algorithm.

In order to verify the effectiveness of the proposed algorithm, Figure 6 shows the effect of different learning rates on the DH-MDQN and DH-DQN algorithms. It is ensured that the algorithms achieve fast convergence and maintain high stability in the proposed NOMA-based multi-UAV-assisted D2D communication network environment.

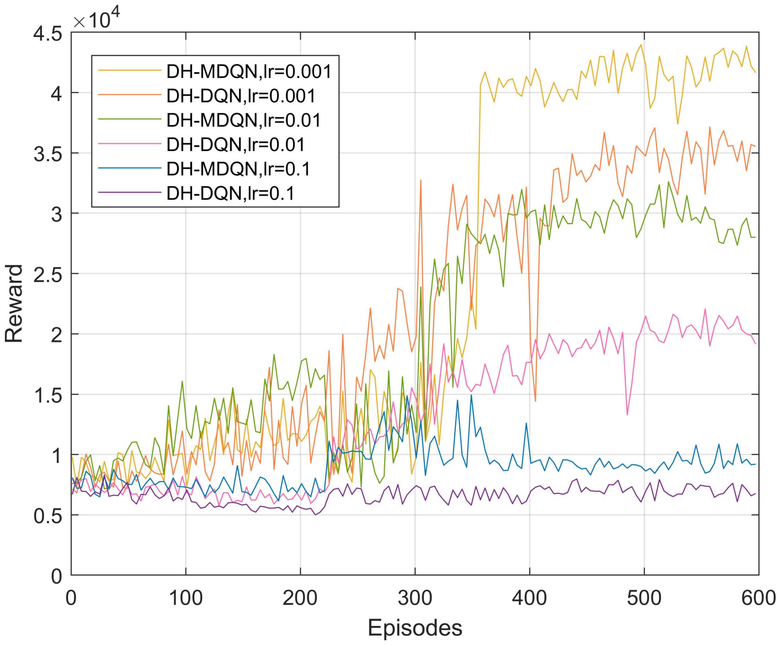


Figure 6. Convergence of DH-MDQN and DH-DQN algorithms with different learning rates.

From the figure, it can be seen that different learning rates affect the learning performance and convergence efficiency of the two algorithms. The cumulative reward values of the algorithms at different learning rates grow slowly at the beginning of training due to the random initialization of the neural network parameters, a phenomenon that can be attributed to the high percentage of random actions and the fact that the replay memory buffer is not completely filled, which results in a model that has not yet entered into the effective training phase. Both DH-MDQN and DH-DQN show the best learning performance when the learning rate is 0.001. The DH-MDQN algorithm converges faster and stabilizes at about 360 episodes, while the DH-DQN algorithm converges with a relative lag and does not stabilize until close to 430 episodes. The main reason for this difference is that the improved mechanism of the DH-MDQN algorithm enables multiple UAVs to share the same neural network, which can adjust the network parameters faster and reach the optimal strategy in a shorter time. In contrast, when the learning rate is 0.1, the algorithm exhibits significant oscillations and instability, resulting in the reward value always remaining at a low level. This is because too high a learning rate will lead to too large a parameter adjustment step, and the neural network will be prone to deviate from the optimal solution or even fail to converge during the optimization process. In contrast, when the learning rate is 0.01, although the convergence stability of the model is improved, the too-large step size still leads to slower convergence, and the model performance fails to reach the optimal performance. Taken together, a learning rate of 0.001 can ensure the convergence speed of the model as well as stabilize the parameter adjustment process to maximize the reward value. Among them, the DH-MDQN algorithm outperforms DH-DQN, mainly due to its efficient utilization of empirical samples and stronger exploration ability, which enables it to achieve a higher cumulative reward value in a shorter training cycle.

(2) Comparative performance analysis of DH-MDQN algorithm.

Figure 7 shows the variation of system throughput of different algorithms (DH-MDQN, DH-DQN, G-MDQN, MDQN, and DH) as the number of training sets increases.

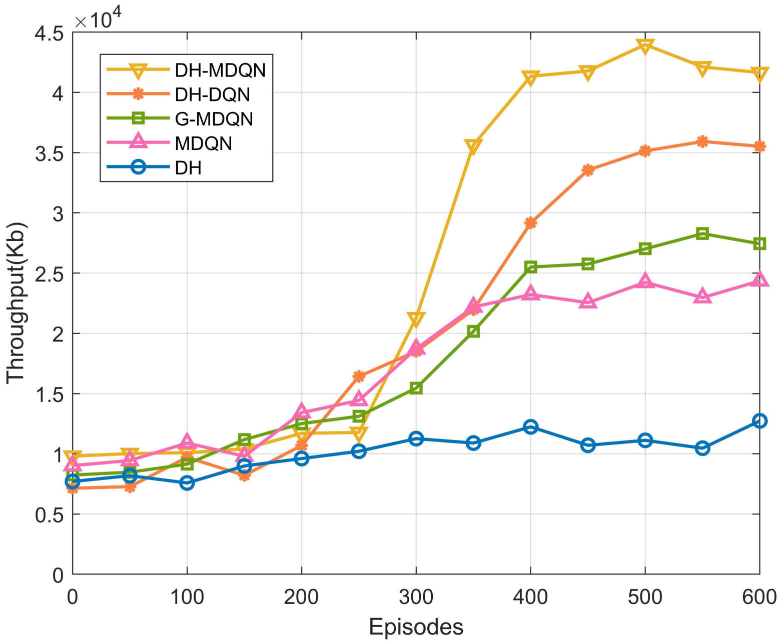


Figure 7. Comparison of system throughput with different algorithms.

From the results, it can be seen that the DH-MDQN algorithm performs the best, the throughput increases rapidly with the increase in the number of training sets and stabilizes after about 360 training sessions, and the final system throughput is about 42,000 Kb. Compared with the DH-DQN algorithm, the DH-MDQN algorithm significantly improves the system throughput. Under training stabilization, the DH-MDQN algorithm improves the system throughput by about 14% compared to the DH-DQN algorithm. In contrast, the system throughput of G-MDQN is lower than that of DH-MDQN, mainly due to the fact that the dynamic hypergraph is able to comprehensively model the cumulative disturbances of the users, which allows for a more fine-grained and flexible resource allocation. In addition, its dynamism permits real-time updating of the network structure to adapt to environmental changes and improve the efficiency and fairness of subchannel allocation. In contrast, traditional graph theory can only establish point-to-point binary relationships, which makes it difficult to capture the complex resource competition and collaboration among multiple users, and thus slightly lacks in allocation flexibility and efficiency. MDQN performs even lower, and its throughput, although improved at the beginning of the training period, has a low final convergence value. This is due to the fact that MDQN only considers power allocation and trajectory planning and does not cover the critical factor of subchannel allocation. Neglecting subchannel allocation leads to less efficient utilization of spectrum resources, which significantly limits the improvement of system throughput. The DH algorithm performs the worst, with its throughput barely improving throughout the training process. This is because DH only focuses on subchannel allocation without involving power control and trajectory planning optimization. Although the subchannel allocation can improve the spectrum utilization efficiency, in the absence of power control and trajectory planning, the overall resource scheduling capability of the system is severely limited and cannot adapt to the demands of the complex dynamic environment, which leads to the throughput always remaining at a low level.

Figure 8 shows the trend of system throughput with the number of D2D clusters under different algorithms.

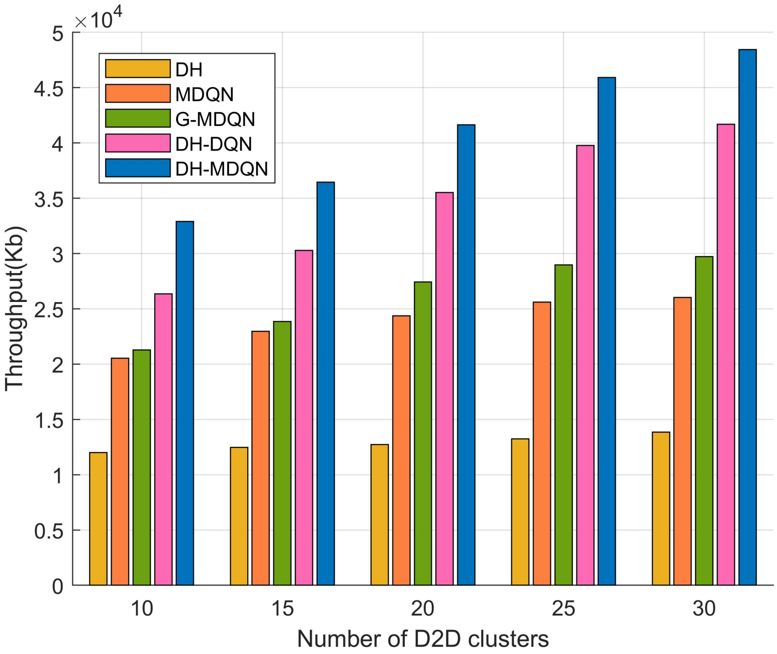


Figure 8. Comparison of system throughput with different numbers of D2D clusters.

It can be seen that the system throughput of all algorithms shows a growing trend as the number of D2D clusters gradually increases from 10 to 30, but the growth rate and the final performance vary depending on the algorithms. Among them, the DH-MDQN algorithm shows the highest system throughput at all D2D cluster sizes; in particular, when the number of D2D clusters is 20, the throughput increase is the most significant. When the number of D2D clusters is small, the allocation and optimization of system resources is relatively simple, and the throughput grows faster at the initial stage; however, as the number of D2D clusters increases further, the interference in the system also increases, resulting in a flattening out of the throughput increase. For different numbers of D2D clusters, the system throughput of the DH-MDQN algorithm is about 13% higher than that of the DH-DQN algorithm on average; in particular, when the number of D2D clusters is 30, the DH-MDQN algorithm produces a system throughput of about 48,400 Kb. This further verifies the significant advantage of the DH-MDQN algorithm in enhancing the system throughput at different numbers of D2D clusters.

Figure 9 demonstrates the comparison of system throughput at different D2D cluster communication distances.

As the communication distance increases, all system throughput shows a certain trend. It can be observed that when the communication distance of the D2D cluster is small, the system throughput is higher, which is because the shorter communication distance can reduce the signal attenuation and interference, which makes the resource allocation more efficient and the transmission rate better guaranteed. As the communication distance increases, the signal attenuation and interference gradually increase, resulting in a decrease in system throughput. In addition, the performance of different algorithms at different communication distances varies. For different D2D cluster communication ranges, the system throughput of the DH-MDQN algorithm is about 15% higher than that of the DH-DQN algorithm on average; in particular, when the D2D cluster spacing is 10 m, the DH-MDQN algorithm produces a system throughput of about 42,900 Kb, which suggests that the DH-MDQN is able to better balance the resource optimization and the interference management, especially in medium-distance scenarios, and the advantages are fully reflected. While other algorithms, such as G-MDQN and MDQN, can achieve better performance at small distances, the decreases in throughput and rate are more obvious at larger distances, which indicates that they are less capable of handling large-scale interference.

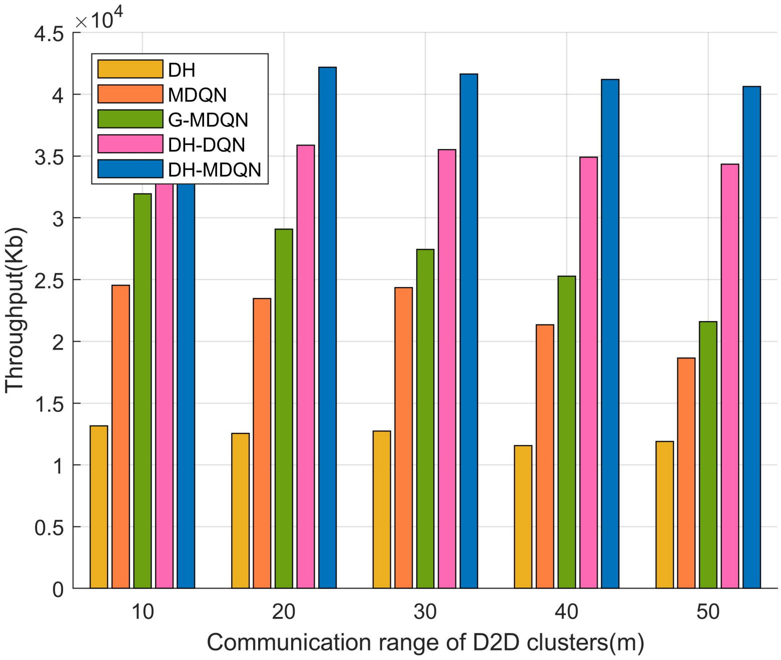


Figure 9. Comparison of system throughput at different D2D cluster communication distances.

(3) Comparative analysis of multiple access under DH-MDQN algorithm.

In order to illustrate the impact of NOMA as well as power allocation on system throughput, Figure 10 shows the comparison of the system throughput under both algorithms, NOMA, OMA, and NOMA without decoding order constraints, for both DH-MDQN and DH-DQN.

It can be seen that the throughput of each algorithm gradually improves as the number of training times increases, indicating that the model gradually converges in continuous training. The simulation results illustrate that removing decoding order has a significant effect on the system throughput in the case of NOMA. Compared to the case without decoding order constraints, the system throughput improves by 34% on average. This shows that decoding order plays a key role in NOMA. The absence of decoding order constraints may lead to less efficient inter-user interference handling, which further affects the performance of the decoding process and thus reduces the system throughput. In addition, the throughput of NOMA is consistently higher than that of OMA. This is due to the fact that NOMA supports more users to transmit simultaneously through spectrum multiplexing, which significantly improves the spectrum utilization, and thus the system throughput is much higher than that of OMA. In contrast, OMA is limited by the adoption of an orthogonal resource allocation strategy, where the resources are allocated independently among different users, which restricts the resource utilization efficiency of the system and results in a limited growth of its throughput.

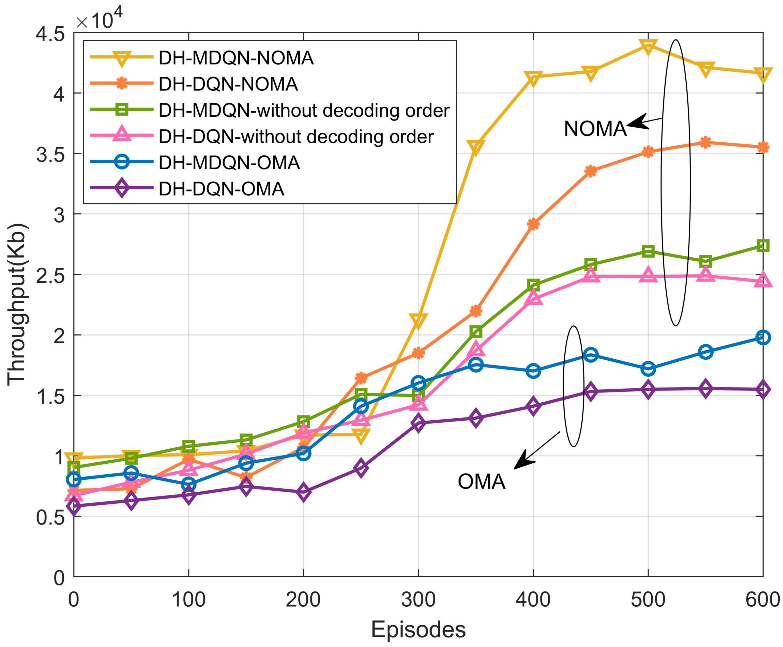


Figure 10. System throughput comparison between NOMA and OMA with different number of sets.

(4) Comparative analysis of D2D under DH-MDQN algorithm.

Figure 11 shows the impact of considering different D2D clusters on the system throughput under both the DH-MDQN and DH-DQN algorithms. Three specific cases are compared: (1) where D2D clusters are present and power allocation is performed, (2) where D2D clusters are present but no power allocation is performed, and (3) where there are no D2D clusters.

The simulation results show that under both algorithms, the system throughput is significantly higher when D2D clusters are present and power allocation is performed than the other two cases. Comparing the two cases without power allocation and without D2D clusters, the system throughput is improved by an average of 54% and 67%, respectively. This is due to the fact that D2D communication can significantly reduce the transmission delay and signal interference and improve the spectral efficiency through direct transmission between users in close proximity. In the case where D2D clusters exist but there is no power allocation, although D2D communication can still improve the throughput to a certain extent, the lack of power optimization may lead to an increase in interference, thus affecting the overall throughput improvement. The system throughput is generally higher in the presence of D2D clusters compared to the case without D2D clusters, which further validates the positive effect of D2D communication on system performance. In the absence of D2D clusters, the system relies on the traditional transmission mode between the base station and the user, which results in less efficient utilization of spectrum resources and limited throughput.

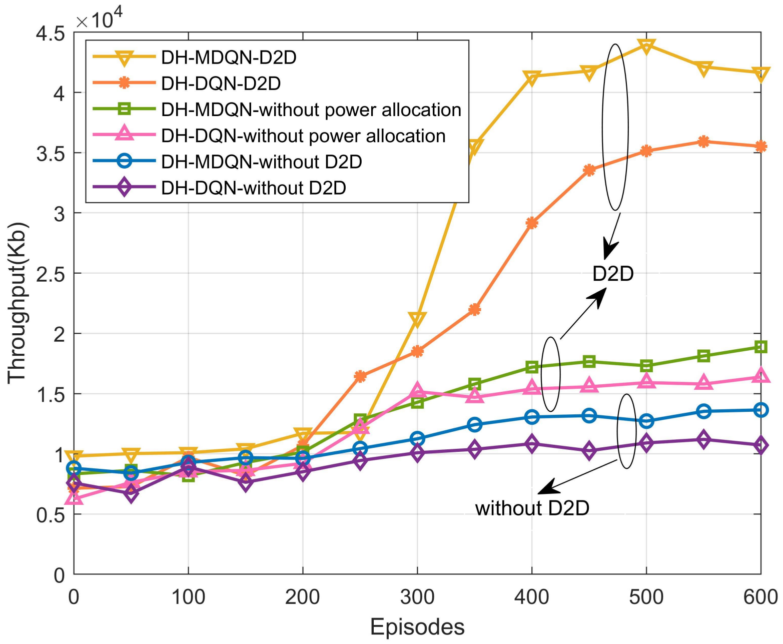


Figure 11. Comparison of system throughput with and without D2D clusters for different numbers of sets.

(5) Comparative analysis of trajectory settings under DH-MDQN algorithm.

Figure 12 shows the flight trajectory of a single UAV in a scenario with six cellular users and six pairs of D2D cluster users. The overall trend of the trajectory shows that the UAV gradually approaches each part of the cellular users and D2D cluster users in the initial stage and effectively avoids interference with all types of users by adjusting its flight path in real time. Initially, as the UAV moves, it continuously adjusts its relative position to the cellular users and D2D clusters to ensure that good communication quality is maintained while avoiding interference. Eventually, the flight trajectory of the UAV converges to the optimal configuration point of the system throughput, which is capable of efficiently scheduling communication resources among multiple users, reducing resource conflicts, and ensuring the efficiency and stability of the communication link. Simulation results show that the DH-MDQN algorithm proposed in this paper is able to realize efficient trajectory planning in complex multi-UAV and multi-D2D environments, which verifies the effectiveness and practicality of the proposed algorithm in improving system throughput.

Figure 13 shows the flight trajectories of multiple UAVs in 3D space and the locations of ground D2D clusters, from which it can be seen that each UAV always maintains a safe and collision-free distance from each other during flight. Thanks to the joint dynamic hypergraph modeling and multi-intelligent body reinforcement learning decision-making mechanism of the proposed DH-MDQN algorithm, each UAV can not only intelligently provide efficient communication services for the ground users and D2D clusters but also flexibly adjust the flight paths in the three-dimensional environment, which not only improves the throughput of the system but also avoids path conflicts, realizing the collaborative optimization and efficient coverage in the dynamic and complex communication scenarios, which verifies that the proposed algorithm in this paper can be used in future wireless networks. The feasibility and superiority of the algorithm proposed in this paper for the cooperative optimization of multi-UAV and D2D communication in future wireless networks is demonstrated.

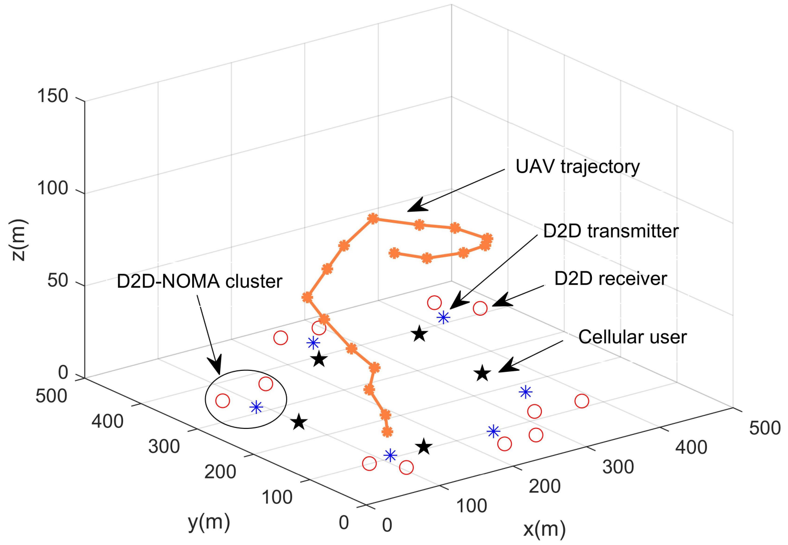


Figure 12. Single-drone 3D flight trajectory map.

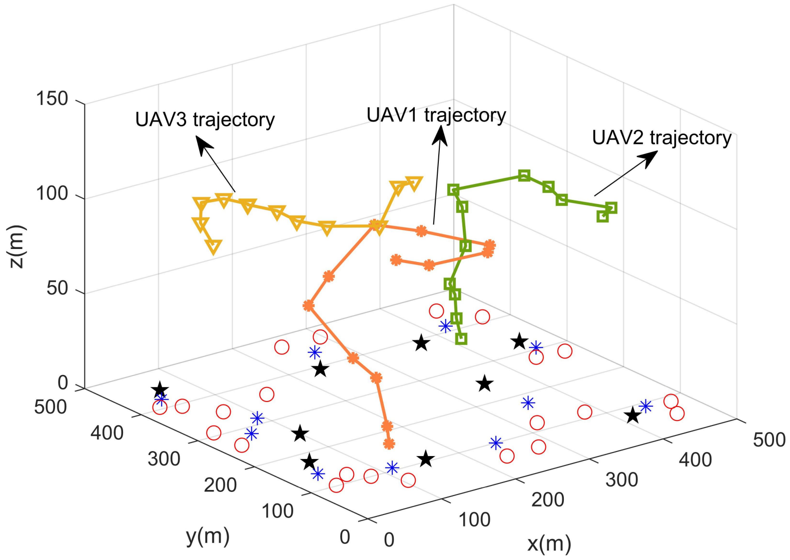


Figure 13. Multi-drone 3D flight trajectory map.

To further illustrate the effect of trajectories on system throughput, Figure 14 shows the throughput versus the number of training sets under different trajectory designs.

It can be seen that the system throughput under 3D trajectory optimization is better than that of 2D trajectory optimization, randomly deployed trajectories, and circular trajectories. This result indicates that 3D trajectory optimization is able to adapt more flexibly to the complexities in the environment, thus improving the spectrum utilization and throughput of the system. Compared with other trajectory design approaches, 2D trajectory optimization improves throughput to some extent, but it fails to take full advantage of the flexibility of the spatial dimension due to its restriction to the plane, resulting in a relatively small increase in throughput, and the average increase in system throughput for 3D trajectories over 2D trajectories is 27%. Random deployment trajectories and circular trajectories do not need to consider trajectory optimization, so convergence can be reached very quickly, and the throughput performance of both of them is smoother and lower, mainly due to the fact that these two trajectory designs fail to be effectively optimized for the distribution of users, which further proves the validity of the three-dimensional trajectories of the researched UAVs.

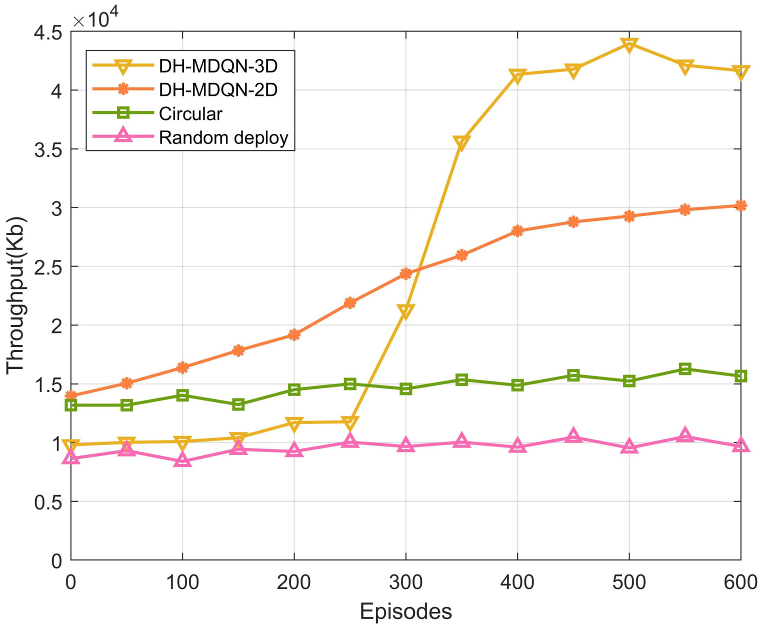


Figure 14. Comparison of system throughput with different trajectory designs.

Figure 15 shows the system throughput for different numbers of drones. In general, more drones can provide more services to users and a higher system and rate can be obtained. In the single-UAV case, convergence is easily reached. Furthermore, in terms of the increase in system throughput, there is an average increase of 43% in system throughput when increasing from two drones to three drones, while there is an average increase of 18% when increasing from three drones to four drones. This result indicates that from three drones onward, the increase in system throughput gradually decreases as the number of drones increases. This is because when the number of UAVs exceeds a certain threshold, resource and interference constraints in the system gradually become apparent, leading to a gradual decrease in the gains from more UAVs. The simulation results show that three UAVs can improve the system throughput while avoiding the marginal effect caused by too many UAVs.

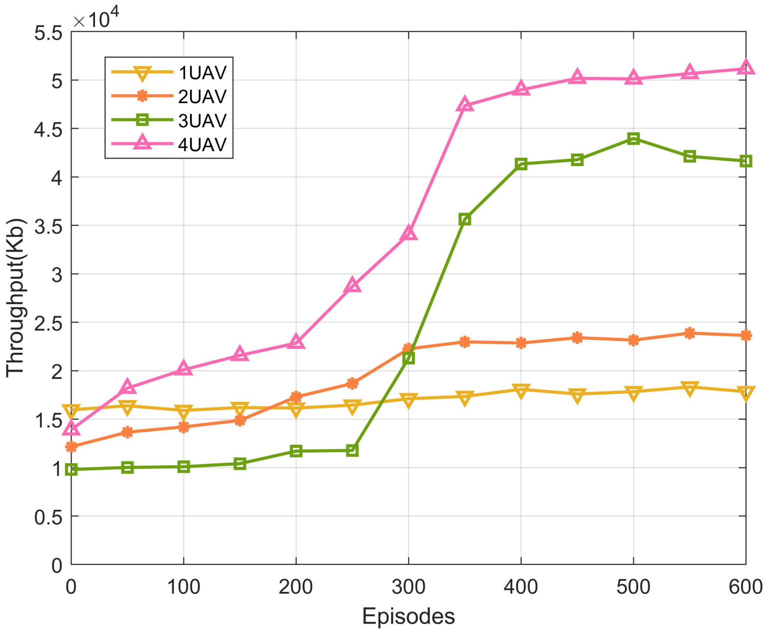


Figure 15. Comparison of system throughput with different numbers of UAVs.

6. Conclusions

In this paper, we discuss a multi-UAV-assisted D2D communication model incorporating NOMA to address the channel allocation, trajectory planning, and power control problems, respectively, and design a DH-MDQN algorithm, in which the channel allocation problem is solved by a dynamic hypergraph and the MDQN algorithm is utilized to solve the trajectory planning and power control problems. Our simulation evaluates the performance of the proposed algorithm through numerical results in terms of the number of D2D clusters, D2D cluster communication spacing, UAV size, trajectory planning, and NOMA decoding order in multiple dimensions. These results also demonstrate the superiority of the UAV-assisted D2D-NOMA framework, and the proposed DH-MDQN algorithm is able to achieve higher system throughput compared to other benchmark algorithms. In future work, the research will be further extended to the sky–ground integrated network system to explore more heterogeneous and dynamic environments in depth in order to meet the practical needs of next-generation wireless networks in terms of high efficiency, high reliability, and strong adaptability.

Author Contributions: Conceptualization, G.C.; methodology, G.W.; validation, G.W.; formal analysis, X.G.; investigation, X.G.; data curation, G.W.; writing—original draft, G.W.; visualization, G.W.; project administration, G.C.; funding acquisition, G.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant No. 61540022, Weapons and Equipment Advance Research Programs No. 6141B021322, No. 6141B012826, and Jilin Provincial Special Project for Research and Development of Industrial Technology No. 2022C047-8.

Data Availability Statement: The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Al-Fuqaha, A.; Guizani, M.; Mohammadi, M.; Aledhari, M.; Ayyash, M. Internet of things: A survey on enabling technologies, protocols, and applications. *IEEE Commun. Surv. Tutor.* **2015**, *17*, 2347–2376. [CrossRef]
2. Islam, S.R.; Avazov, N.; Dobre, O.A.; Kwak, K.S. Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges. *IEEE Commun. Surv. Tutor.* **2016**, *19*, 721–742. [CrossRef]
3. Ding, Z.; Liu, Y.; Choi, J.; Sun, Q.; Elkashlan, M.; Chih-Lin, I.; Poor, H.V. Application of non-orthogonal multiple access in LTE and 5G networks. *IEEE Commun. Mag.* **2017**, *55*, 185–191. [CrossRef]
4. Ding, Z.; Yang, Z.; Fan, P.; Poor, H.V. On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users. *IEEE Signal Process. Lett.* **2014**, *21*, 1501–1505. [CrossRef]
5. Yadav, A.; Quan, C.; Varshney, P.K.; Poor, H.V. On performance comparison of multi-antenna HD-NOMA, SCMA, and PD-NOMA schemes. *IEEE Wirel. Commun. Lett.* **2020**, *10*, 715–719. [CrossRef]
6. Liu, J.; Kato, N.; Ma, J.; Kadowaki, N. Device-to-device communication in LTE-advanced networks: A survey. *IEEE Commun. Surv. Tutor.* **2014**, *17*, 1923–1940. [CrossRef]
7. Qiao, J.; Shen, X.S.; Mark, J.W.; Shen, Q.; He, Y.; Lei, L. Enabling device-to-device communications in millimeter-wave 5G cellular networks. *IEEE Commun. Mag.* **2015**, *53*, 209–215. [CrossRef]
8. Zhao, J.; Liu, Y.; Chai, K.K.; Chen, Y.; Elkashlan, M. Joint subchannel and power allocation for NOMA enhanced D2D communications. *IEEE Trans. Commun.* **2017**, *65*, 5081–5094. [CrossRef]
9. Wang, L.; He, Y.; Chen, B.; Hassan, A.; Wang, D.; Yang, L.; Huang, F. Joint Phase Shift Design and Resource Management for a Non-Orthogonal Multiple Access-Enhanced Internet of Vehicle Assisted by an Intelligent Reflecting Surface-Equipped Unmanned Aerial Vehicle. *Drones* **2024**, *8*, 188. [CrossRef]
10. Zhang, Z.; Ma, Z.; Xiao, M.; Ding, Z.; Fan, P. Full-duplex device-to-device-aided cooperative nonorthogonal multiple access. *IEEE Trans. Veh. Technol.* **2016**, *66*, 4467–4471.

11. Aggarwal, S.; Kumar, N.; Tanwar, S. Blockchain-envisioned UAV communication using 6G networks: Open issues, use cases, and future directions. *IEEE Internet Things J.* **2020**, *8*, 5416–5441. [CrossRef]
12. Pan, G.; Lei, H.; An, J.; Zhang, S.; Alouini, M.S. On the secrecy of UAV systems with linear trajectory. *IEEE Trans. Wirel. Commun.* **2020**, *19*, 6277–6288. [CrossRef]
13. Wang, B.; Zhang, R.; Chen, C.; Cheng, X.; Yang, L.; Li, H.; Jin, Y. Graph-based file dispatching protocol with D2D-enhanced UAV-NOMA communications in large-scale networks. *IEEE Internet Things J.* **2020**, *7*, 8615–8630. [CrossRef]
14. Liu, X.; Yang, B.; Liu, J.; Xian, L.; Jiang, X.; Taleb, T. Sum-Rate Maximization for D2D-Enabled UAV Networks with Seamless Coverage Constraint. *IEEE Internet Things J.* **2024**. [CrossRef]
15. Pan, H.; Liu, Y.; Sun, G.; Wang, P.; Yuen, C. Resource scheduling for UAVs-aided D2D networks: A multi-objective optimization approach. *IEEE Trans. Wirel. Commun.* **2023**. [CrossRef]
16. Zhang, Y.; Hou, X.; Du, H.; Zhang, L.; Du, J.; Men, W. Joint Trajectory and Resource Optimization for UAV and D2D-enabled Heterogeneous Edge Computing Networks. *IEEE Trans. Veh. Technol.* **2024**. [CrossRef]
17. Marani, M.R.; Mirrezaei, S.M.; Mirzavand, R. Joint throughput and coverage maximization for moving users by optimum UAV positioning in the presence of underlaid D2D communications. *AEU-Int. J. Electron. Commun.* **2023**, *161*, 154541. [CrossRef]
18. Tang, R.; Wang, J.; Zhang, Y.; Jiang, F.; Zhang, X.; Du, J. Throughput Maximization in NOMA Enhanced RIS-Assisted Multi-UAV Networks: A Deep Reinforcement Learning Approach. *IEEE Trans. Veh. Technol.* **2024**. [CrossRef]
19. Hosny, R.; Hashima, S.; Hatano, K.; Zaki, R.M.; El Halawany, B.M. UAV trajectory planning in NOMA-aided UAV-mounted RIS networks: A budgeted Multi-armed bandit approach. *J. Phys. Conf. Ser.* **2024**, *2850*, 012008. [CrossRef]
20. Amhaz, A.; Elhattab, M.; Sharafeddine, S.; Assi, C. Uav-assisted cooperative downlink noma: Deployment and resource allocation. *IEEE Trans. Veh. Technol.* **2024**. [CrossRef]
21. Nguyen, T.T.; Tran, M.H.; Tran, X.N. Joint Resource and Trajectory Optimization For Secure UAV-Based Two-way Relay System. *Digit. Signal Process.* **2024**, *153*, 104626. [CrossRef]
22. Jabbari, A.; Khan, H.; Duraibi, S.; Budhiraja, I.; Gupta, S.; Omar, M. Energy Maximization for Wireless Powered Communication Enabled IoT Devices with NOMA Underlaying Solar Powered UAV Using Federated Reinforcement Learning for 6G Networks. *IEEE Trans. Consum. Electron.* **2024**. [CrossRef]
23. Yang, X.; Qin, D.; Liu, J.; Li, Y.; Zhu, Y.; Ma, L. Deep reinforcement learning in NOMA-assisted UAV networks for path selection and resource offloading. *Ad Hoc Netw.* **2023**, *151*, 103285. [CrossRef]
24. Rezwan, S.; Chun, C.; Choi, W. Federated Deep Reinforcement Learning-Based Multi-UAV Navigation for Heterogeneous NOMA Systems. *IEEE Sens. J.* **2023**. [CrossRef]
25. Qin, P.; Fu, Y.; Zhang, J.; Geng, S.; Liu, J.; Zhao, X. DRL-Based Resource Allocation and Trajectory Planning for NOMA-Enabled Multi-UAV Collaborative Caching 6 G Network. *IEEE Trans. Veh. Technol.* **2024**, *73*, 8750–8764. [CrossRef]
26. Docomo, N.T.T. *5G Channel Model for Bands up to100 GHz*; Technical Report; 2016. Available online: https://prepareforchange.net/wp-content/uploads/2018/12/5G_Channel_Model_for_bands_up_to100_GHz2015-12-6.pdf (accessed on 13 December 2024).
27. Liang, L.; Li, G.Y.; Xu, W. Resource allocation for D2D-enabled vehicular communications. *IEEE Trans. Commun.* **2017**, *65*, 3186–3197. [CrossRef]
28. Zhang, H.; Song, L.; Han, Z. Radio resource allocation for device-to-device underlay communication using hypergraph theory. *IEEE Trans. Wirel. Commun.* **2016**, *15*, 4852–4861. [CrossRef]
29. Yuan, L.; Qin, L.; Lin, X.; Chang, L.; Zhang, W. Effective and efficient dynamic graph coloring. *Proc. VLDB Endow.* **2017**, *11*, 338–351. [CrossRef]
30. Wang, B.; Sun, Y.; Sun, Z.; Nguyen, L.D.; Duong, T.Q. UAV-assisted emergency communications in social IoT: A dynamic hypergraph coloring approach. *IEEE Internet Things J.* **2020**, *7*, 7663–7677. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.