

RESEARCH

Open Access



# Unmanned aerial vehicle-assisted wideband cognitive radio network based on DDQN-SAC

Leibing Yan<sup>1\*</sup>, Yiqing Cai<sup>1</sup> and Hui Wei<sup>1</sup>

\*Correspondence:  
leibingy@hait.edu.cn

<sup>1</sup> School of Electronic  
Information Engineering,  
Henan Institute of Technology,  
XinXiang 453003, China

## Abstract

Cognitive radio (CR) systems have emerged as effective tools for improving spectrum efficiency and meeting the growing demands of communication. This study focuses on a flexible CR system based on opportunistic spectrum access technology, which enables secondary networks to efficiently utilize unoccupied spectrum resources for information transmission by actively sensing the spectrum utilization of primary networks. Specifically, we introduce unmanned aerial vehicles (UAV) technology into the CR system to further enhance its flexibility and adaptability, which enables the transmission efficiency of low-altitude UAV networks. In this CR system, UAVs are employed for more flexible spectrum management. The objective of this research is to maximize the average achievable rate of SUs by jointly optimizing the trajectories of secondary UAV, the trajectories of primary UAV, the beamforming of secondary UAV, subchannel allocation and sensing time. To achieve this goal, we employ deep reinforcement learning (DRL) algorithms to optimize these variables. Compared to traditional optimization algorithms, DRL algorithms not only have lower computational complexity but also achieve faster convergence. To address the mixed-action space problem, we propose a Dueling DQN-Soft Actor Critic algorithm. Simulation results demonstrate that the proposed approach in this paper significantly enhances the performance of the CR system compared to traditional baseline schemes. This is manifested in higher spectrum efficiency and data transmission rates, while minimizing interference with the primary network. This innovative research combines drone technology and DRL algorithms, bringing new opportunities and challenges to the future development of cognitive communication systems.

**Keywords:** Cognitive radio (CR), Opportunistic spectrum access (OSA), Unmanned aerial vehicles (UAV), Target location, Wireless sensor network

## 1 Introduction

With the continuous advancement of communication technology, expectations for the next-generation mobile communication system, 6 G, are on the rise. 6 G is seen as the pinnacle of technological revolution, set to usher in higher data speeds, lower latency, and broader connectivity, reshaping various industries [1, 2] However, the widespread adoption and successful deployment of 6 G technology face a significant challenge—spectrum scarcity. Spectrum scarcity has long been a challenging issue in the field of communication. With the explosive growth of wireless communication devices and the

continuous emergence of mobile applications, existing spectrum resources have gradually become scarce, making it difficult to meet the growing communication demands. Therefore, the quest for innovative solutions to address the problem of spectrum scarcity has become an urgent task in the field of communication.

Cognitive radio (CR) technology has already emerged as a key technology for addressing the issue of spectrum scarcity [3–7]. This technology enables secondary networks to intelligently share the spectrum resources of primary networks by real-time spectrum sensing, analysis, and management, thereby enhancing spectral efficiency and meeting communication demands. CR technology commonly operates under three paradigms: spectrum sharing [8–10], sensing-based spectrum sharing [11–13], and opportunistic spectrum access (OSA) [14–17]. In [8], the authors introduced a full-rate cooperative spectrum sharing protocol for bandwidth-efficient cognitive networks and discussed the enhancement of spectral efficiency through spectrum sharing techniques. Additionally, in [9], the authors proposed an intelligent reflecting surface-assisted cognitive radio system. This literature also employs spectrum sharing techniques, aiming to maximize the achievable rates of secondary users (SUs) through the joint optimization of the transmission power of SUs and the reflectivity coefficients of intelligent reflectors while adhering to the maximum interference constraint of primary users. The aforementioned literature effectively enhances spectral efficiency. However, due to the lack of spectrum sensing, the secondary network may introduce excessive interference to the primary users, consequently affecting the information transmission of the primary network. Simultaneously, the primary network can also interfere with the SUs, thereby impacting the achievable rates of SUs.

The above-mentioned techniques, based on sensing-driven spectrum sharing and OSA, are widely employed to address the aforementioned issues. In the literature [11], the authors have employed sensing-driven spectrum sharing techniques to enhance spectrum utilization efficiency. This approach typically consists of a sensing phase and a transmission phase. Initially, the secondary network needs to monitor the subcarrier usage by primary users (PUs). Subsequently, during the transmission phase, the secondary network adjusts its transmission power based on the sensing results. When an idle state is detected on a subcarrier, the secondary network can freely utilize that subcarrier for information transmission, thereby maximizing the achievable data rate for SUs. However, when activity is detected on a subcarrier, the SUs are required to implement interference management measures, such as reducing transmission power, to minimize interference with PUs. On the other hand, in the literature [14], authors have employed OSA techniques. This approach is similar to sensing-driven spectrum sharing but differs in that when the secondary network detects activity on a subcarrier, secondary users will refrain from using that subcarrier. This means that secondary users avoid selecting subcarriers that may interfere with PUs. The simulation results from the aforementioned literature consistently demonstrate that, compared to traditional spectrum sharing methods, these approaches outperform in terms of enhancing spectrum efficiency and reducing interference.

Recently, research on unmanned aerial vehicles (UAV) has also garnered widespread attention. Using UAV instead of traditional base stations offers several advantages [7]. Firstly, UAV can flexibly adjust their position and altitude to accommodate various

communication needs, overcoming signal blind spots and providing improved communication link conditions. This allows for better geographical coverage, thereby enhancing system performance. Secondly, UAV can be rapidly deployed in areas requiring temporary communication coverage without the need for time-consuming traditional infrastructure setup. In [18], the authors proposed a multi-UAV-assisted wireless communication system. They jointly optimized the UAV's trajectory, ground terminal scheduling, and power allocation to maximize the total throughput of ground terminals. In [19], a drone-assisted millimeter-wave communication system was introduced. They jointly optimized the UAV's position and robust hybrid beamforming to maximize the user's minimum achievable rate. In [20], a dual-drone scenario was presented, where one UAV transmitted confidential information to ground users, while the other UAV cooperated in sending artificial noise to interfere with eavesdropping, ensuring data transmission security. Simulation results from these references indicate that UAV can enhance system performance.

Leveraging the advantages of UAV, integrating UAV with cognitive radio systems can further enhance spectral efficiency and system performance [7, 21–25]. In [21], the authors introduced an intelligent reflecting surface (IRS)-assisted drone-enhanced cognitive radio system. They jointly optimized the trajectory of the UAV, the passive beamforming of the IRS, and the power allocation of the UAV to maximize the throughput of secondary users. In [23], the study investigated the impact of UAV-assisted interference in a secure cognitive radio network, validating its influence on security. In [24], research focused on the performance of a cognitive radio-supported UAV network configuration. In this network, UAVs are allowed to communicate with secondary ground terminals in the underlying mode of the licensed spectrum. The objective is to maximize the overall network throughput while meeting constraints related to interference with the primary network and the throughput of each secondary user.

The aforementioned references all employ traditional convex optimization algorithms for variable optimization. However, compared to traditional convex optimization methods, deep reinforcement learning (DRL) algorithms offer significant advantages. Firstly, DRL algorithms excel in addressing complex, high-dimensional, and nonlinear problems. They possess the capability of autonomous learning and strategy improvement without the need for prior knowledge or detailed system models, making them suitable for a wide range of complex real-world problems [7, 10, 26]. Secondly, DRL demonstrates remarkable generalization abilities, enabling it to learn universal strategies that can be applied to multiple tasks, thereby reducing the complexity of modeling specific problems. The algorithm can handle various types of action spaces, including discrete and continuous actions, making it suitable for tasks across different domains. As a result, there is a growing trend in the industry toward the adoption of DRL algorithms [27–30]. The literature [27] investigates an IRS-assisted covert communication system for UAV. To optimize variables in high-dimensional data, the authors propose an optimization algorithm for UAV 3D trajectories and IRS phase shifts based on a Twin-Deep Q Network (TAP-DDQN). In [29], the focus is on optimizing UAV base station trajectories for full-duplex communication. The article introduces a DRL-based method to optimize UAV-BS trajectories, enabling efficient full-duplex communication in disaster scenarios. Another literature [30] explores a multi-user, multiple-input, single-output aerial

IRS-assisted communication system. The problem framework is optimized using the Deep Deterministic Policy Gradient (DDPG) algorithm. To enhance the action decision accuracy of the DDPG algorithm, a mapping function is proposed to mitigate the impact of noise variations on performance during the exploration process.

Based on the considerations of the aforementioned technologies, this paper investigates a drone-assisted broadband cognitive radio network. We jointly optimize the trajectories of the primary UAV, the trajectories of the secondary UAV, the beamforming of the secondary UAV, and the perception time to maximize the total achievable rate of secondary users. We employ the Dueling DQN-Soft Actor Critic (DDQN-SAC) algorithm to address the optimization problem proposed in this study. The main contributions of this article are summarized as follows:

- This is the first article addressing resource allocation in UAV-assisted wideband cognitive radio networks based on OSA technology using DRL algorithms. The joint optimization of the primary UAV's trajectory, the secondary UAV's trajectory, the beamforming for the secondary UAV, subcarrier allocation, and sensing time is aimed at maximizing the total achievable rate of secondary users. The application of UAVs is crucial for enhancing the communication link, reducing signal attenuation, and significantly improving performance in both the sensing and transmission phases.
- Due to the action space containing both continuous and discrete variables, we propose a DDQN-DDPG algorithm. Furthermore, in addition to addressing the challenges of mixed-action spaces, this algorithm significantly reduces computational complexity, enhances training speed, and improves stability compared to traditional optimization algorithms.
- The simulation results indicate that, compared to the baseline approach, our proposed drone-assisted broadband cognitive radio network significantly enhances the system's perceptual performance and the total achievable rate of secondary users.

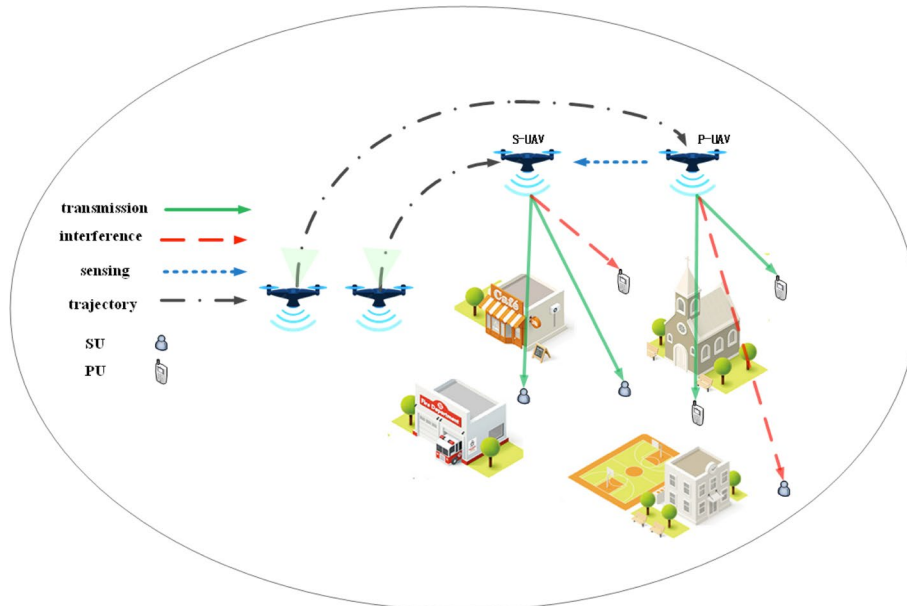
The remaining sections of this paper are as follows. Section 2 presents the system model of the proposed UAV-assisted broadband cognitive radio network and its problem framework. Section 3 describes the algorithms used for optimizing the variables. Section 4 provides simulation results. Section 5 concludes the entire paper.

## 2 System model and problem formulation

In this study, we propose a UAV-assisted wideband cognitive radio network, as illustrated in Fig. 1. The system comprises a primary UAV, a secondary UAV,  $Q$  PUs, and  $K$  SUs. In the system, the primary UAV is equipped with a single antenna, while the secondary UAV is equipped with  $N$  antennas. Both primary and secondary users are equipped with a single antenna. Let  $\mathcal{Q} = \{1, \dots, Q\}$ ,  $\mathcal{K} = \{1, \dots, K\}$ , and  $\mathcal{C} = \{1, \dots, C\}$  represent the sets of PU, SU, and subcarrier, respectively.

### 2.1 Channel model

In this study, we introduce an assumption that the primary UAV and the secondary UAV consistently maintain a fixed altitude  $H$  during their flight. Simultaneously,



**Fig. 1** A UAV-aided wideband cognitive radio network

we predefine sets of stop points (SPs) for the primary UAV and the secondary UAV, denoted as  $Q^p = [q_1, \dots, q_p, \dots, q_p]$  and  $Q^s = [q_1, \dots, q_s, \dots, q_s]$ , respectively. These sets of hover points represent the potential hover locations that the primary UAV and the secondary UAV may choose at different time instances. The coordinates of the SP  $q_p$  for the primary UAV can be represented as  $q_p = (x_p, y_p, z_p)$ , and the coordinates of the SP  $q_s$  for the secondary UAV can be represented as  $q_s = (x_s, y_s, z_s)$ . The coordinates of the primary user  $q$  are represented as  $q_q^R = (x_q^R, y_q^R, 0)$ , and the coordinates of the secondary user  $k$  are represented as  $q_k^R = (x_k^R, y_k^R, 0)$ . Based on the aforementioned information, we can represent the distances between the primary UAV at stopping point  $q_p$  and the secondary UAV at stopping point  $q_s$ , the UAV at stopping point  $q_i$  and the primary user  $q$  and the UAV at stopping point  $q_i$  and the secondary user  $k$  using the concept of Euclidean distance can be, respectively, expressed as

$$d_{ps}^{q_p, q_s} = \sqrt{(x_p - x_s)^2 + (y_p - y_s)^2 + (z_p - z_s)^2}, \tag{1a}$$

$$d_{iq}^{q_i} = \sqrt{(x_i - x_q^R)^2 + (y_i - y_q^R)^2 + z_i^2}, \tag{1b}$$

$$d_{ik}^{q_i} = \sqrt{(x_i - x_k^R)^2 + (y_i - y_k^R)^2 + z_i^2}, \tag{1c}$$

where  $i \in \{p, s\}$ .

In the system under investigation in this study, it is assumed that both the perception and transmission phases take place in a mmWave communication environment. So, the channel gains from the primary UAV to the secondary UAV, from the UAV at stopping point  $q_i$  to the PU  $q$  and from the UAV at stopping point  $q_i$  to the SU  $k$  on subcarrier  $c$  are

$$h_{ps}^{q_p, q_s} = \sqrt{\alpha \left(d_{ps}^{q_p, q_s}\right)^{-\beta}}, \tag{2a}$$

$$h_{iq,c}^{q_i} = \sqrt{\alpha \left(d_{iq}^{q_i}\right)^{-\beta}}, \tag{2b}$$

$$h_{ik,c}^{q_i} = \sqrt{\alpha \left(d_{ik}^{q_i}\right)^{-\beta}}, \tag{2c}$$

respectively, where  $\alpha$  is the path loss at a reference distance  $d_0 = 1$  m, and  $\beta$  is the path loss exponent.

### 2.2 Spectrum sensing

Spectrum sensing serves the purpose of assessing the spectral condition of the Primary Network and categorizing it into one of two states: “idle” ( $\mathcal{H}_0^c$ ) or “occupied” ( $\mathcal{H}_1^c$ ). In the event that the spectrum is identified as “idle,” the Secondary Network is able to make use of it for transmitting information. Conversely, should the spectrum be detected as “occupied,” the SN abstains from active operation to prevent any interference with the PUs. The expressions for the two scenarios mentioned above are, respectively, given as

$$H_c^0 : y_{s,c} = n_s, \tag{3a}$$

$$H_c^1 : y_{s,c} = h_{s,c}^{q_p, q_s} \sqrt{P}x + n_s, \tag{3b}$$

where  $n_s \sim \mathcal{CN}(0, \sigma_s^2)$  is indicative of the additive white Gaussian noise (AWGN) present at the secondary UAV. The variable  $h_{s,c}^{q_p, q_s}$  denotes the baseband equivalent channel from the stopping point  $q_p$  of the primary UAV to the stopping point  $p_s$  of the secondary UAV. Furthermore,  $P$  stands for the transmission power at the primary UAV, while  $x$  represents the complex baseband signal emitted by the primary UAV.

The system operates during a frame duration  $T$ , which is divided into a sensing period  $\tau$  and an information transmission period  $T - \tau$  for the Secondary Network. The detection probability and false alarm probability for subcarrier  $c$  using energy detection are provided as follows

$$P_{d,c} = Q\left(\frac{\eta^c - \sigma_c^2(N + \gamma_c)}{(\sigma_c^2 / \sqrt{\tau f_s}) \sqrt{2\gamma_c + N}}\right), \tag{4a}$$

$$P_{f,c} = Q\left(\frac{\eta^c - \sigma_c^2 N}{(\sigma_c^2 / \sqrt{\tau f_s}) \sqrt{N}}\right), \tag{4b}$$

where  $f_s$  represents the sampling frequency, while  $\gamma_c = h_{s,c}^{q_i, q_j} \frac{P}{\sigma_s^2}$  denotes the received signal-to-noise ratio (SNR) from the primary UAV at  $q_i$  to the secondary UAV at  $q_j$  in the subcarrier  $c$ , and  $Q(\cdot)$  is the right tail function of the standard normal distribution,  $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ . Additionally,  $\eta^c$  is the detection threshold in the subcarrier  $c$ , which can be expressed as

$$\eta^c = (Q^{-1}(\overline{P}_d))\sqrt{\frac{2\gamma_c + N}{\tau f_s}} + \gamma_c + N)\sigma_s^2, \tag{5}$$

where  $\overline{P}_d$  is the target detection probability.

### 2.3 Secondary network information transmission

When the secondary UAV detects that the spectrum is in an idle state, it utilizes the available idle spectrum for information transmission. However, the detection results of the secondary UAVs are not always perfectly accurate, and two scenarios can occur. The first scenario is when the secondary UAV successfully detects the idle spectrum subcarrier  $c$ . In this case, the secondary UAV uses subcarrier  $c$  to transmit the signal to the secondary user  $k$ , which can be represented as

$$y_{k,c} = \rho_{k,c} \mathbf{h}_{sk,c}^{q_s^c} f_{k,c} s_k^c + n_k, \tag{6}$$

where  $\mathbf{h}_{sk,c}^{q_s^c} \in \mathbb{C}^{1 \times N}$  represents the baseband equivalent channel from the SP  $q_s^c$  of the secondary UAV to the SU  $k$  on subcarrier  $c$ . Let  $T^P = \{q_p^1, q_p^2, q_p^3\}$  and  $T^S = \{q_s^1, q_s^2, q_s^3\}$  represent the trajectories of the primary UAV and secondary UAV, where  $q^0$  denotes the starting point,  $q^1$  represents the trajectory of the UAV during the spectrum sensing phase, and  $q^2$  represents the trajectory of the UAV during the information transmission phase.  $\mathbf{f}_k^c \in \mathbb{C}^{L \times 1}$  is the beamforming vector of the  $k$ th SU in the subcarrier  $c$ ,  $s_k^c \sim \mathcal{CN}(0, 1)$  represents the complex baseband modulated signal transmitted to secondary user  $k$  on subcarrier  $c$ , and  $n_k \sim \mathcal{CN}(0, \sigma_k^2)$  is indicative of the AWGN present at the SU  $k$ . Furthermore,  $\rho_{k,c}$  represents the subcarrier allocation status, where  $\rho_{k,c} = 1$  indicates that the secondary UAV employs subcarrier  $c$  to transmit information to SU  $k$ , while  $\rho_{k,c} = 0$  signifies that the SU  $k$  does not utilize subcarrier  $c$ .

The second scenario is when the secondary UAV experiences a missed detection during spectrum sensing. In this case, the interference signal from the primary UAV to the SU  $k$  is given by

$$I_{k,c} = \rho_{q,c} \mathbf{h}_{pk,c}^{q_p^2} \sqrt{P} x, \tag{7a}$$

where  $\mathbf{h}_{pk,c}^{q_p^2} \in \mathbb{C}$  denotes the baseband equivalent channel from the SP  $q_p^2$  of the primary UAV to the SU  $k$  on subcarrier  $c$ . The transmission power of the primary UAV is  $P$ , and  $\rho_{q,c}$  represents the subcarrier allocation status. In the two aforementioned scenarios, the received signal-to-interference-plus-noise ratio (SINR) at SU  $k$  on subcarrier  $c$  can be expressed as  $\text{SINR}_{k,c}^0$  and  $\text{SINR}_{k,c}^1$ , respectively, and they can be written as

$$\text{SINR}_{k,c}^0 = \rho_{k,c} \frac{|\mathbf{h}_{sk,c}^{q_s^c} f_{k,c}|^2}{\sigma_k^2}, \tag{8a}$$

$$\text{SINR}_{k,c}^1 = \rho_{k,c} \frac{|\mathbf{h}_{sk,c}^{q_s^c} f_{k,c}|^2}{\sum_{q=1}^Q \rho_{q,c} |\mathbf{h}_{pk,c}^{q_p^2}|^2 P + \sigma_k^2}, \tag{8b}$$

The probabilities of the two aforementioned scenarios occurring are

$$\psi_c^0 = \Pr(\mathcal{H}_c^0)(1 - P_{f,c}), \tag{9a}$$

$$\psi_c^1 = \Pr(\mathcal{H}_c^1)(1 - P_{d,c}), \tag{9b}$$

where  $\Pr(\mathcal{H}_c^0)$  represents the probability of subcarrier  $c$  being in an idle state, and  $\Pr(\mathcal{H}_c^1)$  represents the probability of subcarrier  $c$  being in an occupied state. Therefore, the achievable rate of the  $k$ th SU in the  $c$ th subcarrier can be written as

$$R_{k,c} = \frac{T - \tau}{T} \rho_{k,c} [\psi_c^0 \log(1 + \text{SINR}_{k,c}^0) + \psi_c^1 \log(1 + \text{SINR}_{k,c}^1)], \tag{10}$$

Based on the above scenarios, the total achievable rate of the secondary network can be expressed as

$$R_s = \sum_{c=1}^C \sum_{k=1}^K R_{k,c}. \tag{11a}$$

#### 2.4 Problem formulation

The problem of system average rate maximization by jointly optimizing  $T^p$ ,  $T^s$ ,  $\mathbf{F} = \{\mathbf{f}_{k,c}\}_{\forall k \in \mathcal{K}, c \in \mathcal{C}}$ ,  $\tau$  and  $\rho = \{\rho_{k,c}\}_{\forall k \in \{\mathcal{K}, \mathcal{Q}\}, c \in \mathcal{C}}$ , in the SN of UAV-assisted wideband CR system can be modeled as

$$\text{P1: } \max_{T^p, T^s, \mathbf{F}, \tau, \rho} R_s \tag{12a}$$

$$\text{s.t. } P_{d,c} \geq \tilde{P}_d, \tag{12b}$$

$$P_{f,c} \leq \tilde{P}_f, \tag{12c}$$

$$\frac{T - \tau}{T} \sum_{c=1}^C \sum_{k=1}^K \rho_{k,c} (\psi_c^0 + \psi_c^1) |f_{k,c}|^2 \leq P_T, \forall c \in \mathcal{C}, \forall k \in \mathcal{K}, \tag{12d}$$

$$\sum_{c=1}^C R_{k,c} \geq R_{th}, \forall c \in \mathcal{C}, \tag{12e}$$

$$\frac{T - \tau}{T} \sum_{c=1}^C \sum_{k=1}^K \rho_{k,c} \rho_{q,c} \psi_c^1 |h_{sq,c}^q f_{k,c}|^2 \leq P_q^I, \forall q \in \mathcal{Q}, \tag{12f}$$

$$0 \leq \tau \leq T, \tag{12g}$$

$$\rho_{i,c} \in \{0, 1\}, \forall i \in \{\mathcal{K}, \mathcal{Q}\}, \tag{12h}$$



$$\sum_{c=1}^C \rho_{i,c} \leq 1, \tag{12i}$$

$$\|q_i - q_l^R\| \leq S_{\max}^2, q_i \in \{Q^p, Q^s\}, l \in \{Q, K\}, \tag{12j}$$

where  $\overline{P_f}$  represents the maximum allowable false alarm probability,  $P_T$  is the maximum transmission power of the secondary UAV,  $P_q^I$  is the maximum interference to the PU  $q$  caused by the secondary UAV, and  $S_{\max}$  is the maximum distance covered by a UAV in a single move. Equation (12b) represents the minimum detecting probability constraint. Equation (12c) represents the maximum tolerable false alarm probability constraint. Equation (12d) represents the maximum transmission power constraint for the secondary UAV. Equation (12e) represents the minimum achievable rate constraint for secondary user  $k$ . Equation (12f) represents the maximum interference constraint imposed by the secondary UAV on the primary user  $q$ . Equation (12g) represents the sensing time constraint. Equation (12h) denotes the subcarrier allocation scenario. Equation (12i) indicates that each subcarrier can be allocated to at most one user for information transmission. Equation (12j) represents the maximum UAV movement distance constraint.

### 3 Resource optimization scheme

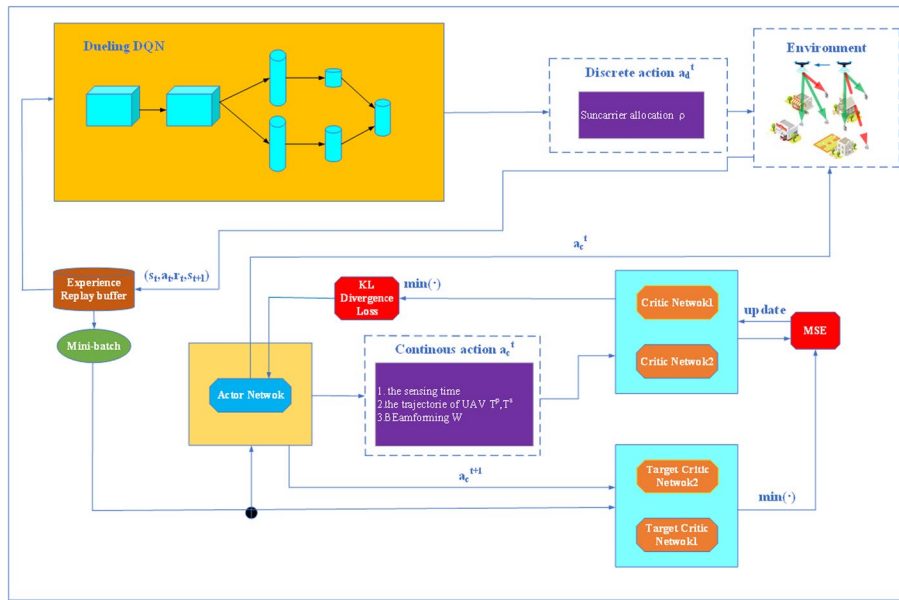
In this section, we utilize the DDQN-SAC algorithm to optimize the variables in problem P1. In comparison with traditional optimization algorithms, this algorithm offers significant advantages. Firstly, it effectively handles mixed-action spaces, addressing optimization problems with both continuous and discrete variables simultaneously. Secondly, it combines the strengths of DDQN and SAC algorithms, providing improved stability and convergence. Additionally, it significantly reduces computational complexity, enhances training efficiency, and reduces resource requirements compared to traditional optimization algorithms. Lastly, it doesn't necessitate detailed prior knowledge or precise system models, making it suitable for a wide range of complex real-world problems.

#### 3.1 Problem transformation

We consider our formulated optimization problem as a reinforcement learning problem. In reinforcement learning, the UAV-assisted broadband cognitive radio network is regarded as an environment, and the central controller is seen as an agent. Besides the environment and the agent, reinforcement learning also encompasses elements such as state space, action space, transition probabilities, and reward functions (Fig. 2).

*State space* The state space consists of all possible states in the environment, which includes information relevant to the agent for making decisions based on the current state. At time slot  $t$ , the state space includes the action space of the previous time slot  $a^{(t-1)}$ , the channel state information of the current time slot  $\mathbf{H}^{(t)}$ , and the achievable rate of the secondary network from the previous time slot  $R_{\text{ave}}^{(t-1)}$ . So, the state space at time slot  $t$  can be represented as  $s^{(t)} = \{a^{(t-1)}, \mathbf{H}^{(t)}, R_{\text{ave}}^{(t-1)}\}$ .

*Action space* The action space defines the set of all possible actions available to an intelligent agent. The composition of the action space is contingent upon the inherent



**Fig. 2** The framework of the proposed DDQN-SAC algorithm

characteristics of the problem and the specific application context, and it may manifest as continuous, discrete, or hybrid. In the context of this paper, the action space  $a^t$  at time slot  $t$  is configured as a hybrid action space. It encompasses a discrete action space  $a_d^t$  that encapsulates trajectory optimization for the primary UAV, trajectory optimization for the secondary UAV, and subcarrier allocation. Furthermore, it incorporates a continuous action space  $a_c^t$  dedicated to beamforming at the secondary UAV and perception time. These distinct actions collectively constitute the hybrid action space, succinctly denoted as  $a^t = \{a_d^t, a_c^t\}$ .

*Transition probability* The transition probability  $Pr(s^{(t+1)}|s^{(t)}, a^{(t)})$  represents the probability of transitioning from state  $s^{(t)}$  to state  $s^{(t+1)}$  by taking action  $a^{(t)}$ . In this paper, the transition probability follows the changes of channel information.

*Reward function* The reward function in DRL is a function used to evaluate the goodness of the actions taken by the intelligent agent in its interaction with the environment. In this paper, the reward function is composed of the total rate of the secondary network and the maximum interference constraint on the primary user, represented as

$$r = \alpha_1 R_s + \alpha_2 \delta, \tag{13}$$

where  $\alpha_1$  and  $\alpha_2$  are constants, and  $\delta$  is a penalty term, which can be represented as

$$\delta = \begin{cases} 0, & \text{if } P_{I,q} \leq P_T, \forall q \in Q, \\ \chi, & \text{otherwise,} \end{cases} \tag{14}$$

where  $\chi$  is a negative constant, and  $P_{I,q}$  represents the interference caused by the secondary network to the primary user  $q$ , which can be expressed as

$$P_{I,q} = \frac{T - \tau}{T} \sum_{c=1}^C \sum_{k=1}^K \rho_{k,c} \rho_{q,c} \psi_c^1 |h_{sq,c}^{q_s} f_{k,c}|^2, \forall q \in Q. \tag{15}$$

### 3.2 Problem optimization based on DDQN-SAC

*Using the DDQN algorithm to handle a discrete action space* In order to more accurately estimate the values of different actions in various states, improve training stability and efficiency, and decompose the Q values into state-value functions and action-advantage functions, the Q value can be obtained by

$$Q(s, a; \theta, \alpha, \beta) = V(s; \theta, \alpha) + (A(s, a; \theta, \beta) - \frac{1}{|A|} \sum_{a'} A(s, a'; \theta, \beta)), \tag{16}$$

where  $\theta$  represents the weight parameters of the neural network used to approximate the Q value function,  $\alpha$  represents the weight parameters of the state-value function, and  $\beta$  represents the weight parameters of the action-advantage function. Furthermore,  $|A|$  represents the number of possible actions in the action space.

Furthermore, the target Q value is estimated through the Bellman equation, representing the Q value that the agent expects to achieve. Using the mean squared error loss function, the error between the actual Q value and the target Q value can be measured, i.e., the loss function, which is expressed as

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N (Q(s_i, a_i; \theta) - (r + \gamma \max_{a'} Q(s_{i+1}, a'; \theta^-)))^2, \tag{17}$$

where  $N$  represents the tuple size extracted from the experience replay buffer, and  $\theta^-$  represents the weight parameters of the target neural network.

*Using the SAC algorithm to handle a continuous action space* The network structure of SAC consists of five neural networks, namely the actor network, two critic networks, and two target value networks, with their network parameters being  $\theta_a, \theta_{c1}, \theta_{c2}, \theta_{c1}^t$ , and  $\theta_{c2}^t$ , respectively.

Sample small batches of tuples  $N$  from the experience replay buffer in order to calculate the policy loss. SAC algorithm utilizes maximum entropy reinforcement learning, with the policy loss function which can be expressed as

$$L(\theta_a) = \frac{1}{N} \sum_{n=1}^N [\alpha' \log \pi(a_n | s_n; \theta_a) - \min_{i=1,2} Q_i(s_n, a_n; \theta_{ci})]. \tag{18}$$

where  $\theta_a$  and  $\theta_{ci}$  correspond to the actor network parameters and critic network parameters, respectively. Additionally,  $\alpha'$  represents the temperature parameter used to control exploration. And the above equation includes a minimization operation designed to address the issue of network overestimation.

The target Q value for any tuple  $n$  sampled from the experience replay buffer can be expressed as

$$y_n = r_n + \gamma [\min_{i=1,2} Q_i(s_{n+1}, a_{n+1}; \theta_{ci}^t) - \alpha' \log \pi(a_{n+1}|s_{n+1}; \theta_a)], \quad (19)$$

where  $\gamma$  represents the discount factor ( $\gamma \in (0, 1)$ ). Furthermore,  $\theta_{ci}^t$  represents the parameters of the target critic network. By comparing the estimated  $Q$  values generated by the current policy with the target  $Q$  values, the loss function provides a signal to guide policy improvement for the system to achieve higher rewards. The loss function can be given as:

$$L(\theta_{ci}^t) = \frac{1}{N} \sum_{n=1}^N (y_n - Q(s_n, a_n; \theta_{ci}))^2, \quad (20)$$

In summary, the algorithm starts with the initialization of neural network parameters for both DDQN and SAC. Next, the UAV-assisted broadband cognitive radio network is taken as input, and the generated actions are applied to the environment, resulting in rewards and the next state. These experiences are stored as tuples in the experience replay buffer. After accumulating a certain number of tuples in the replay buffer, small batches of tuples are randomly sampled. Formulas (16) and (18) are then used to calculate the loss function for updating the value network parameters. In the case of the SAC algorithm, KL divergence is calculated using Formula (17) and the actor network parameters are updated. Detailed steps of the DDQN-SAC algorithm are provided in Algorithm 1, outlining the specific optimization process.

**Algorithm 1** Optimization of UAV-assisted wideband cognitive radio network based on DDQN-SAC.

- 
- 1: **Initialization:** Experience buffer with size  $D$ , network parameter  $\theta$ ,  $\alpha$ ,  $\beta$ ,  $\theta^t$ ,  $\theta_a$ ,  $\theta_{c1}$ ,  $\theta_{c2}$ ,  $\theta_{c1}^t$  and  $\theta_{c2}^t$ .
  - 2: **Input:** The Channel state information  $\mathbf{H}$  at time slot  $t$  and the achievable rate of SU at time slot  $t - 1$ .
  - 3: **Output:** The optimal action  $a^*$  obtained from the policy network.
  - 4: **for** each episode  $I = 1, 2, \dots, N$  **do**
  - 5:     Observe the initial system state  $s_0$ .
  - 6:     **for** each step  $t = 1, 2, \dots, T$  **do**
  - 7:         In time step  $t$ , the discrete action  $a_d^t$  is determined using a greedy strategy.
  - 8:         In time step  $t$ , continuous action  $a_c^t$  is generated from actor network in SAC algorithm.
  - 9:         Applying action  $a^t = \{a_d^t, a_c^t\}$  to state  $s^t$  results in obtaining a reward  $r^t$  and the next state  $s^{t+1}$ . Store tuple  $(s^t, a^t, r^t, s^{t+1})$  in the experience replay buffer.
  - 10:         Extract mini-batch tuples of size  $N$  from the experience replay buffer.
  - 11:         Obtain the mean squared error based on equation (17) and update  $\theta$ ,  $\alpha$ , and  $\beta$ .
  - 12:         Obtain the mean squared error based on equation (20) and update  $\theta_{c1}$  and  $\theta_{c2}$ .
  - 13:         Update the action network parameters  $\theta_a$  using the KL divergence based on equation (18).
  - 14:         Update the target network parameters  $\theta^t$  using the soft update strategy.
  - 15:         Update the target evaluation network parameters  $\theta_{c1}^t$  and  $\theta_{c2}^t$  using the soft update strategy.
  - 16:     **end for**
  - 17: **end for**
-

#### 4 Simulation results

In this section, we present simulation results to assess the performance of the UAV-assisted wideband cognitive radio system proposed in this study. The simulation experiments were conducted on a simulation platform using Python 3.9 and PyTorch 1.10.2. The simulation parameters were configured based on references [11, 15].

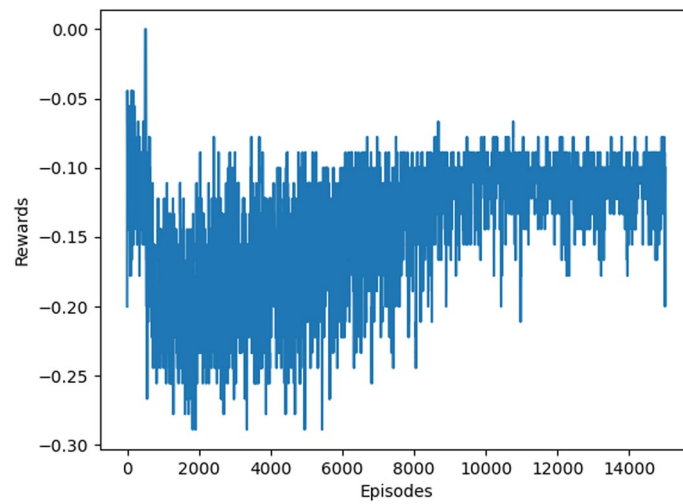
The secondary UAV is equipped with  $N = 6$  antennas, and there are  $Q = 6$  primary users and  $K = 6$  secondary users. The UAV flies at a fixed altitude of  $H = 50$  m. In addition, the AWGN at the secondary UAV and SU  $k$  locations has  $\sigma_s^2 = \sigma_k^2 = 0.01$ , the sampling frequency is  $f_s = 6$  MHz, and the transmit power of the primary base station is  $P = 30$  dBm. The target detection probability is  $\overline{P_d} = 0.9$ , and the maximum allowable false alarm probability is  $\overline{P_f} = 0.1$ .

In the simulation section, we introduce two benchmark schemes to assess the performance of the method proposed in this paper. Benchmark 1: The beamforming at the secondary UAV is fixed at a constant value, with all other parameters kept identical to those in the proposed method. Benchmark 2: The number of channels is constrained to 1, maintaining consistency with all other aspects specified in the proposed method.

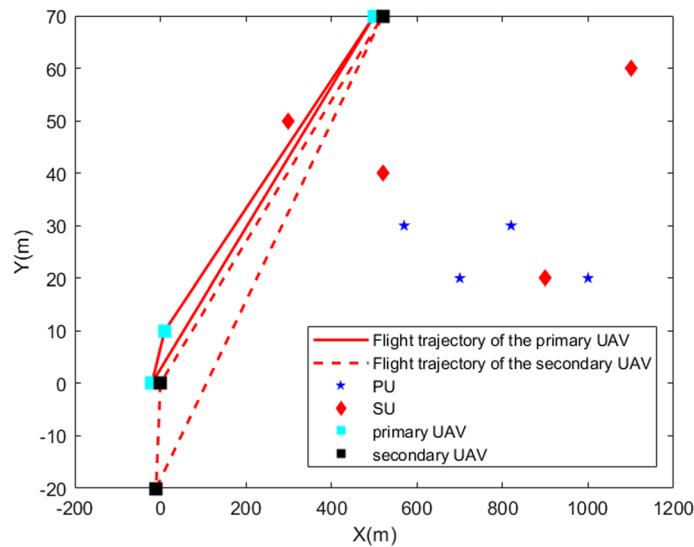
Figure 3 illustrates the convergence of the optimization of the UAV-assisted cognitive radio network using the DDQN-SAC algorithm proposed in this work. From Fig. 3, it is evident that our proposed algorithm performs in a total round of 14,000 iterations and started to converge around the 8000 iterations. Furthermore, apart from the initial fluctuations, it can be observed that as the number of iterations increased, the algorithm's rewards steadily increased, eventually reaching convergence. It is that the proposed algorithm exhibits excellent convergence ability and stability, this is due to the powerful exploration capability of the DRL.

Figure 4 illustrates the trajectory movements of unmanned aerial vehicles (UAVs) at different stages, namely during the perception phase and the transmission phase. After the information transmission is complete, the UAVs will return to their initial positions. It is worth noting that the movement direction of the main UAV during the perception phase may be influenced by the trajectory of the secondary UAV. This is done to minimize potential instances of false detection or false alarms during the perception phase. Furthermore, during the information transmission phase, the secondary UAV moves away from the location of the primary user to ensure that there is no interference with the primary user, thus meeting the communication requirements. In this manner, the UAV system can effectively balance the demands of perception and communication, reduce interference with the primary user, and enhance the system performance.

Figure 5 illustrates the impact of varying secondary UAV transmission power on the average achievable rate of the secondary users. The observations reveal a substantial increase in the average achievable rate of secondary users in all schemes as the transmission power of the secondary UAV increases. It is noteworthy that our proposed method demonstrates significantly enhanced system performance when compared to the other benchmark approaches. As analyzed numerically, our proposed method obtains the rate improvement up to 316 % compared to the benchmark 2 and up to 47% compared to benchmark 1, which vividly demonstrates the superior performance of our propose method.

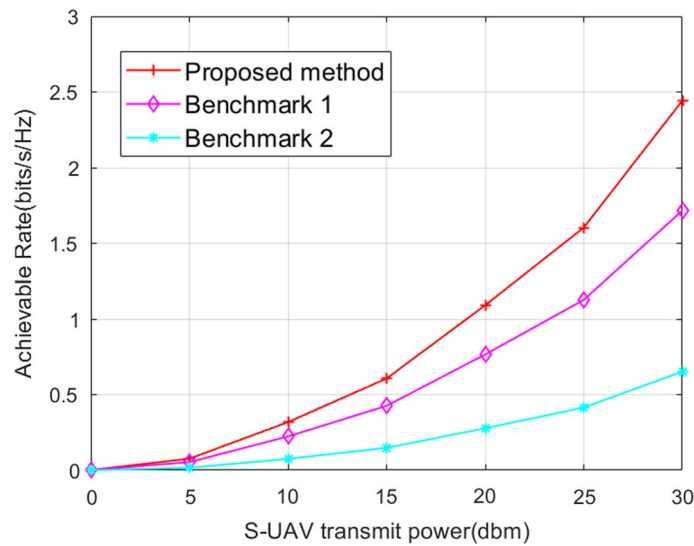


**Fig. 3** The convergence of the proposed algorithm

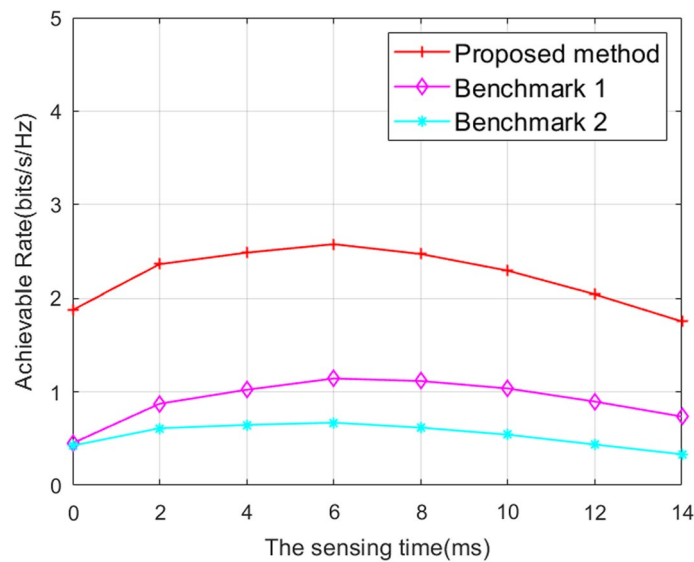


**Fig. 4** The flight trajectory of the UAVs

Figure 6 illustrates the impact of varying sensing time on the average achievable rate of secondary users in different scenarios. According to the graph, it can be observed that with an increase in perception time, the average achievable rate of secondary users shows an initial increase followed by a decrease. This phenomenon can be explained by the fact that, as perception time increases, the system’s perceptual performance gradually improves, resulting in a reduction in missed detections and false alarms. However, when the perceptual performance reaches a certain threshold, further increasing the perception time no longer significantly enhances perceptual performance. Instead, it consumes additional information transmission time in the secondary network, leading to a decrease in the average achievable rate of secondary users. Furthermore, it is evident from the graph that, in comparison with the baseline



**Fig. 5** The average achievable rate of the SUs versus the transmit power of the secondary UAV



**Fig. 6** The average achievable rate of the SUs versus sensing time

approach, the solution proposed in this paper consistently demonstrates superior performance across various scenarios.

### 5 Conclusion

In this study, we proposed a UAV-assisted broadband cognitive radio network scheme. The objective of this scheme was to jointly optimize the trajectories of the primary UAV and secondary UAV, the beamforming patterns of secondary UAV, and the subcarrier allocation to maximize the achievable rate of secondary users while adhering to maximum interference constraints on the primary user. To address the challenge of dealing with a hybrid action space problem, we employed the

DDQN-SAC algorithm for problem optimization. Through the presentation of simulation results, we observed a significant improvement in system performance compared to the baseline approaches.

#### Acknowledgements

We appreciate the editors and reviewers who processed and reviewed our manuscript to provide the detailed professional comments on the technical contributions, logical structure, and content presentation of this paper.

#### Author contributions

LY conducted conceptualization, writing, original draft preparation, and software; YC conducted software, methodology, validation, and figures; HW conducted formal analysis and editing; and all authors reviewed the manuscript.

#### Funding

This work was supported in part by the key scientific and technological project of Henan province (Grant Nos. 212102210558, 232102320050); The Nature Science Foundation of Henan province (Grant No. 202300410101); and Doctoral research start project of Henan Institute of Technology (Grant No. KQ1852, KQ813).

#### Availability of data and materials

Data sharing is not applicable to this study.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

All authors consent to the publication of this manuscript.

##### Competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Received: 12 November 2023 Accepted: 20 March 2024

Published online: 02 April 2024

#### References

1. P. Ahokangas, M. Matinmikko-Blue, S. Seppo, Envisioning a future-proof global 6G from business, regulation, and technology perspectives. *IEEE Commun. Mag.* **61**(2), 72–78 (2023)
2. G.P. Fettweis, H. Boche, 6G: the personal tactile internet and open questions for information theory. *IEEE BITS Inf. Theory Mag.* **1**(1), 71–81 (2021)
3. W.S. Ahmad, N.A. Radzi, F.S. Samidi, A. Ismail, F. Abdullah, M.Z. Jamaludin, M. Zakaria, 5G technology: towards dynamic spectrum sharing using cognitive radio networks. *IEEE Access* **8**, 14460–14488 (2020)
4. H. Sun, F. Zhou, R.Q. Hu, L. Hanzo, Robust beamforming design in a NOMA cognitive radio network relying on SWIPT. *IEEE J. Sel. Areas Commun.* **37**(1), 142–155 (2019)
5. M. Amjad, M.H. Rehmani, S. Mao, Wireless multimedia cognitive radio networks: a comprehensive survey. *IEEE Commun. Surv. Tutor.* **20**(2), 1056–1103 (2018)
6. L. Wang, W. Wu, F. Zhou, Intelligent resource allocation for IRS-assisted sensing-enhanced secure communication CRNs, in 2023 International Conference on Ubiquitous Communications (2023)
7. L. Wang, W. Wu, F. Tian, H. Hu, Intelligent resource allocation for UAV-enabled spectrum sharing semantic communication networks, in: Proceedings of IEEE International Conference on Communication Technology (ICCT), pp. 1359–1363 (2023)
8. S. Dhanasekaran, T. Reshma, Full-rate cooperative spectrum sharing scheme for cognitive radio communications. *IEEE Commun. Lett.* **14**(8), 97–100 (2018)
9. X. Guan, Q. Wu, R. Zhang, Joint power control and passive beamforming in IRS-assisted spectrum sharing. *IEEE Commun. Lett.* **24**(7), 1553–1557 (2020)
10. L. Wang et al., Intelligent resource allocation for transmission security on IRS-assisted spectrum sharing systems with OFDM. *Phys. Commun.* **58**, 102013 (2023)
11. Y. Wu, F. Zhou, Q. Wu, Y. Huang, R. Q. Hu, Resource allocation for IRS-assisted sensing-enhanced wideband CR networks, in Proceedings of IEEE International Conference on Communications (2021)
12. X. Kang, Y.-C. Liang, H.K. Garg, L. Zhang, Sensing-based spectrum sharing in cognitive radio networks. *IEEE Trans. Veh. Technol.* **58**(8), 4649–4654 (2009)
13. L. Wang, W. Wu, F. Zhou, Q. Wu, O. A. Dobre, T. Q. Quek, Hybrid hierarchical DRL enabled resource allocation for secure transmission in multi-IRS-assisted sensing-enhanced spectrum sharing networks, *IEEE Trans. Wirel. Commun.* (2023)
14. Y.H. Bae, J.W. Baek, Achievable throughput analysis of opportunistic spectrum access in cognitive radio networks with energy harvesting. *IEEE Trans. Commun.* **64**(4), 1399–1410 (2016)
15. W. Wu et al., Joint sensing and transmission optimization for IRS-assisted cognitive radio networks. *IEEE Trans. Wirel. Commun.* **22**(9), 5936–5941 (2023)
16. S. Stotas, A. Nallanathan, On the throughput and spectrum sensing enhancement of opportunistic spectrum access cognitive radio networks. *IEEE Trans. Wirel. Commun.* **11**(1), 97–107 (2012)



17. O. Altrad, S. Muhaidat, A. Al-Dweik, A. Shami, P.D. Yoo, Opportunistic spectrum access in cognitive radio networks under imperfect spectrum sensing. *IEEE Trans. Veh. Technol.* **63**(2), 920–925 (2014)
18. Y. Gao et al., Robust trajectory and communication design for angle-constrained multi-UAV communications in the presence of jammers. *China Commun.* **19**(2), 131–147 (2022)
19. K. Liu et al., Deployment and robust hybrid beamforming for UAV MmWave communications. *IEEE Trans. Commun.* **71**(5), 3073–3086 (2023)
20. C. Zhong, J. Yao, J. Xu, Secure UAV communication with cooperative jamming and trajectory control. *IEEE Commun. Lett.* **23**(2), 286–289 (2019)
21. Y. Yu et al., Joint trajectory and resource optimization for RIS assisted UAV cognitive radio. *IEEE Trans. Veh. Technol.* **72**(10), 13643–13648 (2023)
22. H. Hu et al., Optimization of energy management for UAV-enabled cognitive radio. *IEEE Wirel. Commun. Lett.* **9**(9), 1505–1508 (2020)
23. Y. Wang et al., Resource allocation and trajectory design in UAV-assisted jamming wideband cognitive radio networks. *IEEE Trans. Cogn. Commun. Netw.* **7**(2), 635–647 (2021)
24. S.K. Nobar et al., Resource allocation in cognitive radio-enabled UAV communication. *IEEE Trans. Cogn. Commun. Netw.* **8**(1), 296–310 (2022)
25. Y. Jiang, J. Zhu, Three-dimensional trajectory optimization for secure UAV-enabled cognitive communications. *China Commun.* **18**(12), 285–296 (2021)
26. L. Wang, et al., Adaptive Resource Allocation for Semantic Communication Networks. [arXiv:2312.01081](https://arxiv.org/abs/2312.01081) (2023)
27. T.P. Turong et al., Flyreflect: joint flying IRS trajectory and phase shift design using deep reinforcement learning. *IEEE Internet Things J.* **10**(5), 4605–4620 (2023)
28. H. Wang et al., Joint UAV placement optimization, resource allocation, and computation offloading for THz band: A DRL approach. *IEEE Trans. Wirel. Commun.* **22**(7), 4890–4900 (2023)
29. S. Bi et al., Deep reinforcement learning for IRS-assisted UAV covert communications. *China Commun.*, to be published (2023)
30. J. Moon et al., Joint UAV placement optimization, resource allocation, and computation offloading for THz band: a DRL approach. *IEEE Internet Things J.* **8**(20), 15441–15455 (2021)

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.