A Deep Reinforcement Learning Framework for Spectrum Management in Dynamic Spectrum Access

Hao Song, Lingjia Liu, *Senior Member, IEEE*, Jonathan Ashdown, *Member, IEEE*, and Yang Yi

***Abstract*—Dynamic spectrum access (DSA) has the great potential to alleviate spectrum shortage and promote network capacity. However, two fundamental technical issues have to be addressed, namely, interference coordination between DSA users and interference suppression for primary users (PUs). These two issues are very challenging since generally there is no powerful infrastructures in DSA networks to support centralized control. As a result, DSA users have to perform spectrum management individually, including spectrum access and power allocation, without accurate channel state information and centralized control. In this article, a novel spectrum management framework is proposed, in which *Q*-learning, a type of reinforcement learning, is utilized to enable DSA users to carry out effective spectrum management individually and intelligently. For more efficient process, neural networks (NNs) are employed to implement *Q*-learning processes, so-called deep *Q*-network (DQN). Furthermore, we also investigate the optimal way to construct DQN considering both the performance of wireless communications and the difficulty of NN training. Finally, extensive simulation studies are conducted to demonstrate the effectiveness of the proposed spectrum management framework.**

***Index Terms*—Deep *Q*-network (DQN), dynamic spectrum access (DSA), echo state networks (ESNs), reinforcement learning (RL), spectrum management.**

I. Introduction

THE white paper released by CISCO foresees that global mobile data traffic will witness exponential growth from 2017 to 2022 with a 46% compound annual growth rate and a sevenfold total increase [1]. Such a tremendous growth makes spectrum resources extremely scarce and costly, as all mobile operators seek for spectrum extensions to meet the future mobile data traffic demand. However, an opposite fact is disclosed by relevant practical measurements and investigations that many precious spectrum resources are underutilized. Even in some crowded cities, such as New York and Chicago, the utilization rate of spectrum resources is low, below 40% [2]. These measurement results spur the Federal Communication Commission (FCC) to review the legacy static spectrum allocation policy, which limits potential users to obtain spectrum access opportunities [3]. Hence, a concept of dynamic spectrum access (DSA) is raised, where spectrum resources will be shared by different users rather than just licensed users. DSA is an enabling and supporting technology for distributed Internet of Things (IoT) networks, where IoT devices have to manage their wireless resources individually, including spectrum access and transmit power, as no powerful infrastructure exists to provide centralized control [4].

Many frequency bands are opened up to support DSA. One of the representative applications of DSA is unlicensed spectrum access, such as industrial, scientific, and medical (ISM) bands, and unlicensed national information infrastructure (UNII) bands. LTE systems have been encouraged to extend their system bandwidth by accessing 5.8-GHz ISM bands, such as licensed-assisted access (LAA) and LTE-unlicensed (LTE-U) systems [5]. The relevant enabling technologies have been widely studied, such as resource allocation and co-existence between LTE-U and Wi-Fi [6], [7]. However, crowded Wi-Fi devices have made unlicensed bands extremely congested and detrimental interference environments may be encountered on these bands. Therefore, the FCC searches for more available spectrum to satisfy the demand of DSA users by exploring ultra-wideband millimeter-wave (mmWave) bands. To facilitate the use of mmWave, the license of 14-GHz contiguous mmWave from 57- to 71-GHz bands have been opened up [8]. Unfortunately, effective wireless transmissions on mmWave bands require complicated and dedicated designed signal processing technologies and hardware, causing severe overhead. Therefore, to provide more DSA opportunities on superior low-frequency bands, the FCC has decided to further exploit underutilized licensed bands. For example, in 2015, an auction was held by the FCC for the license of advanced wireless services (AWS-3) bands, including 1695–1710-MHz, 1755–1780-MHz, and 2155–2180-MHz bands. The winner in the auction is allowed to access AWS-3 bands as secondary users. In addition, 3550–3700-MHz bands, also referred to as 3.5-GHz bands will be available for DSA in the future [9].

Despite many advantages, DSA on opened licensed bands is very challenging for the following reasons. First, on most opened licensed bands, incumbents, also referred to as primary users (PUs) exist, which possess higher priorities and should

Manuscript received September 9, 2020; revised December 14, 2020; accepted January 6, 2021. Date of publication January 27, 2021; date of current version July 7, 2021. This work was supported by the U.S. National Science Foundation under Grant ECCS-1802710, Grant ECCS-1811497, Grant CNS-1811720, and Grant CCF-1937487. Approved for public release (Case Number: 88ABW-2019-5765). This article was presented in part at IEEE Conference on Computer Communications Workshops (2019 INFOCOM WKSHPS), Paris, France, 2019. *(Corresponding author: Lingjia Liu.)*

Hao Song, Lingjia Liu, and Yang Yi are with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA (e-mail: ljliu@ieee.org).

Jonathan Ashdown is with the Information Directorate, Air Force Research Laboratory, Rome, NY 13441 USA.

Digital Object Identifier 10.1109/JIOT.2021.3052691

be protected from harmful interference. For example, on AWS-3 bands, the federal meteorological-satellite (MetSat) systems are PUs and should be protected by any DSA user that intends to access AWS-3 bands [9]. Second, generally, there is no powerful infrastructure to provide centralized control for DSA users. DSA users have to conduct spectrum management individually [10]. As a result, DSA users may suffer from severe interference. Unfortunately, the interference issue in DSA networks cannot be addressed by the traditional methods, such as interference coordination [11] and interference cancellation [12], which depend upon cooperation between users or accurate channel state information (CSI) estimations for other users.

As powerful tools, applying machine learning and deep learning technologies in the wireless communication field has been widely studied to improve system performance or efficiency, such as beamforming management [13] and resource allocation [14]. However, these proposed methods are based on supervised learning, which require training data. Training data could be acquired by measurements or generated by the particular model of application scenarios. However, practical measurements are very costly, since tremendous amounts of data need to be collected and processed. Moreover, in a dynamic system like DSA networks, the model is normally unknown, as the information regarding network layout and channel states is unavailable. In this article, a novel framework is proposed to leverage deep reinforcement learning (RL) in spectrum management, including both spectrum access and power allocation, in DSA networks. In the proposed framework, Q-learning with the model-free nature is utilized to enable a DSA user to learn wireless environments, which are dominated by behaviors of PUs and other DSA users. Furthermore, a DSA user will carry out spectrum management just by interacting with environments without depending on any CSI and cooperation with other users [15]. Nonetheless, Q-learning cannot handle large exploration space. With a large amount of states and actions, the training of Q-learning will become difficult [16]. For fast convergence, neural networks (NNs) are utilized to perform Q-learning processes, including approximating the expected cumulative reward and exploring optimal state–action pairs, so-called deep Q-network (DQN), which is a type of deep RL [17]. In our earlier work, a deep RL approach is introduced for spectrum access in DSA networks, where power allocation is not considered in [18] and [19]. In this article, we apply deep RL in both spectrum access and power allocation of DSA networks. The key contributions of this article are summarized as follows.

1) A framework of spectrum management is developed based on DQN for DSA networks, enabling DSA users to perform proper spectrum management individually and intelligently without relying on accurate channel estimations and centralized control. In the DQN, the spectrum management strategies, including spectrum access and power allocation, are defined as states, while the adjustment for spectrum management is defined as actions, which is conducted based on the reward obtained through interacting with environments directly. Additionally, both the co-existence between DSA users and the protection for PUs are taken into account in the framework.
2) We provide a comprehensive investigation of the proper way to constitute DQN. Especially, we focus on studying the performance of echo state networks (ESNs), a special type of recurrent NNs (RNNs), in spectrum management of DSA, which possess the temporal correlation attribute and are easier to be trained compared to traditional RNNs. Through simulations, comparison, and analysis, the optimal selection of NNs is found, which can bring in an excellent performance in terms of both achievable data rate, PU protection, and convergence behaviors.

The remainder of this article is organized as follows. In Section II, the system model of DSA networks is described. Section III presents the design of interference information feedback methods to support spectrum management in DSA. In Section IV, a novel framework of spectrum management leveraging Q-learning is proposed. Then, a DQN-enabled spectrum management scheme is introduced and the selection of NNs is discussed in Section V. In Section VI, simulation results are plotted along with performance analysis. Finally, this article is summarized in Section VII.

II. System Model and Problem Formulation

A DSA network consisting of multiple DSA users and PUs is considered, which is constructed in the distributed fashion without powerful infrastructures and centralized control. Without loss of generality, it is assumed that there are a transmitter (TX) and a receiver (RX) in each DSA user, so-called a DSA user pair. DSA users opportunistically access wireless channels shared with other DSA users and PUs. For simplification, a reasonable assumption is made that each PU only uses one wireless channel and PUs occupy different channels to avoid making interference to each other.

The main notations used in this article are shown as follows. Let $\mathbf{N} = \{n|n = 1, 2, \ldots, N\}^T$ and $\mathbf{M} = \{m|m = 1, 2, \ldots, M\}^T$ represent the sets of DSA users and wireless channels, respectively. Under the assumption that different PUs occupy different channels, the channel set $\mathbf{M}$ can also denote the set of PUs. The sets of channels allocated to DSA user n and DSA users accessing channel m are denoted by $\mathbf{\Omega}_n = \{m|m = 1, 2, \ldots, M_n\}^T$ and $\mathbf{\Phi}_m = \{n|n = 1, 2, \ldots, N_m\}^T$, respectively.

Due to lack of centralized control in DSA networks, DSA users may suffer from the interference caused by other DSA users and PUs. Accordingly, the received signals of DSA user n on channel m are shown in Fig. 1, which could be given by

$$y_n^m = x_n^m \cdot h_{nn}^m + x_m^m \cdot h_{mn}^m + \sum_{j \in \Phi_m, j \neq n} x_j^m \cdot h_{jn}^m + z_n^m \qquad (1)$$

where x_n^m is the desired signal of DSA user n on channel m, and x_m^m and x_j^m denote interference signals from PU m and DSA user j, respectively. Correspondingly, h_{nn}^m, h_{mn}^m, and h_{jn}^m stand for the channel gains of the links from the TX to the RX of DSA user n, from PU m to DSA user n, and from DSA user j to DSA user n, respectively. Besides, the additive white

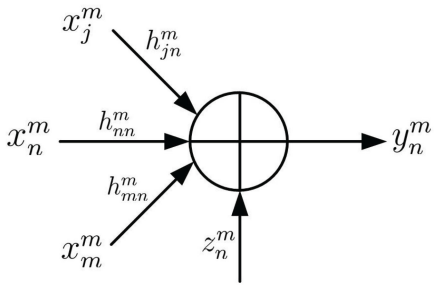


Fig. 1. Received signals of a DSA user.

Gaussian noise (AWGN) received on channel m is denoted by z_n^m.

The corresponding signal-to-interference-plus-noise ratio (SINR) is given by

$$r_n^m = \frac{p_n^m \cdot \left|h_{nn}^m\right|^2}{\underbrace{p_m^m \cdot \left|h_{mn}^m\right|^2}_{\text{Interference from PU m}} + \underbrace{\sum_{j\in\Phi_m, j\neq n} p_j^m \cdot \left|h_{jn}^m\right|^2}_{\text{Interference from other DSA users}} + \underbrace{B\cdot N_0}_{\text{noise}}} \tag{2}$$

where p_n^m, p_m^m, and p_j^m represent the transmit power of n, m, and j on channel m, respectively. B and N_0 are channel bandwidth and noise spectral density, respectively. The corresponding spectral efficiency is $\log_2(1 + r_n^m)$.

With r_n^m, the spectrum management problem in DSA with respect to variables $\{p_n^m\}_{n\in\mathbf{N},m\in\mathbf{M}}$ can be formulated as

$$\begin{aligned} \max_{\{p_n^m\}} \quad & \sum_{n\in\mathbf{N}}\sum_{m\in\mathbf{M}} \log_2\left(1 + r_n^m\right), \\ \text{subject to} \quad & \sum_{n\in\mathbf{N}} I_n^m \le \Gamma, \quad \text{for any } m \in \mathbf{M} \\ & \sum_{m\in\mathbf{M}} p_n^m \le \bar{P}, \quad \text{for any } n \in \mathbf{N} \\ & p_n^m \ge 0, \quad \text{for any } n \in \mathbf{N} \text{ and } m \in \mathbf{M} \end{aligned} \tag{3}$$

where Γ and $\bar{P}$ denote the interference threshold on a channel and the transmit power constraint of a DSA user, respectively. I_n^m is the interference received by PU m, which is caused by DSA user n. In the optimization problem (3), the objective function aims at maximizing all DSA users' spectral efficiency by optimizing power allocations of DSA users on each channel. The first two constraints are used to restrict the interference to a PU and the total transmit power of a DSA user, respectively. It is important to note that with the third constraint variables (transmit power) could be 0. If $p_n^m = 0$, this indicates DSA user n will not access wireless channel m.

III. System Design

Generally, no powerful infrastructure, such as base stations (BSs) or control centers, is deployed in distributed DSA networks to provide centralized control, so that DSA users have to carry out their spectrum managements individually. Moreover, a DSA user can only obtain very limited CSI, only channel states of the link between its own TX and RX. CSI of other DSA users and PUs is normally unavailable. Thus, it is difficult for DSA users to perform spectrum management through centralized resource allocation algorithms, which require accurate CSI. To facilitate DSA users to make proper spectrum management and protect PUs from harmful interference, the feedback related to the received interference should be provided by PUs, which is necessary for our developed spectrum management framework using DQN. However, DSA users and PUs may be operated by different mobile systems, which cannot connect to each other directly. Therefore, two possible interference information feedback methods of PUs are analyzed and corresponding system procedures are designed for viability.

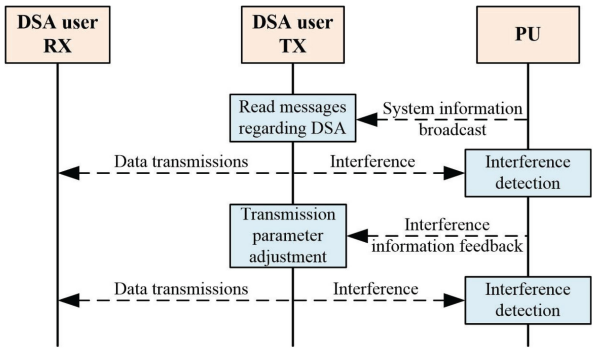


Fig. 2. System procedure of interference information feedback.

A. Procedure of Interference Information Feedback

The preliminary condition of effective information exchange between DSA users and PUs is the synchronization in time and frequency domain. In other words, DSA users need to know the frequency–time resource blocks that carry interference feedback information. Therefore, the system procedure of interference information feedback is designed. It is worth noting that since DSA users and PUs may be controlled by different wireless communication systems, the message exchange between them should be minimal to make the design interference information feedback processes less complicated and easy to realize.

Fig. 2 describes the interference information feedback process. To ensure that DSA users are aware of configurations regarding DSA, PUs should add the corresponding information in their system information (SI) and broadcast it periodically. SI is a proper carrier for DSA configurations, since SI is used to carry common control information that are fundamental and indispensable for all users to conduct wireless transmissions, and generally delivered upon fixed wireless channels [20]. Thus, in our design, when a DSA user attempts to access a frequency band, it needs to receive the SI from the corresponding PUs to read DSA configurations. DSA configurations should provide the information related to transmit power constraints and the dedicated time–frequency resources to transmit interference feedback information. According to the received DSA configurations, the DSA user carries out data transmissions. Then, PUs measure received interference caused by DSA users and feed the corresponding interference information back to DSA users through the dedicated time–frequency resources indicated in DSA configurations. Based on the interference information feedback, DSA users adjust

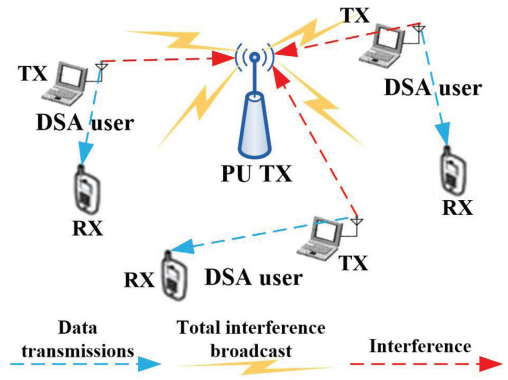


Fig. 3. Broadcast the total interference to all DSA users.

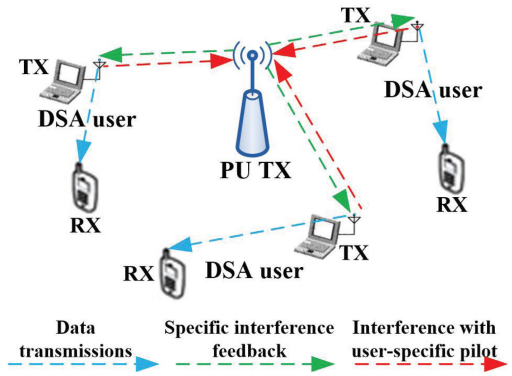


Fig. 4. Feed the specific interference back to each DSA user.

their DSA parameters to improve their own performance and guarantee PU protections.

B. Interference Information Feedback Method

There is no doubt that DSA users would be able to make the more appropriate decision on DSA parameter adjustments if they can obtain more precise interference information feedback. However, the accuracy of interference information feedback is dominated by the way that PUs measure interference from DSA users. Here, based on the designed system procedure of interference information feedback, we discuss two possible methods of PUs performing interference measurements and the corresponding interference information that can be attained by DSA users.

1) Method 1: In a general way, a PU is only able to measure the total received interference, which can be realized by sensing blank time–frequency slots embedded in their occupied channels. Then, the PU broadcasts the measurement results to DSA users. The method is presented in Fig. 3. Obviously, with this method, the overhead of interference information feedback is relatively small, while interference information that DSA users can get is really rare, only the total interference level that PUs are suffering conveyed.

2) Method 2: As shown in Fig. 4, PUs identify and detect the interference caused by each individual DSA users, and feed the specific interference level back to each DSA user. Unfortunately, PU can only receive the mixed interference signals of all DSA users sharing the same channels. To distinguish the interference signals from different DSA users, each DSA user needs to be configured with user-specific pilots, by detecting which PUs can acquire the specific interference caused by different DSA users [21]. However, in DSA networks, there is no powerful infrastructure, like BSs, to conduct centralized measurement configurations for DSA users and PUs. Therefore, a low-complexity and efficient user-specific pilot assignment method is proposed, which is described as follows. To avoid pilot contamination, the user-specific pilots of different DSA users should be transmitted on different time–frequency resource blocks. A PU includes the information of unused user-specific pilots and the corresponding time–frequency resource blocks used to send different user-specific pilots in its SIs and broadcast to all DSA users. If a DSA user attempts to access the channels occupied by this PU, it needs to receive and read the PU's SI first. Then, the DSA user randomly selects a user-specific pilot and sends the chosen user-specific pilot on the corresponding time–frequency resource blocks. The PU needs to keep monitoring the time–frequency resource blocks used for user-specific pilot transmissions. If the PU find that a user-specific pilot is transmitted on the corresponding time–frequency resource blocks, the PU should remove the user-specific pilot from its SIs, and measure the user-specific pilot to obtain interference information. In this way, interference measurements for each particular DSA user could be achieved without relying on centralized measurement configurations supported by powerful infrastructures.

Although this method is able to provide more precise interference information feedback to DSA users, considerable overhead would also be aroused. Compared to method 1, more time and frequency resources are consumed to perform user-specific pilot assignments, transmissions, and measurements.

IV. Reinforcement-Learning-Based Spectrum Management

To accomplish better performance under the condition of no centralized control and channel estimations, RL will be employed, enabling DSA users to perform spectrum management individually and intelligently.

A. Reinforcement Learning

Due to the model-free nature, RL has been applied in many fields, including wireless communications. To be specific, RL enables agents to learn environments and optimal actions according to accumulated rewards rather than labeled data or training data. In RL, the environment is defined as practical environments where a system practically conducts its operations. Therefore, the reward information is collected by taking different actions in practical environments directly. Moreover, an action may be taken multiple times to attain the knowledge of the relationship between actions and states. After fully exploring environments, the action that could bring in maximum rewards for the current state will be selected [15].

Q-learning, a basic RL method, holds the model-free attribute and low-complexity process. With Q-learning, agents learn optimal actions through directly interacting with environments without relying on any environment model information

and cooperation with other agents [22]. Besides, the reward information is represented by the Q-value, which is updated by iteratively taking various actions into environments. In Q-learning, there are two main stages, namely, exploration and exploitation. In the exploration stage, different actions should be are tried even if an action is known that it is not the optimal choice for the current state in order to fully collect the reward knowledge. On the other hand, in the exploitation stage, only the action that is expected to provide maximum rewards for the current state will be chosen for an excellent performance. Apparently, the tradeoff between exploration and exploitation is very important, both of which directly dominate the performance of Q-learning.

In some circumstances, the optimal action of the current state is selected, which could be expressed by the optimal policy π^* as

$$A_t^* = \arg \max_{A_t} Q^{\pi^*}(S_t, A_t) \tag{4}$$

where S_t and A_t^* are the current state and its optimal action, respectively.

However, to fully exploring environments and adapt to the variation of environments, the actions that have not been tried or do not have maximum rewards should be taken into environments in the exploration stage. The exploration is the key that Q-learning could be applied in dynamic systems, like DSA networks, by which Q-values could be updated according to the change of environments. Here, the ε-greedy method is employed to carry out the exploration in Q-learning, where $\varepsilon \in [0, 1]$ is a predefined probability to control the chance of randomly selecting actions or following the optimal policy π^* [23]. The ε-greedy method could be expressed as

$$A_t = \begin{cases} \arg\max_{A_t} Q^{\pi^*}(S_t, A_t), \text{ with the probability of } 1-\varepsilon, \\ \text{Randomly select actions, with the probability of } \varepsilon. \end{cases} \tag{5}$$

It is worth noting that during exploration stages PUs may suffer from considerable interference and DSA users may experience unsatisfied communication qualities, as inappropriate actions may be taken with random selections. To cope with this issue, the exploration of Q-learning should be triggered and performed only if a DSA user cannot maintain a preferable communication quality to restrict the frequency of explorations. Furthermore, PUs could be protected by adjusting the total transmit power constraint in Q-table. For example, if a PU using a wireless channel is suffering from intolerable interference, it can mandate DSA users to have stricter transmit power constraint on this channel.

After taking a selected action in environments, the Q-value of the selected action and the current state will be updated by

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha \cdot \left[R_{t+1} + \gamma \cdot \max_{A_{t+1}} Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t) \right] \tag{6}$$

where R_{t+1} is the corresponding reward of taking the selected action in environments. $\alpha \in (0, 1)$ denotes a learning rate used to control the step size of Q-value update. $\gamma \in [0, 1]$ stands for a discounted rate used to adjust the weight between the immediate reward and the future reward. For example, if the Q-value update should heavily depend on the immediate reward, a small γ should be adopted. From (6), it can be seen that the Q-value is used to measure how good an action is for a state and reflect both the immediate reward if taking an action under a state and the future expected reward.

B. Spectrum Management Using Q-Learning

Here, a Q-learning-based spectrum management scheme is proposed. The DSA will be formulated as a Q-learning problem, in which the essential components of Q-learning, including agents, states, and actions, are defined. The details of using Q-learning in spectrum management are elaborated as follows.

1) Each DSA user will be regarded as an agent that carries out Q-learning processes independently, including action selections and Q-value update.
2) The state of DSA user n is defined as the transmit power on wireless channels, which could be represented by a transmit power vector, $S = (p_1, p_2, \ldots, p_M)^T$, where p_i, $i = 1, 2, \ldots, M$, is the transmit power of the ith channel. As transmit power is a continuous value, possible states would be infinite. Here, the number of states is restricted by discretizing the transmit power into different levels. For example, assume that the total transmit power of a DSA user is limited to 300 mW, the transmit power on a single channel has four levels, which are 0, 100, 200, and 300 mW.
3) The action of a DSA user is designed as the transmit power change for each channel, which can also be denoted by a vector, $A = (a_1, a_2, \ldots, a_M)^T$, where a_i, $i = 1, 2, \ldots, M$, is the adjustment of the ith channel's transmit power. Similarly, the amount of actions should be restricted. Accordingly, only three types of actions will be selected, namely, increasing transmit power to the next higher level, decreasing transmit power to the next lower level, and no change, notated by *In*, *De*, and *Un*, respectively.

Based on the configuration of Q-learning, the corresponding Q-table is designed in Table I, where the total transmit power constraint, the number of channels, and the number of transmit power levels are assumed to be 300 mW, 2, and 4, respectively. For fairness, each DSA user accesses at least one channel, and accordingly the transmit power of at least one channel is nonzero. It is noticeable that for an initial state, some actions should avoid been selected, since taking these actions will make the next state unavailable or out of the scope of defined states. For example, at state 7 (100 mW, 200 mW), actions 5, 6, and 9 should not be chosen, which makes the total transmit power exceeds the constraint. Actions 2, 4, and 8 are inapplicable for state 1 (100 mW, 0 mW), after taking which the transmit power of the second channel becomes a negative value (−100 mW). Therefore, for feasibility, some operating mechanisms need to be designed to tackle these issues, which are described as follows.

TABLE I
Q-TABLE IN A DSA USER

	$A1:(Un,Un)$	$A2:(Un,De)$	$A3:(De,Un)$	$A4:(De,De)$	$A5:(Un,In)$	$A6:(In,Un)$	$A7:(De,In)$	$A8:(In,De)$	$A9:(In,In)$
$S1:(100mW,0mW)$	$Q(S1,A1)$	LR	LR	LR	$Q(S1,A5)$	$Q(S1,A6)$	$Q(S1,A7)$	LR	$Q(S1,A9)$
$S2:(0mW,100mW)$	$Q(S2,A1)$	LR	LR	LR	$Q(S3,A5)$	$Q(S2,A6)$	LR	$Q(S3,A8)$	$Q(S2,A9)$
$S3:(100mW,100mW)$	$Q(S3,A1)$	$Q(S3,A2)$	$Q(S3,A3)$	LR	$Q(S3,A5)$	$Q(S3,A6)$	$Q(S3,A7)$	$Q(S3,A8)$	$Q(S3,A9)$
$S4:(200mW,0mW)$	$Q(S4,A1)$	LR	$Q(S4,S3)$	LR	$Q(S4,A5)$	$Q(S4,A6)$	$Q(S4,A7)$	LR	LR
$S5:(0mW,200mW)$	$Q(S5,A1)$	$Q(S5,A2)$	LR	LR	$Q(S5,A5)$	$Q(S5,A6)$	LR	$Q(S6,A8)$	LR
$S6:(200mW,100mW)$	$Q(S6,A1)$	$Q(S6,A2)$	$Q(S6,A3)$	$Q(S6,A4)$	LR	LR	$Q(S6,A7)$	$Q(S6,A8)$	LR
$S7:(100mW,200mW)$	$Q(S7,A1)$	$Q(S7,A2)$	$Q(S7,A3)$	$Q(S7,A4)$	LR	LR	$Q(S7,A7)$	$Q(S7,A8)$	LR
$S8:(300mW,0mW)$	$Q(S8,A1)$	LR	$Q(S8,A3)$	LR	LR	LR	$Q(S8,A7)$	LR	LR
$S9:(0mW,300mW)$	$Q(S9,A1)$	$Q(S9,A2)$	LR	LR	LR	LR	LR	$Q(S9,A8)$	LR

1) The corresponding Q-values of the inappropriate actions should be set to a very small value, represented by *LR* in Table I, to reduce the chance of choosing these actions.
2) If an action will result in the transmit power vector unable to match any state in the Q-table, the initial state is adopted as the next state. For example, after taking action 6 with the state 6 (200 mW, 100 mW) as the initial state, the transmit power vector will be changed to (300 mW, 100 mW), which is unavailable in the Q-table. Therefore, state 6 will be adopted as the next state.

Obviously, the proposed spectrum management based on Q-learning can provide both spectrum access and power allocation strategies for DSA users. A DSA user will access a wireless channel only if the transmit power of this channel in the state is not equal to 0. Moreover, the state expressed by the transmit power vector could indicate power allocation strategies directly.

C. Definition of Reward

The definition of reward directly determines the performance of spectrum management, which should consider both data rate enhancement and PU protections. In Section III, two potential interference information feedback methods are discussed for future DSA networks, based on which the reward used in Q-learning will be defined.

If method 1 as shown in Fig. 3 is applied, PUs are merely able to broadcast total received interference to all the DSA users, the reward of DSA user n is defined as

$$R_n = \sum_{m\in\Omega_n} \log_2\left(1+\frac{\left|h_{nn}^m\right|^2\cdot p_n^m}{\left|h_{mn}^m\right|^2\cdot p_m^m+\sum\limits_{j\in\Phi_m, j\neq n}\left|h_{jn}^m\right|^2\cdot p_j^m+B\cdot N_0}\right) - \kappa\cdot\sum_{m\in\Omega_n} e^{\frac{I^m}{\hat{I}^m}} \tag{7}$$

where the first term and the second term are spectral efficiency and the penalty caused by the interference to PUs, respectively. The penalty is determined by κ. I^m and $\hat{I}^m$ are the total interference suffered by the PU m and the reference interference level, respectively. The penalty will exponentially increase with the growth of I^m. κ is a weight to adjust the influence of the penalty on the reward. Apparently, for a DSA user, the only information needed to calculate the defined reward is the interference feedback from PUs, while its spectral efficiency can be monitored by itself.

Under method 2, a DSA user can obtain more detailed interference information feedback from PUs, namely, the specific interference caused by it. Accordingly, the reward of DSA user n is defined as

$$R_n = \sum_{m\in\Omega_n} \log_2\left(1+\frac{\left|h_{nn}^m\right|^2\cdot p_n^m}{\left|h_{mn}^m\right|^2\cdot p_m^m+\sum\limits_{j\in\Phi_m, j\neq n}\left|h_{jn}^m\right|^2\cdot p_j^m+B\cdot N_0}\right) - \kappa\cdot\sum_{m\in\Omega_n} e^{\frac{I_n^m}{\hat{I}_n^m}} \tag{8}$$

where I_n^m is the interference received by PU m, which is caused by DSA user n. $\hat{I}_n^m$ denotes the reference interference level of DSA user n on channel m.

For feasibility and practicability, the feedback processes of spectral efficiency information and the interference to PUs need to be designed. Spectral efficiency information could be fed back to DSA TXs along with channel measurement and acknowledgment (ACK) feedback, which are used to preserve communication quality and assist packet retransmissions, respectively. On the other hand, a feasible and simple way to achieve interference feedback is that PUs broadcast interference information on their fixed broadcasting channels with fixed modulation and coding schemes, so that DSA users can retrieve interference information by listening PUs' broadcast channels periodically. Since interference information normally only comprises limited data, the corresponding overhead to both PUs and secondary users can be restricted in a low level. Relatively, the interference information feedback method 2 will cause more severe overhead, where more broadcasting resources need to be occupied, as more detailed and specific interference information is fed back. Besides, method 2 has more complicated system procedures especially in interference measurement, which has been designed and described in Section III. However, method 2 can provide more precise interference information feedback to each DSA user, facilitating more efficient and accurate Q-learning process. To alleviate the overhead of interference information feedback, unnecessary feedback should be avoided. For example, a PU will feed interference information back to DSA users only when its received interference is measured to be changed. For a DSA user, if it does not receive a new interference feedback, the DSA user will keep using the latest interference feedback to calculate the reward.

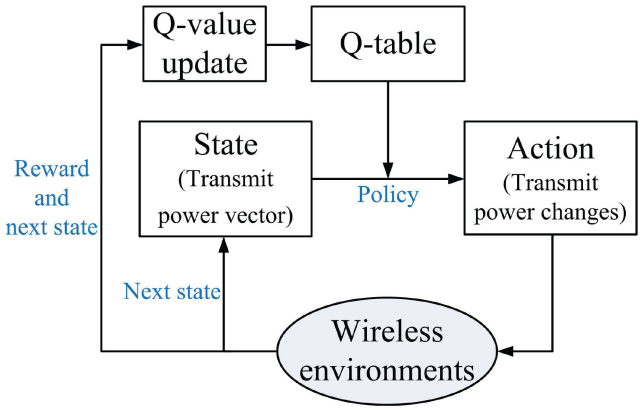


Fig. 5. Q-learning-based spectrum management.

D. Process of Q-Learning-Based Spectrum Management

Based on the aforementioned design, the system procedure of the spectrum management using Q-learning is described as follows. As shown in Fig. 5, a DSA user selects an action (transmit power changes) according to its current state (its transmit power vector) and Q-table, as well as the applied policy expressed by (5). With the chosen action, the DSA user adjusts its transmit power and updates its transmit power vector, based on which wireless transmissions are performed in wireless environments. Then, the updated transmit power vector will be used as the next state, and a reward is calculated based on the performance of its wireless transmission and the interference information feedback from PUs according to (7) or (8). Finally, the next state substitutes the initial one to be the current state of the DSA user, and the Q-value related to the initial state and the selected action in the Q-table is updated according to (6). Obviously, the training of Q-learning could be integrated into the real wireless transmissions of DSA users, which does not require any extra training process or training data with limited system overhead, low complexity, and high feasibility.

V. Deep Q-Network-Based Spectrum Management

Since Q-learning is processed by Q-value updates, Q-learning is not able to handle the Q-table with a large size. The large size of Q-table makes Q-learning very hard or even impossible to converge [16]. Thus, powerful NNs will be utilized to address this issue and support efficient Q-learning processes, also referred to as DQN.

A. Deep Q-Network

The DQN process of a DSA user is illustrated in Fig. 6, in which two NNs are utilized, including an evaluated NN and a target NN. It is assumed that the initial state is S_t and there are totally L actions in Q-table. After inputting the initial state S_t into the evaluated NN, the Q-values of S_t with respect to all the actions, $Q(S_t, A1), Q(S_t, A2), \ldots, Q(S_t, AL)$, will be output. Afterward, according to the applied policy, an action A_t is chosen, which is taken to updated the transmit power vector and determine the next state S_{t+1}. Then, the DSA user carries out wireless transmissions in environments based on the updated transmit power vector and obtain the reward according to the

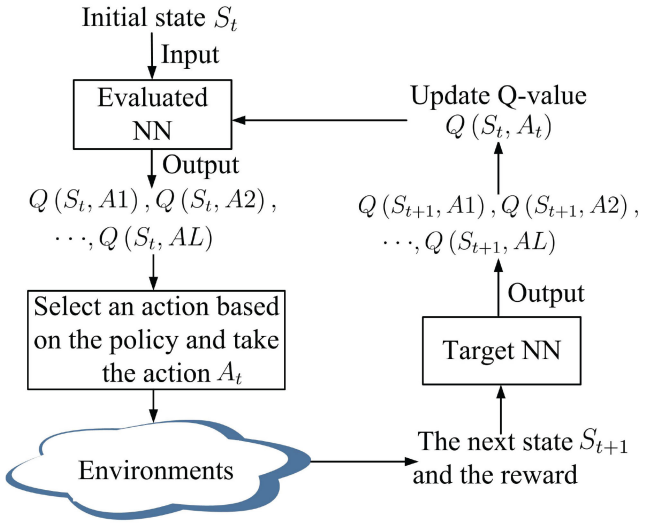


Fig. 6. Iteration of DQN.

reaction of environments, including its spectrum efficiency and interference feedback from PUs. The next state S_{t+1} is input into the target NN to generate Q-values of S_{t+1} with respect to all the actions, $Q(S_{t+1}, A1), Q(S_{t+1}, A2), \ldots, Q(S_{t+1}, AL)$. With the generated Q-values of S_{t+1} and the obtained reward, the Q-value of the initial state S_t and the selected action A_t, $Q(S_t, A_t)$, is updated according to (5). The updated Q-value $Q(S_t, A_t)$ will be regarded as a target value used in the training of the evaluated NN with the backpropagation method. The target NN is updated periodically by employing the evaluated NN as a new target NN [17], [24].

B. Selection of Neural Networks

The selection of NNs is crucial for the performance of DQN, which should be chosen based on application scenarios. Feedforward NNs (FFNNs) have been widely applied in various fields, which possess a simple structure and are easy to be trained [25]. Nevertheless, RNNs may be able to bring in better performance for DSA networks due to their temporal correlation attribute. In an RNN, activation values of neurons are determined not only by current input data but also by previous activation values of recurrent neurons and output neurons. These feedback connections make RNNs capable of capturing temporal correlations, which are very useful and meaningful in a system with dynamic environments [25], [26]. For instance, a typical application of RNN is the natural language processing, as languages normally need to be comprehended considering both current words and previous words [27]. In DSA networks, it is better for DSA users to make the decision on spectrum management based on a series of environment status rather than the environment at the current moment. However, the complicated structure makes RRNs very hard to train [16]. To cope with that, ESNs, a simplified type of RNNs, is introduced, the structure of which is plotted in Fig. 7. In an ESN, the weights of input and reservoir layers are predefined and fixed, while only the weights of an output layer need to be trained [28], [29]. As a result, compared to traditional RNNs,

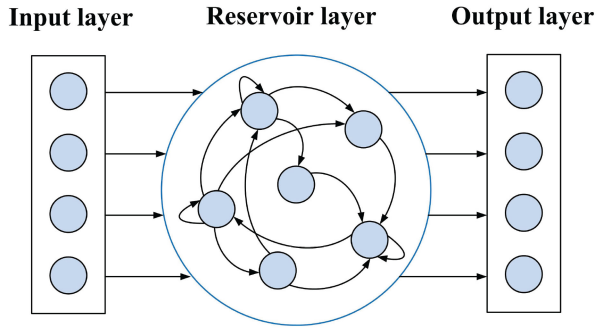


Fig. 7. Echo state network.

ESNs are easier to train, meanwhile, the temporal correlation attribute could be preserved.

VI. Simulation Results and Analysis

Through simulations, the performance of our proposed spectrum management scheme is investigated. Additionally, the optimal way to constitute DQN is studied through simulations, which will take into account both system performance and the convergence of machine learning/deep learning methods.

A. Simulation Setup

In the simulation, a DSA network consisting of two wireless channels and four DSA users is considered. Moreover, four DSA users and four PUs are randomly distributed in a 150 m $\times$ 150 m area. For fairness, it is assumed that each DSA user is at least access one channel and the transmit power constraint for a DSA user is 300 mW. The WINNER II channel model and the Rician channel model are adopted to calculate channel gains [30]. According to the aforementioned analysis, the ε in the ε-greedy method is a critical parameter, which dominates the tradeoff between exploration and exploitation of Q-learning or DQN. In the simulation, the total number of training is 8000, in which 4000 times training is used for exploration with a relatively large ε, 0.5, facilitating DSA users to sufficiently explore all the possible spectrum management strategies. Then, ε will be adjusted to be 0 to let DSA users select their spectrum management strategies with optimal rewards. The detailed simulation parameters are listed in Table II. For comparison, Q-learning, FFNN-based DQN, and ESN-based DQN will be used to simulate the proposed spectrum management scheme, respectively. All the simulations are conducted by Python and Tensorflow is utilized to execute the training of NNs. For the Q-value update in (6), the learning rate α and the discounted rate γ are set to be 0.01 and 0.9, respectively.

B. Performance With the Total Interference Broadcast

First, the performance of the proposed spectrum management scheme is investigated under the condition that only the total interference broadcast can be provided by PUs as shown in Fig. 3. Accordingly, the reward of a DSA user is given by (7) and the reference interference level $\hat{I}^m$ is set to 8×10^{-6} mW. Fig. 8 presents the total reward, the summation of all DSA users' rewards, versus training steps. Obviously,

TABLE II
Simulation Parameters

Parameters	Values
Transmit power of PUs	400mW
Transmit power constraint of DSA users	300mW
Channel bandwidth B	2MHz
Noise spectral density N_0	-174dBm/Hz
Center frequency	5GHz
Path-loss model (WINNER II)	$41 + 22.7 \cdot \log_{10}(d[m])$ $+20 \cdot \log_{10}(f_c[GHz]/5)$
K-factor	8
Penalty weight κ	0.3

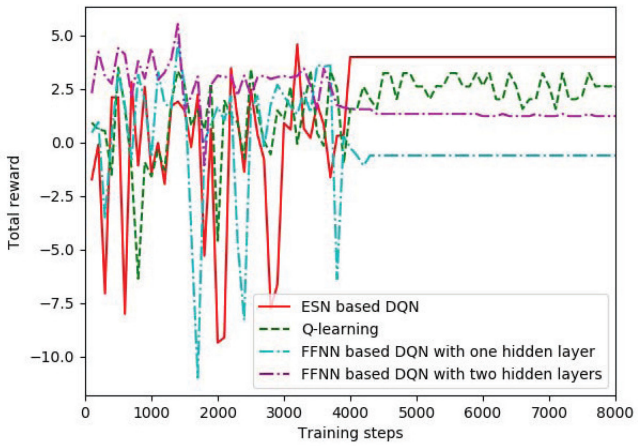


Fig. 8. Total reward with the total interference broadcast.

ESN-based DQN has a better performance on the reward, as owning to the temporal correlation nature of ESN, DSA users can learn dynamic wireless environments better and make the more appropriate decision on spectrum management. Besides, it can be seen that after the exploration stage the total reward of ESN-based DQN becomes stable, indicating the excellent convergence behaviors of ESN-based DQN.

Fig. 9 illustrates the total data rate of all DSA users with the unit of Mbits/s. Due to no centralized control, each DSA user attends to acquire more benefits in the competition with others. As a result, when a DSA user is experiencing the low reward, it may take the action of raising transmit power to boost its data rate. However, afterward, the DSA user may encounter more severe interference from other DSA users, causing low data rate. This is because the high transmit power of a DSA user will incur more serious interference to other DSA users, which may also use the same method of rising their transmit power to preserve communication quality. Hence, the spectrum management scheme should enable DSA users to reach a balance on transmit power rather than unboundedly raising transmit power against serious interference. It can be observed from Fig. 9 that ESN-based DQN is able to let DSA users reach a balance fast. In addition, spectrum management with ESN-based DQN could bring in higher data rate, indicating that an excellent balance is achieved between DSA users.

Fig. 10 shows the simulation results of total interference to PUs caused by DSA users. According to (7), the interference is regarded as a penalty in the defined reward utility. DSA users are encouraged to lower their transmit power. Apparently, with

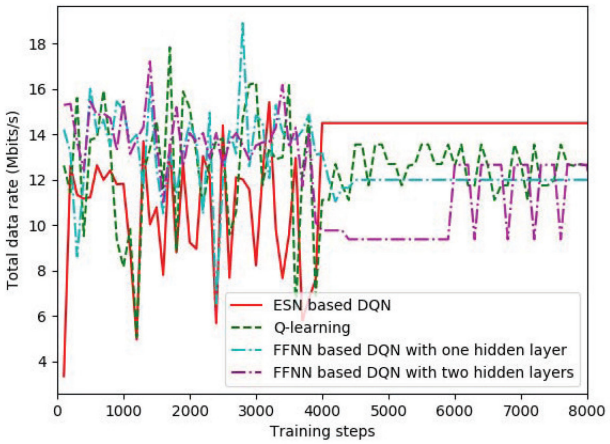


Fig. 9. Total data rate with the total interference broadcast.

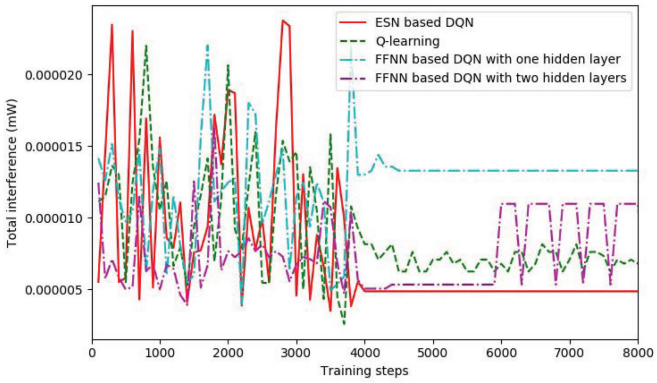


Fig. 10. Total interference with the total interference broadcast.

ESN-based DQN, DSA users are capable of effectively suppressing the interference to PUs in a relatively low level. The reason is that the powerful ESN could enable DSA users to learn the interference tolerable level of PUs through interacting with environments and the received reward, so that more proper spectrum managements are performed to protect PUs from detrimental interference.

C. Performance With the Specific Interference Feedback

We also study the performance of the proposed spectrum management when DSA users can get more accurate interference feedback from PUs as shown in Fig. 4. In this case, the reward is calculated according to (8) and the reference interference level $\hat{I}_n^m$ is set to 2×10^{-6} mW. Figs. 11–13 show the simulation results of total reward, the total data rate, and the total interference, respectively. It is easy to observe that ESN-based DQN can converge immediately once stepping into the exploitation stage, since ESN promotes the environment learning ability of users and an excellent balance between users can rapidly be achieved. Additionally, ESN-based DQN possesses the higher reward and the lower total interference than other methods. It is noted that in Fig. 12 the total data rate of ESN-based DQN is lower than that of FFNN-based DQN with two hidden layers. This phenomenon manifests that ESN-based DQN can make better use of interference information feedback from PUs when the feedback is more specific and

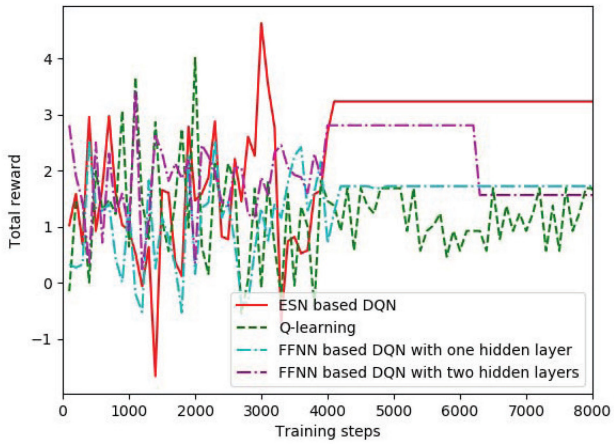


Fig. 11. Total reward with the specific interference feedback.

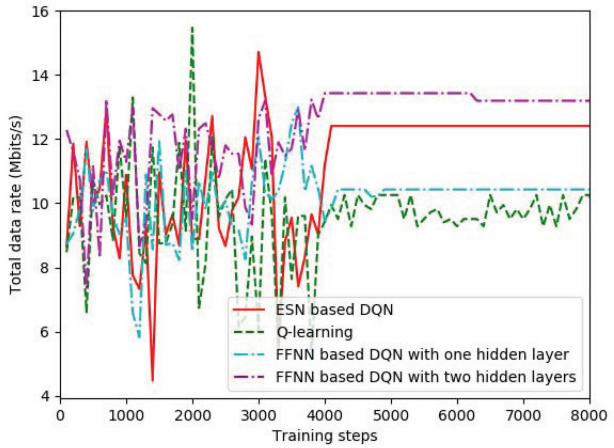


Fig. 12. Total data rate with the specific interference feedback.

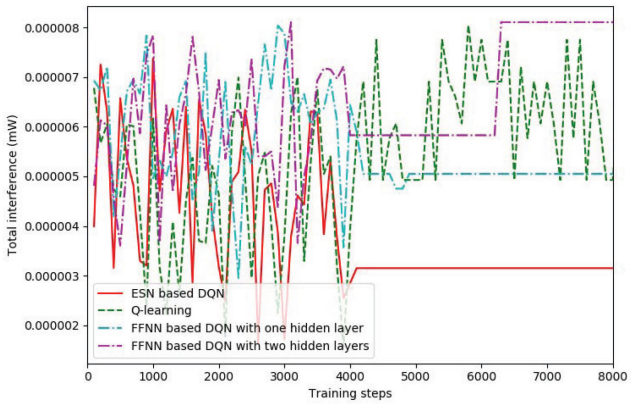


Fig. 13. Total interference with the specific interference feedback.

detailed. For the reward enhancement, ESN-based DQN mitigates the interference to PUs by reducing transmit power and sacrificing the data rate.

D. Optimality Evaluation

To investigate the optimality of the proposed framework in this article, centralized optimization is employed as a compared method for performance comparison. Through centralized optimization, global optimal or suboptimal can be achieved, which could be deemed as "upper bound" for the

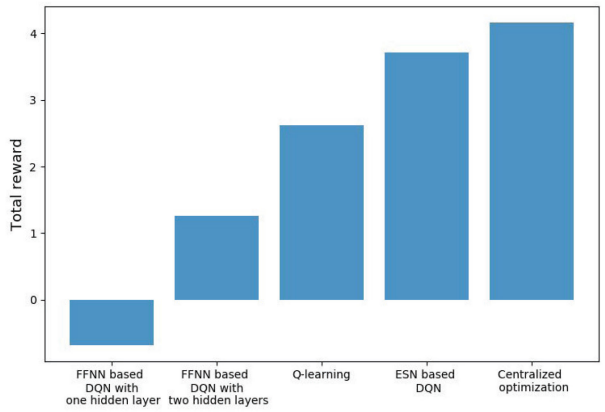


Fig. 14. Optimality evaluation through total reward.

proposed framework. Note that it is impossible to realize centralized optimization in distributive DSA, which requires powerful infrastructures, centralized control, and accurate and instantaneous channel estimations. The simulation results of the centralized optimization are obtained by solving the optimization problem (3), where Γ and $\bar{P}$ are set to be 8×10^{-6} and 300 mW, respectively, same with the setup of the proposed framework. As the optimization problem (3) is not convex, the iterative water filling is used to solve the optimization problem [31].

Fig. 14 shows the simulation results of our proposed spectrum management framework and the compared method on the defined reward, which can reflect the performance of both spectral efficiency of DSA users and the interference to PUs. The reward is calculated according to (7). For the proposed framework, the simulation is conducted under the condition of PUs broadcasting total interference to all DSA users. Besides, the simulation results are the average values of 1000 times training of exploitation after exploration. Note that a negative value will be represented by a bar below 0 in Fig. 14. It can be observed that the centralized optimization has the best performance on the total reward compared to others. Apart from global optimal/suboptimal achieved by centralized optimization, another reason is that with centralized optimization each DSA user can be served with accurate power allocation leading to global optimal/suboptimal. However, in distributive DSA, power allocation is rougher, where the transmit power is discretized into different levels and a DSA user can only transmit with one of these levels to limit the size of Q-learning states. Clearly, the performance of the proposed framework using ESN-based DQN is closest to that of the centralized optimization, indicating that it is able to enable an excellent balance between different DSA users and effective PU protection, resulting in a better system performance.

VII. Conclusion

In a DSA network, the spectrum management is very challenging, as lack of centralized control makes DSA users have to carry out the spectrum management independently. To enable effective spectrum management in DSA, in this article, a spectrum management scheme leveraging Q-learning is proposed, in which DSA users carry out their spectrum management, including both spectrum access and power allocation, by directly interacting with environments without depending on channel estimations and training date. However, a new challenge arises that Q-learning is not able to handle a large Q-table size. In other words, when a DSA network consists of a large amount of wireless channels and DSA users, Q-learning is very hard to be trained, causing instability. Thus, the ESN, a type of RNN, is adopted to realize Q-learning, named the DQN, for better performance, efficient training, and fast convergence. Through extensive simulation studies, it has been proven that the proposed spectrum management scheme with ESN-based DQN can achieve the higher reward with both the achievable data rate and PU protections considered. In addition, using ESN in the proposed scheme has better convergence behaviors.

Acknowledgment

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of AFRL. This article is an extension work of our conference paper [32].

References

[1] "Cisco visual networks index: Global mobile data traffic forecast update, 2017–2022 white paper," San Jose, CA, USA, CISCO, Whitepaper, Feb. 2019.

[2] *Spectrum Occupancy Measurements*, Shared Spectr. Company, Vienna, VA, USA, 2007. [Online]. Available: http://www.sharedspectrum.com/wp-content/uploads/Loring_Spectrum_Occupancy_Measurements_v2_3.pdf

[3] "Spectrum policy task force," Federal Commun. Comm., Washington, DC, USA, Rep. ET Docket 02–135, Nov. 2002.

[4] H. Song, J. Bai, Y. Yi, J. Wu, and L. Liu, "Artificial intelligence enabled Internet of Things: Network architecture and spectrum access," *IEEE Comput. Intell. Mag.*, vol. 15, no. 1, pp. 44–51, Feb. 2020.

[5] H. Song, X. Fang, L. Yan, and Y. Fang, "Control/user plane decoupled architecture utilizing unlicensed bands in LTE systems," *IEEE Wireless Commun.*, vol. 24, no. 5, pp. 132–142, Oct. 2017.

[6] H. Song, X. Fang, and C.-X. Wang, "Cost-reliability tradeoff in licensed and unlicensed spectra interoperable networks with guaranteed user data rate requirements," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 1, pp. 200–214, Jan. 2017.

[7] H. Song and X. Fang, "A spectrum etiquette protocol and interference coordination for LTE in unlicensed bands (LTE-U)," *Proc. IEEE Int. Conf. Commun. Workshop (ICCW)*, London, U.K., Jun. 2015, pp. 2338–2343.

[8] *Fact Sheet: Spectrum Frontiers Rules Identify, Open Up Vast Amounts of New High-Band Spectrum for Next Generation (5G) Wireless Broadband*, Federal Commun. Comm., Washington, DC, USA, 2016. [Online]. Available: http://transition.fcc.gov/Daily_Releases/Daily_Business/2016/db0714/DOC-340310A1.pdf

[9] S. Bhattarai, J.-M. J. Park, B. Gao, K. Bian, and W. Lehr, "An overview of dynamic spectrum sharing: Ongoing initiatives, challenges, and a roadmap for future research," *IEEE Trans. Cogn. Commun. Netw.*, vol. 2, no. 2, pp. 110–128, Jun. 2016.

[10] B. Jabbari, R. Pickholtz, and M. Norton, "Dynamic spectrum access and management [dynamic spectrum management]," *IEEE Wireless Commun.*, vol. 17, no. 4, pp. 6–15, Aug. 2010.

[11] H. Song, X. Fang, and Y. Fang, "Unlicensed spectra fusion and interference coordination for LTE systems," *IEEE Trans. Mobile Comput.*, vol. 15, no. 12, pp. 3171–3184, Dec. 2016.

[12] M. Tang, M. Vehkapera, X. Chu, and R. Wichman, "LI cancellation and power allocation for multipair FD relay systems with massive antenna arrays," *IEEE Wireless Commun. Lett.*, vol. 8, no. 4, pp. 1077–1081, Aug. 2019.

[13] Y. Wang, A. Klautau, M. Ribero, A. C. K. Soong, and R. W. Heath, "MmWave vehicular beam selection with situational awareness using machine learning," *IEEE Access*, vol. 7, pp. 87479–87493, 2019.

[14] S. Xu, P. Liu, R. Wang, and S. S. Panwar, "Realtime scheduling and power allocation using deep neural networks," *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, 2019, Marrakesh, Morocco, pp. 1–5.

[15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, Nov. 2017.

[16] R. Pascanu, T. Mikolov, and Y. Bengio, "On the difficulty of training recurrent neural networks," *Proc. ICML*, Feb. 2013, pp. 1310–1318.

[17] M. Sewak, "Deep Q-network (DQN), double DQN, and dueling DQN," in *A Step Towards General Artificial Intelligence*. Singapore: Springer, Jun. 2019, pp. 95–108.

[18] H. H. Chang, H. Song, Y. Yi, J. Zhang, H. He, and L. Liu, "Distributive dynamic spectrum access through deep reinforcement learning: A reservoir computing-based approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 1938–1948, Apr. 2019.

[19] H. H. Chang, L. Liu, and Y. Yi, "Deep echo state Q-network (DEQN) and its application in dynamic spectrum sharing for 5G and beyond," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Nov. 2, 2020, doi: 10.1109/TNNLS.2020.3029711.

[20] *A Radio Resource Control (RRC) Protocol Specification, v10.0.0*, 3GPP Standard TS 36.331, Mar. 2013.

[21] *Physical Layer Measurements, v11.1.0*, 3GPP Standard TS 36.214, Dec. 2012.

[22] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, pp. 279–292, May 1992.

[23] E. R. Gomes and R. Kowalczyk, "Dynamic analysis of multiagent Q-learning with ε-greedy exploration," *Proc. 26th Annu. Int. Conf. Mach. Learn. (ICML)*, Jun. 2009, pp. 369–376.

[24] H. V. Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double Q-learning" *Proc. 30th AAAI Conf. Artif. Intell. (AAAI)*, 2016, pp. 2094–2100.

[25] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[26] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms Architecture and Stability*. New York, NY, USA: Wiley, 2001.

[27] T. Mikolov, M. Karafiát, L. Burget, J. Cernocky, and S. Khudanpur, "Recurrent neural network based language model," *Proc. Inter-Speech*, Sep. 2010, pp. 1045–1048.

[28] M. Lukosevicius, "A practical guide to applying echo state networks," *Neural Networks: Tricks of the Trade*, Berlin, Germany: Springer, 2012, pp. 659–686.

[29] Z. Zhou, L. Liu, V. Chandrasekhar, J. Zhang, and Y. Yi, "Deep reservoir computing meets 5G MIMO-OFDM systems in symbol detection," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, 2020, pp. 1266–1273.

[30] P. Kyosti *et al.*, "IST-4-027756 WINNER II channel models D1.1.2, V1.2," EBITG, TUI, UOULU, CU/CRC, NOKIA, Espoo, Finland, Rep., Sep. 2007.

[31] Z. Wang, V. Aggarwal, and X. Wang, "Iterative dynamic water-filling for fading multiple-access channels with energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 382–395, Mar. 2015.

[32] H. Song, L. Liu, H. Chang, J. Ashdown, and Y. Yi, "Deep Q-network based power allocation meets reservoir computing in distributed dynamic spectrum access networks," *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Paris, France, 2019, pp. 774–779.

Hao Song received the B.E. degree in electronic information engineering from Zhengzhou University, Zhengzhou, China, in 2011, and the Ph.D. degree in information and communication engineering from Southwest Jiaotong University, Chengdu, China, in 2018. He is currently pursuing the Ph.D. degree with the Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, USA.

His current research interests include 5G networks, swarm UAV networks, networking, machine learning and its applications in wireless communications, resource allocation, optimization, and dynamic spectrum access.

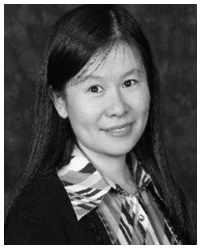
Lingjia Liu (Senior Member, IEEE) is an Associate Professor with the Bradley Department of Electrical Engineering and Computer Engineering, Virginia Tech (VT), Blacksburg, VA, USA. He is also the Associate Director of Wireless@VT. Prior to joining VT, he was an Associate Professor with the EECS Department, University of Kansas (KU), Lawrence, KS, USA. He spent more than four years working in the Mitsubishi Electric Research Laboratory and the Standards and Mobility Innovation Laboratory, Samsung Research America (SRA). He was leading Samsung's efforts on multiuser MIMO, CoMP, and HetNets in LTE/LTE-Advanced standards. His general research interests mainly lie in emerging technologies for beyond 5G cellular networks, including machine learning for wireless networks, massive MIMO, massive MTC communications, and mmWave communications.

Dr. Liu received the Global Samsung Best Paper Award in 2008 and 2010 from SRA, the Air Force Summer Faculty Fellow from 2013 to 2017, the Miller Scholar at KU in 2014, the Miller Professional Development Award for Distinguished Research at KU in 2015, the 2016 IEEE GLOBECOM Best Paper Award, the 2018 IEEE ISQED Best Paper Award, the 2018 IEEE TAOS Best Paper Award, the 2018 IEEE TCGCC Best Conference Paper Award, and the 2020 WOCC Charles Kao Best Paper Award.

Jonathan Ashdown (Member, IEEE) was born in Niskayuna, NY, USA, in 1984. He received the B.S., M.S., and Ph.D. degrees in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2006, 2008, and 2012, respectively.

His Ph.D. dissertation was on a high-rate ultrasonic through-wall communication system using MIMO-OFDM in conjunction with interference mitigation techniques. From 2012 to 2015, he worked as an Electronics Engineer with the Department of Defense (DoD), SPAWAR Systems Center Atlantic, Charleston, SC, USA, where he was involved in several basic and applied research projects for the U.S. Navy, mainly in the area of software-defined radio. In 2015, he transferred within DoD. He is currently a Senior Electronics Engineer with Air Force Research Laboratory, Rome, NY, USA, where he is involved in the research and development of advanced emerging communications and networking technologies for the U.S. Air Force.

Dr. Ashdown was a recipient of the Best Unclassified Paper Award at the IEEE Military Communications Conference in 2012.



Yang Yi received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiao Tong University, Shanghai, China, and the Ph.D. degree in computer engineering from Texas A&M University, College Station, TX, USA.

She is an Associate Professor with the Department of Electrical and Computer Engineering, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA, where she is also serving as the Director of Multifunctional Integrated Circuits and Systems Center. Her current research interests include very-large-scale integrated circuits and systems, computer-aided design, and neuromorphic computing.