Intelligent Beam Management Based on Deep Reinforcement Learning in High-Speed Railway Scenarios

Yuanyuan Qiao, *Graduate Student Member, IEEE*, Yong Niu, *Senior Member, IEEE*, Xiangfei Zhang, Sheng Chen, *Life Fellow, IEEE*, Zhangdui Zhong, *Fellow, IEEE*, Ning Wang, *Member, IEEE*, and Bo Ai, *Fellow, IEEE*

***Abstract*—Millimeter-wave (mm-wave) communications can fundamentally solve the problem of spectrum shortage in wireless communication systems, and many progresses have been made in standardization, which laid the foundation for the application of mm-wave in high-speed railway (HSR) scenarios. However, the HSR channel is fast time-varying and difficult to model. Also beamforming is essential to improve the directional gain of the antenna and offset the high path loss of mm-wave. But the high-speed movement of train makes the beam management extremely challenging, and the trade-off between achievable performance and beam training overhead is unavoidable. Reinforcement learning (RL) can offer new solutions to this problem, as it does not need a large number of training samples and other system information, and is capable of achieving high performance with low complexity. In this paper, we propose an intelligent beam management scheme based on a deep RL algorithm called deep Q-network (DQN), and our main idea is to exploit the hidden patterns of mm-wave train-ground communication system to improve the downlink signal-to-noise ratio (SNR), while ensuring a certain communication stability and imposing a minimal training overhead. Through extensive simulations, we demonstrate that the proposed DQN-based scheme has better performance than the four baseline schemes, and it also offers great advantages in SNR stability and implementation complexity.**

***Index Terms*—Train-ground communications, high-speed railway, millimeter-wave communications, beam management, deep reinforcement learning**

 This work was supported by the National Key Research and Development Program of China under Grant 2021YFB2900301; in part by the National Key Research and Development Program of China under Grant 2020YFB1806903; in part by the National Natural Science Foundation of China under Grant 62221001, Grant 62231009, Grant U21A20445; This study was supported by the Fundamental Research Funds for the Central Universities, China, under grant number 2022JBQY004 and 2022JBXT001; in part by the Fundamental Research Funds for the Central Universities 2023JBMC030. The review of this article was coordinated by Rongqing Zhang. (*Corresponding authors: Y. Niu, B. Ai.*)

Y. Qiao is with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, and also with Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications, Beijing Jiaotong University, Beijing 100044, China (email: qiaoyuanyuan@bjtu.edu.cn).

Y. Niu is with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing 100044, China (E-mail: niuy11@163.com).

X. Zhang is with the State Key Laboratory of Advanced Rail Autonomous Operation, and also with the Frontiers Science Center for Smart High-speed Railway System, Beijing Jiaotong University, Beijing 100044, China (E-mail: zxfei78@163.com).

S. Chen is with School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, U.K. (E-mail: sqc@ecs.soton.ac.uk).

Z. Zhong, B. Ai are with the State Key Laboratory of Advanced Rail Autonomous Operation, and also with the Beijing Engineering Research Center of High-speed Railway Broadband Mobile Communications, Beijing Jiaotong University, Beijing 100044, China (e-mails: zhdzhong@bjtu.edu.cn; boai@bjtu.edu.cn).

N. Wang is with the School of Information Engineering, Zhengzhou University, Zhengzhou 450001, China (e-mail: ienwang@zzu.edu.cn).

I. Introduction

A. Motivation

Looking back at the past decade, the rapid development of high-speed rail (HSR) has driven many technological innovations and changed people's lives. By the end of 2020, the operating mileage of Chinese HSR has reached 39,000 kilometers, ranking the first in the world [1]. Since many passengers are accustomed to broadband wireless access in their daily living, more and more people hope to have high-quality broadband wireless access on mobile terminals when traveling by train. At present, the demand rate of passengers in each carriage is about 37.5 Mbps, and with the growth of business volume and quality of service (QoS), it may reach 0.5-5 Gbps in the future [2]. But the wireless transmission scheme now widely used in HSR scenarios, such as GSM-R, obviously cannot meet these requirements. Therefore, it is important to carry out relevant research on broadband wireless communication technology in HSR scenarios.

At present, the research on HSR communications mainly includes two categories: intra-train and train-ground communications. In order to overcome the huge penetration loss caused by the metal body of carriage, it is usually considered to deploy a mobile relay (MR) on the roof of the train and a wireless access point (AP) inside the carriage. The AP collects users' data and transmit them to the track-side base station (BS) through the MR [3]. At this time, analyzing the intra-train communications is similar to the indoor communications. But the problems inherent in HSR communications related to handover, Doppler spread, non-stationary and fast time-varying are not solved [4]. Therefore, these issues related to train-ground communications are the main factors that affects the communication performance in HSR scenarios.

Consistent with the main problems of other wireless communication systems, the current performance of the train-ground communications in HSR scenarios is mainly limited by spectrum resources. For example, the available spectrum resources of GSM-R are only 4 MHz, but it needs to deal with emergency calls, train scheduling and so on, leaving very little to meet the need of passengers. The millimeter-wave (mm-wave) frequency band located at 30-300 GHz, which has abundant spectrum resources, can fundamentally solve this problem and provide passengers with high-capacity broadband wireless access [5]. At the same time, the rapid development of CMOS radio frequency integrated circuits has also promoted the production of related electronic devices and

the formulation of standards [6], making it possible to apply mm-wave communications in HSR scenarios.

However, mm-wave communication suffers from much higher path loss than its low-frequency counterpart. For example, in free space, the path loss in 60 GHz is 28 dB higher than that in 2.4 GHz [7]. Furthermore, the rain attenuation and atmospheric absorption of mm-wave signals further reduce its effective propagation distance. In order to solve this problem, directional antennas are generally used in mm-wave communications to realize beamforming, thereby increasing the antenna gain. In addition, due to the much shorter wavelength, a large number of antenna elements can be integrated in a small area, and a large-scale antenna array can be deployed at the MR to complete the receive (RX) beamforming. Consequently, a variety of beam training algorithms have been proposed to help reducing the time required for beam training [8].

Compared with traditional beamforming, the beam formed in mm-wave communications is narrower. Since the velocity of high-speed train (HST) is very fast, more frequent beam alignment is required. Traditional schemes, such as traversal scanning and hierarchical scanning, are no longer suitable for train-ground communication system in HSR scenarios. On the other hand, since HST generally travels along long straight rail track, line-of-sight (LoS) transmission dominates in train-ground communications between track-side BS and MR on the roof of the train. Therefore, the optimal beam direction, defined as the direction with the highest received signal power, is highly correlated with the location of the train, and it is possible to search for the optimal beam in a narrow area around the HST. At the same time, the HST travels periodically on the same transportation line based on a fixed timetable, and therefore the train-ground communication system intuitively generates some regular patterns. Since the artificial intelligence (AI) is good at extracting the hidden relationship from data, applying it to the train-ground communication system in the HSR scenarios can fully exploit the relationship between the position of the HST and the corresponding optimal beam direction, thereby simplifying the process of beam management.

Among the many branches of AI algorithms, reinforcement learning (RL) is a hot topic in recent years [9]. Compared with other algorithms, RL is model-free and does not require the desired output data from external supervisors. Furthermore, its complexity mainly exists in the offline training, and the most valuable action is executed directly according to the state of the system in the online training, thereby the complex iterative optimization process is avoided. Applying RL methods to the train-ground communications in the HSR scenarios can effectively reduce the overhead caused by frequent beam alignment, and realize more intelligent beam management. This motivates our work to design an intelligent beam management scheme based on the deep RL (DRL) algorithm, called deep Q-network (DQN), for the mm-wave train-ground communication system to improve the downlink SNR and its stability. Specifically, our contributions are as follows.

- Considering the mm-wave train-ground communication system in the HSR scenarios, we formulate the optimization problem of RX beam management to maximize the downlink SNR, while ensuring a certain SNR stability. Specific path loss, propagation properties, high mobility and Doppler spread inherent in the mm-wave based HSR train-ground communication are taken into consideration in the formulation.
- Exploiting the correlation between the position of HST and the downlink instantaneous received power, we propose an intelligent beam management scheme based on a DRL framework, called DQN, to establish the mapping between the position of HST and the optimal beam direction.
- We evaluate the proposed algorithm in the 30 GHz mm-wave train-ground communication system. Through extensive evaluations, we show that the proposed DQN-based algorithm not only achieves near-optimal performance, but also imposes the lowest online training complexity, in comparison with four baseline schemes. Our work therefore effectively resolves the classical dilemma having to tradeoff between system performance and beam training overhead in HSR scenarios.

The rest of the paper is organized as follows. In Section I-B, we provide an overview of the related work. Section II introduces the mm-wave based train-ground communication system model and formulates the problem of RX beam management aiming at maximizing the downlink received power. The proposed DQN-based algorithm to solve the RX beam management optimization problem is detailed in Section III, and the performance evaluation of the proposed DQN algorithm is provided in Section IV, in comparison with some existing schemes. Section V concludes this paper.

B. Related Work

There are two main research directions for beam management in HSR scenarios, beam switching and beam tracking. Compared with traditional traversal and hierarchical scanning, beam switching has lower complexity and overhead [10]. However, due to the limited number of optional beams and frequent beam switching, the gain of space division multiplexing decreases, i.e., the overhead is improved at the expense of system performance. On the other hand, beam tracking offers high flexibility and system capacity by adjusting the beam direction in real time to track the specific propagation path [3]. Compared with beam switching, beam tracking offers better system performance, while imposing higher complexity. Xiong *et al.* [11] proposed to cover and track low-speed vehicles with a wide beam. However, the system performance near the cell edge experiences significant degradation due to the high path loss of mm-wave. Gao *et al.* [3] proposed a dynamic beam tracking scheme that jointly adjusts the beam direction and width. In this scheme, the beam tracking in mm-wave train-ground communication is formulated as a non-convex optimization problem, and an approximate optimal solution is obtained based on genetic algorithm. However, the scheme contains many restricted assumptions and simplifications, which limit its application. Lu *et al.* [12] formulated the multi-cell coordinated beamforming in HSR scenarios as an optimization problem. By introducing auxiliary variables, the globally optimal beamforming vector

can be obtained iteratively. However, it is difficult for this high-complexity optimization based algorithm to work well in the train-ground communication systems with fast time-varying channels. Therefore, proposing a low-complexity and high-performance beam tracking scheme suitable for HSR scenarios remains a key challenge.

Recently, Yan *et al.* [13] provided a new idea for solving the above problem. In [13], it was mentioned that since the data transmission in the mm-wave train-ground communication is approximately LoS, the optimal beam direction is highly correlated with the position of the train. Yu *et al.* [14] further pointed out that the position of HST is almost predictable, and with the help of the position information of the train, the optimal beam direction can be determined by searching the narrow area around the train, rather than blindly exploring the whole space. In addition, based on a fixed timetable, HSTs run periodically on the same transportation line, and therefore the trackside train-ground wireless communication system intuitively generates some regular patterns. By capturing these patterns, we can obtain more valuable information to design better beam management in HSR scenarios.

It is well known that AI emerging in recent years has the ability to learn the hidden patterns from massive data, and it has been widely used in wireless communications [13], [15]. Feng and Mao [16] pointed out that solving the various combinatorial problems is a major challenge in wireless network design and the existing solutions usually rely on information exchange, thus having to trade off between overhead and system performance. The authors of [16] also mentioned that RL as a new AI paradigm is model-free, and it does not require knowledge of the interdependencies between different nodes. Therefore, it is possible to obtain better system performance with reduced overhead by applying RL. For example, Wang *et al.* [17] applied RL to solve the multi-channel access problem with the goal of maximizing the number of successful transmissions. Luo *et al.* [18] proposed a RL based dynamic power control scheme to maximize the sum rate of the static user equipment (UEs). Wang *et al.* [19] used RL to determine the optimal switching strategy in the Internet of things scenarios. In the scheme of [19], different motion patterns of the UEs are considered, and the goal is to reduce the switching frequency while meeting the system throughput requirement. Both [19] and [20] considered deploying multiple agents to control over complex communication systems.

At present, there exist a lot of research works on applying RL to wireless communications but most of the scenarios are urban cellular networks, where UEs move slowly and the range is small. A few works did investigate HSR scenarios. Wang *et al.* [21] considered multiple track-side BSs in the train-ground communication system, and implemented dynamic spectrum management by the interaction between multiple agents, thereby reducing the failure probability of handover. Cai *et al.* [22] realized the adaptive optimization of handover parameters in LTE-R system based on RL. Xu *et al.* [23] proposed a multi-agent-based power allocation algorithm in the HSR scenarios. In this scheme, the system state is defined as the channel state information (CSI) and beamforming vector, but the optimal beam direction is assumed to have been determined in advance. The paper [23] also emphasized that how to design an efficient and reliable beamforming scheme in the mm-wave train-ground communication system is an unsolved problem. It can be seen that the few existing schemes on the application of RL in train-ground communication systems make decisions based on CSI. But these works can help us to understand the idea of introducing RL in HSR communication networks, and are instructive for designing RL based intelligent beam management schemes in HSR scenarios.

To sum up, the existing studies have not considered the application of combining beamforming and DRL to the mm-wave based train-ground communication system in HSR scenarios. Using DRL to improve the beam management of mm-wave train-ground communication system has great potential, and this motivates our current work. In this paper, we formulate the problem of RX beamforming in the mm-wave train-ground communication system in HSR scenarios, and propose a dynamic beam management algorithm based on DRL to improve the downlink SNR.

II. System Overview and Problem Formulation

The train-ground communication in HSR scenarios is illustrated in Fig. 1, where both BS and MR are equipped with antenna arrays for beamforming. When the HST moves, the beamforming at transmitter aligns the main lobe with the MR. The beamforming at receiver estimates the arrival angle of the received signal in advance, and then form a RX beam according to the propagation direction of the signal.

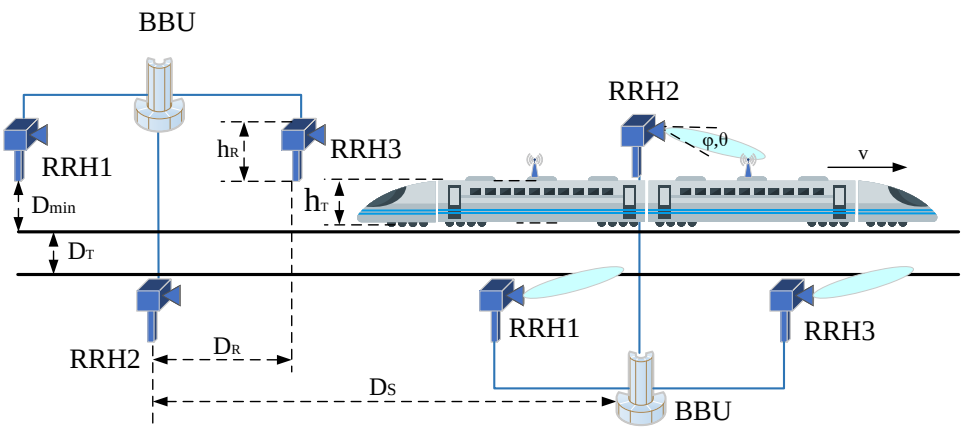


Fig. 2. Illustration of mm-wave train-ground communication system.

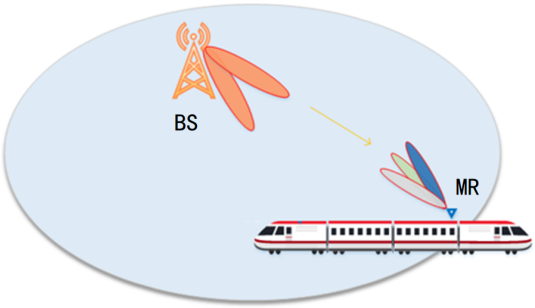


Fig. 1. Illustration of train-ground communication system in HSR scenarios.

RX beamforming can further improve the quality of received signals, thereby improving the throughput and reliability of train-ground communication system. When the MR receives the signal, the information is transmitted to the AP inside the carriage through optical fiber.

A. System Model

The infrastructure of the mm-wave train-ground communication system considered is depicted in Fig. 2, which is also the only mm-wave communication deployment scheme under the HSR scenario considered by 3GPP [24], [25]. Specifically, three remote radio taps (RRHs), denoted by RRH1, RRH2 and RRH3 in Fig. 2 with adjacent RRHs deployed on different sides of the rail track, and one baseband processing unit (BBU) are connected via optical fibers to form a track-side unit. Track-side units are deployed regularly along the entire rail track line as illustrated in Fig. 2, where D_R represents the minimum distance between the RRHs, and $D_S = 3D_R$ is the distance between the BBUs in adjacent two units. Furthermore, in Fig. 2, $D_{\min}$ represents the distance between RRH and the rail, and D_T is the minimum distance between two parallel rail tracks, while h_R is the height of the RRH relative to the rail, and h_T is the height of the train, which is also the height of the MR relative to the rail. In addition, the velocity of train is v, and φ and ϑ denote the azimuth and downtilt angles of transmit (TX) beam, respectively.

Since HSR is mainly deployed in suburbs and viaducts, the signal is rarely blocked by buildings to result in non-LoS transmission, and therefore only the LoS transmission is considered. In this paper, we choose RMa LoS in [26] as the model of path loss:

$$\overline{PL} = \begin{cases} \overline{PL}_1, & 10\,\mathrm{m} \le d_{2\mathrm{D}} \le d_{\mathrm{BP}}, \\ \overline{PL}_2, & d_{\mathrm{BP}} \le d_{2\mathrm{D}} \le 10\,\mathrm{km}, \end{cases} \tag{1}$$

$$\begin{aligned} \overline{PL}_1 =& 20\log_{10}\left(40\pi d_{3\mathrm{D}}\frac{f_{G_c}}{3}\right) - \min\left\{0.044h^{1.72}, 14.77\right\} \\ &+ \min\left\{0.03h^{1.72}, 10\right\}\log_{10}\left(d_{3\mathrm{D}}\right) \\ &+ 0.002\log_{10}(h)d_{3\mathrm{D}}, \end{aligned} \tag{2}$$

$$\overline{PL}_2 = \overline{PL}_1\left(d_{\mathrm{BP}}\right) + 40\log_{10}\left(\frac{d_{3\mathrm{D}}}{d_{\mathrm{BP}}}\right). \tag{3}$$

The carrier frequency f_{G_c} in (2) is measured in GHz. The definitions of $d_{2\mathrm{D}}$ and $d_{3\mathrm{D}}$ are given in Fig. 3, and $h \in [5,\ 50]$ m is the average height of buildings in the propagation environment, while d_{BP} represents the break point distance, which is calculated according to

$$d_{\mathrm{BP}} = 2\pi h_{\mathrm{R}} h_{\mathrm{T}} \frac{f_c}{\mathrm{c}}, \tag{4}$$

where f_c is the carrier frequency in Hz, and $\mathrm{c} = 3.0 \times 10^8$ m/s is the free-space propagation speed of electromagnetic wave.

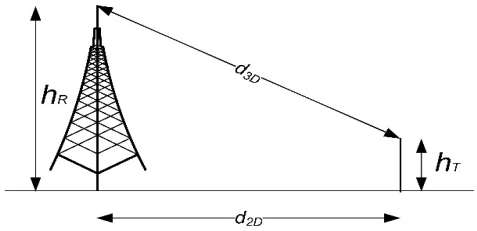


Fig. 3. Definitions of $d_{2\mathrm{D}}$ and $d_{3\mathrm{D}}$.

Next, we set a random shadow X_σ, which is a Gaussian random variable with zero mean and standard deviation σ. Then, the large-scale fading model is given by

$$PL(d) = \overline{PL}(d) + X_\sigma \ [\mathrm{dB}]. \tag{5}$$

In each time slot, the HST can be considered to be stationary. In order to simplify the analysis, we introduce the concept of location bin, i.e., when the projection of the MR relative to the rail is located in a certain location bin $\left[d_0 - \sigma_D,\ d_0 + \sigma_D\right]$, the MR is considered in the same location d_0. If the length of the railway to be observed is d, then the number of location bins on this railway is $L = d/2\sigma_D$. Hence, the algorithm used for beam management has L observation points. Then, the received power sequence can be expressed as

$$\boldsymbol{P}_R = \left[P_R^{(1)}\ P_R^{(2)} \cdots P_R^{(L)}\right]^{\mathrm{T}} \in \mathbb{R}^{L\times 1}. \tag{6}$$

In addition, we denote the TX antenna gain of the BS by G_T and the RX antenna gain of the MR by G_R, which depend on their corresponding beam angles and half-power beamwidths, respectively. Let θ_i and w_i be the Rx beam angle and half-power beamwidth in the i-th location bin. Then according to [26], the received power of the MR can be calculated by

$$P_R^{(i)} = \|\boldsymbol{H}(i)\|_{\mathrm{F}}^2 G_T(i) G_R(\theta_i, w_i) d_{\mathrm{TR}}(i)^{-\gamma} P_T, \tag{7}$$

where P_T is the TX power of the BS, $\boldsymbol{H}(i) \in \mathbb{C}^{N_r \times N_t}$ is the small-scale fading matrix, which depends on the wavelength, velocity of HST and Doppler spread, and $\|\cdot\|_{\mathrm{F}}$ denotes the matrix Frobenius norm, while $d_{\mathrm{TR}}(i)$ is the transceiver distance and γ is the large-scale path loss coefficient. Since the BS TX beam always points to the MR during HST operation and we further assume that the TX beam has a constant beamwidth, G_T is a function of the HST position i.

Due to the fact that the wireless channel in the train-ground communication system is fast time-varying, serious Doppler spread is introduced, which destroys the orthogonality of the sub-carriers in the OFDM signal, resulting in inter-carrier interference (ICI). More specifically, the small-scale fading matrix $\boldsymbol{H}(i)$ can be expressed by

$$H(i) = g_l(i)\, e^{2\pi j t f_{d,\max}\cos(\vartheta_l(i))} F\left(\theta_l^{TX}(i), \theta_l^{RX}(i)\right), \tag{8}$$

where $g_l(i)$ is the complex small-scale fading gain corresponding to the path l when the MR is at the i-th location bin, and $f_{d,\max} = \frac{f_c v}{c}$ is the maximum Doppler spread with f_c and c being the carrier frequency and the light speed, respectively, while $\boldsymbol{F}(\cdot,\cdot) \in \mathbb{C}^{N_r \times N_t}$ represents the TX and RX beamforming matrix [27], and $\vartheta_l(i)$ is the arrival angle of the path l relative to the direction of train movement. Usually, $g_l(i)$ is predefined and its second moment is set to 1 [28].

Consider that the BS and MR both use uniform linear antenna arrays to transmit and receive signals, with N_t and N_r representing their numbers of antenna elements, respectively, where $N_r < N_t$. Further denote Δ_t and Δ_r as the normalized TX and RX antenna spacings, respectively. Then $\boldsymbol{F}\left(\theta_l^{TX}(i), \theta_l^{RX}(i)\right)$ can be expressed as

$$\boldsymbol{F}\left(\theta_l^{TX}(i), \theta_l^{RX}(i)\right) = \boldsymbol{u}^{RX}\left(\theta_l^{RX}(i)\right)\left(\boldsymbol{u}^{TX}\left(\theta_l^{TX}(i)\right)\right)^{\mathrm{H}}, \quad (9)$$

where $(\cdot)^{\mathrm{H}}$ denotes the conjugate transpose operator, $\boldsymbol{u}^{TX}\left(\theta_l^{TX}(i)\right) \in \mathbb{C}^{N_T \times 1}$ is the spatial feature vector of the transmitting unit along $\Omega^{TX} = \cos\left(\theta_l^{TX}(i)\right)$ given by

$$\boldsymbol{u}^{TX}\left(\theta_l^{TX}(i)\right) = \frac{1}{\sqrt{N_t}} \begin{bmatrix} 1 \\ e^{-\mathrm{j}2\pi\Delta_t\Omega^{TX}} \\ \vdots \\ e^{-\mathrm{j}2\pi(N_T-1)\Delta_t\Omega^{TX}} \end{bmatrix}, \quad (10)$$

and $\boldsymbol{u}^{RX}\left(\theta_l^{RX}(i)\right) \in \mathbb{C}^{N_R \times 1}$ is the spatial feature vector of the receiving unit along $\Omega^{RX} = \cos\left(\theta_l^{RX}(i)\right)$ given by

$$\boldsymbol{u}^{RX}\left(\theta_l^{RX}(i)\right) = \frac{1}{\sqrt{N_r}} \begin{bmatrix} 1 \\ e^{-\mathrm{j}2\pi\Delta_r\Omega^{RX}} \\ \vdots \\ e^{-\mathrm{j}2\pi(N_R-1)\Delta_r\Omega^{RX}} \end{bmatrix}. \quad (11)$$

We use the widely adopted directional antenna model from IEEE 802.15.3c [29], which includes a linearly scaled Gaussian main lobe and a constant-level side lobe. Based on this model, the gain of a directional antenna can be expressed as

$$G(\theta) = \begin{cases} G_0 - 3.01\left(\frac{2\theta}{\theta_{-3\mathrm{dB}}}\right)^2, & 0^\circ \le \theta \le \frac{\theta_{\mathrm{ml}}}{2}, \\ G_{\mathrm{sl}}, & \frac{\theta_{\mathrm{ml}}}{2} \le \theta \le 180^\circ, \end{cases} \quad (12)$$

where $G_0 = 10\log_{10}\left(1.6162 \Big/ \sin\left(\frac{\theta_{-3\mathrm{dB}}}{2}\right)\right)^2$ is the maximum antenna gain, $\theta \in [0,\ \pi]$ denotes the beam angle, and $\theta_{-3\mathrm{dB}}$ represents the half-power beam angle, while $\theta_{\mathrm{ml}} = 2.6\theta_{-3\mathrm{dB}}$ denotes the main lobe width in degrees, and $G_{\mathrm{sl}} = -0.4111\ln\left(\theta_{-3\mathrm{dB}}\right) - 10.579$ is the side lobe gain.

We adopt the widely used ICI approximation model given in [30] to quantify the ICI power P_{ICI}:

$$P_{\mathrm{ICI}} = \int_{-1}^{1} (1 - |\tau|)\mathrm{J}_0\left(2\pi f_{d,\max} T_s \tau\right) d\tau, \quad (13)$$

where T_s denotes the symbol duration and $\mathrm{J}_0(\cdot)$ denotes the zero-order Bessel function of the first kind.

B. Problem Formulation

Based on the above model, when the MR is located in the i-th location bin, the received signal-to-noise ratio (SNR) of receiver can be expressed as

$$\mathrm{SNR}(i) = \frac{P_R^{(i)}}{N_0 W + P_{\mathrm{ICI}}}, \quad (14)$$

where the received power $P_R^{(i)}$ of the MR in the i-th location bin is given in (7), N_0 is the one-sided power spectral density of the white Gaussian noise, and W represents the system bandwidth. Then the downlink instantaneous maximum achievable rate of the established mm-wave train-ground communication system can be calculated by

$$R(i) = W\log_2(1 + \mathrm{SNR}(i)). \quad (15)$$

Our goal is to maximize the sum $\sum_{i=0}^{L} \mathrm{SNR}(i)$ of the downlink SNR sequence for the given TX power P_T of the BS, where $\mathrm{SNR}(i)$ is defined in (14), while ensuring the basic QoS requirements of passengers as well as reducing the complexity of beam tracking. Therefore, when the length d of the observed railway and the length $2\sigma_D$ of the location bin are determined, this optimization problem can be formulated as follows

$$\begin{aligned}
(\mathrm{P1}) \quad & \max_{\{\theta_i : 1 \le i \le L\}} \ \sum_{i=0}^{L} \mathrm{SNR}(i), && (16) \\
& \text{s.t.} \ \ \theta_i \in \Theta,\ \forall i = 1, 2, \cdots, L, && (17) \\
& \quad\ \ w_i \in \Omega,\ \forall i = 1, 2, \cdots, L, && (18) \\
& \quad\ \ A_E^{\mathrm{TX}} \le A_{\max}^{\mathrm{TX}}, && (19) \\
& \quad\ \ A_E^{\mathrm{RX}} \le A_{\max}^{\mathrm{RX}}, && (20) \\
& \quad\ \ P_R^{(i)} \ge P_{\mathrm{target}},\ \forall i = 1, 2, \cdots, L. && (21)
\end{aligned}$$

The constraints (17) and (18) are the beam direction and beamwidth constraints for avoiding large scale adjustments, where $\Omega = \left[-\pi/2, \pi/2\right]$ and $\Theta = \left[-\pi/2, \pi/2\right]$, and (19) and (20) represents the constraints on the maximum beam gain of an antenna element, in which A_E^{TX} and A_E^{RX} denote the TX and RX antenna element gains, respective, while $A_{\max}^{\mathrm{TX}}$ and $A_{\max}^{\mathrm{RX}}$ are the upper bounds of A_E^{TX} and A_E^{RX}, respectively. Furthermore, P_{target} represents the threshold of downlink received power.

In addition to high throughput, a consistent throughput is also a measure of quality of experience (QoE). To measure the stability of the system capacity, we define the RX SNR stability of a beam management scheme by

$$S(l) = \frac{|\mathrm{SNR}(l) - E[\mathrm{SNR}]|}{E[\mathrm{SNR}]}, \quad (22)$$

where $SNR(l)$ is the RX SNR at location l, and $E[\mathrm{SNR}]$ denotes the mean of SNR when HST passes through the whole track section. The closer $S(l)$ is to 0, the more stable the SNR and hence the system capacity is.

From the antenna gain model (12), it can be seen that the RX beam direction is the primary variable that determines the

downlink SNR. Therefore, we fix the beamwidth variable, and consider the following optimization problem:

$$\text{(P2)} \quad \max_{\{\theta_i:1\le i\le L\}} \quad \sum_{i=0}^{L} \mathrm{SNR}(i), \tag{23}$$

$$\text{s.t.} \quad \theta_i \in \Theta,\ \forall i = 1, 2, \cdots, L, \tag{24}$$

$$w_i = w,\ \forall i = 1, 2, \cdots, L, \tag{25}$$

$$A_E^{\mathrm{TX}} \le A_{\max}^{\mathrm{TX}}, \tag{26}$$

$$A_E^{\mathrm{RX}} \le A_{\max}^{\mathrm{RX}}, \tag{27}$$

$$P_R^{(i)} \ge P_{\mathrm{target}},\ \forall i = 1, 2, \cdots, L. \tag{28}$$

$$\sum_{i=0}^{L} \mathrm{S}(i) \le L \cdot \mathrm{S}_{\max} \tag{29}$$

Since both the objective function and the constrains (28) and (29) contain non-convex mathematical relations, (P2) is a non-convex nonlinear programming problem. An appropriate value of the beamwidth w can be determined by experiment.

when researchers solving non-convex problems in the past, either the objective function or constraints were approximated to transform the problem to be solved into a convex optimization problem, or greedy algorithms were used to make the solution as close as possible to the optimal solution. The essence of DRL is to search for the optimal solution in all system states and action spaces, which is an exploration mechanism between exhaustive and greedy methods. Using DRL to solve non-convex optimization problems doesn't require any approximation or transformation of the objective function or constraints. And the training of RL does not require a large number of samples and environmental information, which is an advantage that deep learning does not have. However, the training process of RL involves interacting with the environment and learning through policy updates and exploration, which typically demands more time and computational resources. Feedback in RL is often sparse and delayed, making training more challenging. Therefore, the integration of deep learning and RL, forming DRL, is currently a hot research topic. This approach leverages the strengths of deep learning models in perception and decision-making tasks.

III. RX Beamforming Optimization via DRL

As aforementioned, the route of the HST is fixed and its operation has certain periodicity. Therefore, the track-side train-ground wireless communication system exhibits some regular patterns, and AI is good at capturing these hidden patterns. In particular, the correlation between the position of the MR and the optimal beam direction as well as between the optimal beam direction and the maximum received signal power is such a pattern of the train-ground communication system in space. DRL is particularly suitable for our application, and we design the RX beam management scheme in the HSR scenarios based on DRL. Our proposed solution for solving the optimization (P2) consists of two parts: a Q-learning algorithm for online adjustment of beam direction and a convolutional neural network, which is a variant of deep neural networks (DNN), used to estimate the Q-values offline.

A. RX Beamforming Based on Reinforcement Learning

RL is an intelligent decision-making paradigm, which can be implemented with TensorFlow and Keras [31]. The idea is to use the experience obtained from training to learn the pattern of a specific Markov process by interacting with the environment through an agent. In our mm-wave train-ground communication system, a RL agent is deployed at the MR, whose task is to perform the training process and output the optimal policy for adjusting the beam direction with the goal of maximizing $\|\boldsymbol{P}_R\|_2$. The agent interacts with the mm-wave train-ground communication system for a long time, and performs the action by observing the current system state. First we introduce the following three definitions.

1) System State Space: Since the path loss is dominant in the fading experienced by the signal when the HST moves, the change of the received power is large when the traveling distance is long. Hence, there is a certain correlation between the downlink received power and the train position. Therefore, the instantaneous received power of the MR can be marked as the system state, which can be expressed as $S = (s_1, s_2, \cdots, s_P)$, where $P = N_B L$ represents the size of the system state space with N_B denoting the size of the beam codebook.

2) Action Space: When the beam width is determined, the beam direction of the MR is the optimization variable, and the action space is defined as $A = (a_1, a_2, \cdots, a_{N_B})$, i.e., the absolute adjustment of the beam index is performed in the MR side between different location bins.

3) Reward Function: Let a reference scheme be specified, in which the direction of RX beamforming is constant θ_F. Then, the reward function is defined as

$$r\big(s^k, a^k\big) = P_R\big(s^k, a^k\big) - P_R\big(s^k, \theta_F\big), \tag{30}$$

where $P_R\big(s^k, a^k\big)$ denotes the received power of the MR after the selected action a^k is executed in the state s^k of the k-th epoch, and $P_R\big(s^k, \theta_F\big)$ is the received power of the MR when the reference scheme is adopted in the same state.

Due to the mobility of HST, the state of the train-ground communication system changes with the time and the displacement of the MR. At the k-th epoch, the agent obtains the state s^k of the system, and derives an optimal policy π_k through the learning. Then, the system executes the action a^k of adjusting the beam direction based on the policy π_k, and obtains the reward $r\big(s^k, a^k\big)$ corresponding to the action a^k executed in the state s^k according to (30). Finally, the system advances to a new state s^{k+1} and repeats the above operations until the last epoch of training is completed. In addition, the reward returned by the RL is the cumulative discount reward R_k, which is used to measure the pros and cons of executing the action a^k in the state s^k, and its calculation is given by

$$R_k = \sum_{t=0}^{T} \alpha r\big(s^{k+t}, a^{k+t}\big), \tag{31}$$

where T is the maximum number of epochs, and $\alpha \in (0,\ 1]$ is the discount factor. Since the goal of the agent is to maximize the expectation of cumulative rewards, i.e., $\max E\big[R_k | s^k\big]$, the RL based algorithm does not only focus on the immediate benefits obtained after executing action in the current state. Rather

it hopes to maximize the long-term cumulative discounted benefits. In our mm-wave based train-ground communication system, this is equivalent to maximize the expectation of downlink SNR.

Two widely used approaches to obtain $\max E\left[R_k|s^k\right]$ are value function based and policy-based methods. In this paper, we use the value function based algorithm. For such algorithms, Banach fixed point theorem can be used to prove the uniqueness and existence of the convergence solution [32].

In order to determine the value function of state-action pairs, we first define the state value function by $V_\pi\left(s^k\right) = E\left[R_k|s^k\right]$, which is the cumulative reward of the policy π_k. The policy π_k can be defined as the probability distribution of the action a under the given state s^k, i.e., $\pi_k\left(a|s^k\right) = P\left(a|s^k\right)$ or $a = \pi_k\left(s^k\right)$. In our mm-wave based train-ground communication system, the changes of state are independent, and we can rewrite the state value function as follows:

$$V_\pi\left(s^k\right) = r\left(s^k, \pi_k\right) + \alpha \sum_{s^{k'} \in S} P\left(s^{k'} \middle| s^k, \pi_k\right) V_\pi\left(s^{k'}\right), \quad (32)$$

where $P\left(s^{k'}\middle|s^k, \pi_k\right)$ is the state transition probability when the strategy π_k is adopted. Next, we define the Q-function that represents the value of state-action pairs to quantify the expected reward after executing the action a^k when the system is in the state s^k and the strategy π is adopted, namely, $Q_\pi\left(s^k, a^k\right) = E\left[R_k|s^k, a^k\right]$. Since the system state is independent of the adjustment of the beam direction, $E\left[R_k|s^k, a^k\right] = E\left[R_k|s^k\right]$ and we have $Q_\pi\left(s^k, a^k\right) = V_\pi\left(s^k\right)$, which is defined in (32).

The agent's goal becomes to maximize this Q-function to obtain the optimal beam management strategy $\pi^\star$, that is,

$$\pi^\star = \arg\max_{a^k \in A} Q_\pi\left(s^k, a^k\right). \quad (33)$$

Afterward, the agent observes the current state s_k and executes the most valuable action $a^\star$ according to the optimal policy $\pi^\star$, which can be expressed as $a^\star = \pi^\star\left(s^k\right)$. In Q-learning, the Q-function can be determined recursively according to

$$\begin{aligned} Q_\pi\left(s^k, a^k\right) =& (1-\beta) Q_\pi\left(s^k, a^k\right) + \beta\Big(r\left(s^k, a^k\right) \\ &+ \alpha \max_{a'^k \in A} Q_\pi\left(s'^k, a'^k\right)\Big), \end{aligned} \quad (34)$$

where β represents the learning rate. According to (34), each state-action pair needs to be visited and evaluated when updating $Q_\pi\left(s^k, a^k\right)$, which leads to huge complexity and slow convergence. In order to solve (P2) effectively, we further introduce a DNN to achieve more intelligent control.

B. RX Beamforming Based on DQN

In order to speed up the convergence of Q-learning, we introduce an improved RL algorithm, which is a DQN. This algorithm uses the sampled data to train a DNN for estimating Q-values, which maps the inputs of state-action pairs onto their corresponding Q-values. However, directly applying DNN to Q-learning may lead to non-convergence due to the correlation between training samples and the correlation between Q-values and target values. Therefore, the DQN algorithm is adopted to reduce these correlations [31]. In this method, the agent first explores the environment by randomly executing actions and stores the experiences in the target network, where the set of experiences includes the current state, the executed action, the immediate reward, and the next state. Then, we can use a mechanism called experience replay, by randomly sampling the data in mini-batch from the target network to break the correlations in the observation sequence. Using the samples from the target network, the weight parameters of the DQN are updated by minimizing the mean squared error of the DQN and the Q-function of the target network. Usually the stochastic gradient descent method is used for this optimization.

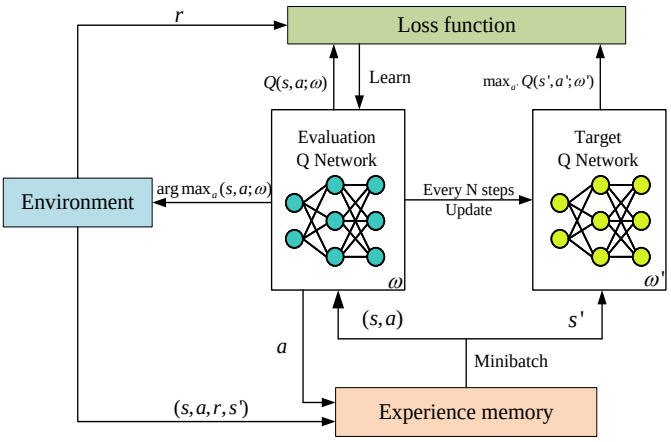


Fig. 4. Flowchart of the DRL approach.

The flowchart of the DRL method is shown in Fig. 4, and the parameter updating process of the DQN is detailed as follows. First, we replace or approximate the value function $Q_\pi\left(s^k, a^k\right)$ with a DQN Q that has the parameters ω^k, i.e., $Q\left(s^k, a^k; \omega^k\right) \approx Q_\pi\left(s^k, a^k\right)$. By setting $\beta = 1$, this approximation is used to define a loss function as:

$$\begin{aligned} L\left(\omega^k\right) =& E\Bigg[\Bigg(\underbrace{r\left(s^k, a^k\right) + \alpha \max_{a'^k \in A} Q_\pi\left(s'^k, a'^k\right)}_{\text{Target}} \\ &- \underbrace{Q\left(s^k, a^k; \omega^k\right)}_{\text{predicted}}\Bigg)^2\Bigg]. \end{aligned} \quad (35)$$

The gradient of $L\left(\omega^k\right)$ with respect to ω^k is given by

$$\begin{aligned} \frac{\partial L\left(\omega^k\right)}{\partial \omega^k} =& -E\Bigg[\Big(r\left(s^k, a^k\right) + \alpha \max_{a'^k \in A} Q_\pi\left(s'^k, a'^k\right) \\ &- Q\left(s^k, a^k; \omega^k\right)\Big) \frac{\partial Q\left(s^k, a^k; \omega^k\right)}{\partial \omega^k}\Bigg]. \end{aligned} \quad (36)$$

The iterative process with gradient descent is repeated until the last epoch is competed. We can update ω^k according to the experience by random sampling from the experience pool to obtain the optimal strategy of beam management on the MR side. After the training of the DQN is completed, the agent directly executes the most valuable action of beam direction adjustment according to the estimated Q-values.

As shown in the Fig. 5, we adopt a fully connected neural network framework in this paper, where the input layer is composed of the system states and output layer is the policy $\pi(s, a)$ for beam direction adjustment, containing a hidden

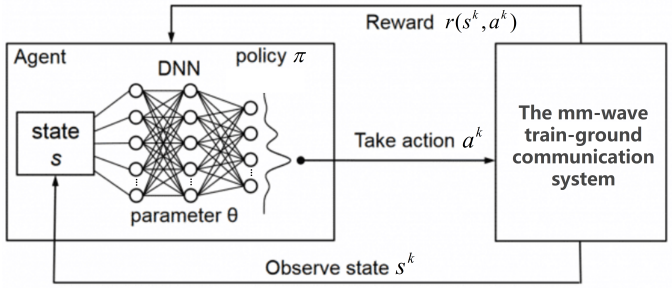


Fig. 5. Framework of the proposed DRL based algorithm.

layer with 10 neurons. Then, we can train the neural network used to estimate the Q-values based on the loss function shown in the (35).

The pseudo code of the DQN algorithm for beam management is given in Algorithm 1, where T is the maximum number of epochs. In each epoch, the algorithm consists of two stages. The first stage is the transition generating stage, which uses the Q-learning to adjust the direction of the RX beam. At this stage, the agent observes the state of the system, randomly selects action with probability ε or the one most likely to obtain the largest discount accumulative reward with probability $1-\varepsilon$, executes the selected action, and gets the corresponding reward. Then the system reaches the next state, and stores this memory into $\mathcal{B}$ as experience. The second stage is the parameter updating stage. At this stage, the algorithm randomly samples minibatch of memories from $\mathcal{B}$, and uses the stochastic gradient descent method to update the parameters of the DNN to make the more accurate prediction of Q-values.

Algorithm 1 Dynamic Beam Management Based on DQN

Initialization:

Initialize RX beamforming: $\theta_l^{RX}(i)=0^\circ, l=1,2,\cdots,L$, $i=1,2,\cdots,T$;

Initialize action-value function Q_π, target action-value function Q, and replay memory buffer $\mathcal{B}$;

for current location bin $l=1:L$ **do**

  Initialize state sequence $s_l^1=0\,\mathrm{dBm}$;

  for $k=1:T$ **do**

    Observe current state $s_l^k=P_R^{(l)}$;

Transition Generating Stage:

    Select action $a_l^k\in A$ based on $\varepsilon-$greedy policy;

    Execute action a_l^k, receive reward $r\left(s_l^k,a_l^k\right)$ and next state s_l^{k+1}; Store $\left\{s_l^k,a_l^k,r\left(s_l^k,a_l^k\right),s_l^{k+1}\right\}$ into $\mathcal{B}$;

Parameter Updating Stage:

    Sample randomly minibatch of experience $\left\{s^m,a^m,r\left(s^m,a^m\right),s^{m+1}\right\}$ from $\mathcal{B}$;

    if k terminates at step $m+1$ **then**

      set $y^m=r\left(s^m,a^m\right)$;

    else

      set $y^m=r\left(s^m,a^m\right)+\alpha\max_{a'\in A}Q_\pi\left(s^{m+1},a'\right)$;

    end if

    Perform gradient descent step to minimize $\left(y^m-Q\left(s^m,a^m;\omega^m\right)\right)^2$;

    Every C steps reset $Q=Q_\pi$;

  end for

end for

Output:

RX power sequence $\left\{P_R^{(l)}\right\}_{l=1}^L$; Trained DNN.

C. Implementing Proposed DQN-Based Algorithm

The implementation of the proposed DQN-based algorithm on HST includes offline training and online prediction.

When the HST is in trial operation, the proposed algorithm is in the offline training mode. After training, a DNN-based trained model is obtained, which maps the receiving power of MR and the position of the HST to the optimal beam direction, as illustrated in Fig. 4. The training complexity can be derived as follows. Assume that given the RX beam angle, the complexity of calculating the downlink received power once is C_P. Then the complexity of the first stage of Algorithm 1 is on the order of $\mathsf{O}\left(L\cdot T\cdot C_P\right)$, where L is the number of rail location bins and T is the maximum number of epochs. Further assume that the size of the data set representing the state-action pairs is D, and the size of the single sampling is m_{E}. Then the complexity of the second stage of Algorithm 1 is $\mathsf{O}\left(L\cdot T\cdot\frac{D}{m_{\mathrm{E}}}\cdot C_P\right)$. Therefore, the training computational complexity of the proposed DQN-based algorithm is on the order of $\mathsf{O}\left(L\cdot T\left(1+\frac{D}{m_{\mathrm{E}}}\right)\cdot C_P\right)$.

After training, HST is in actual operation, and the proposed algorithm is in the online prediction mode. According to the DNN obtained in the offline training phase, the agent directly adjusts the beam direction that is most likely to obtain the optimal average SNR. Therefore, when HST is in operation, the proposed DQN-based algorithm does not traverse any independent variable space. Given the number of location bins L, the online or operational complexity of the proposed algorithm is on the order of $\mathsf{O}\left(L\cdot C_P\right)$.

We now compare the proposed DQN-based algorithm with the classic Q-learning algorithm, in terms of computation and storage requirements. Our DQN-based algorithm has one more stage of parameter updating during offline training, but the introduction of DNN solves the problem of excessive storage space required by the classic Q-learning to construct a RL model for large-scale state-action set. Therefore, the proposed algorithm is advantageous in the storage complexity of both the offline training and online prediction phases. The proposed DQN-based algorithm and the traditional Q-learning algorithm both have the same online computational complexity.

The training of the DRL model proposed in this paper utilizes communication parameters such as MR's instantaneous receive power, MR's beam direction, downlink receive power and SNR. The main differences between various vehicle-to-ground communication scenarios lie in the physical settings of BSs, railway tracks, high-speed trains, and the environment. Therefore, the physical parameters of the vehicle-ground scenarios have a weak effect on the training of the DRL algorithm in this paper, ensuring the applicability of the DRL algorithm across different scenarios. Additionally, the communication characteristics of different scenarios involve choosing between wideband and narrowband mm-wave communication. Wideband communication is suitable for scenarios requiring high data rates and large data transmission,

while narrowband communication prefers to scenarios that need to overcome signal fading, anti-multipath effects and satisfy the requirement of low power consumption [33]. In the narrowband communication systems, since frequency of all subcarriers are close to the carrier frequency, the spatial directions are frequency-independent, whereas wideband mm-wave communication experiences severe beam squint due to the spatial direction of the beam changing with frequency, resulting in a significant impact on system performance [34], [35]. Furthermore, the use of statistical mm-wave CSI for hybrid precoding effectively reduces reliance on hard-to-obtain instantaneous mm-wave CSI, which is a critical consideration in wideband mm-wave communication [36].

IV. Performance Evaluation

A. Simulation Setup

According to the train-ground communication system modeled in Subsection II-A, both the BS and MR work in the 30 GHz mm-wave frequency band. Considering the small coverage of mm-wave BS, the length of the railway to be observed is set to $d = 500\,\mathrm{m}$, and the BS is located at the leftmost of the railway in the scenario of Fig. 2. Other simulation parameters are given in Table I. In order to evaluate the performance of the proposed DQN-based algorithm, the following four algorithms are chosen as baseline schemes.

1) FB (fixed Beam): The azimuth of the RX beam is unchanged during the operation. In this scenario, there exists an azimuth angle $\theta^{\star}_{\mathrm{FB}}$ with which the MR achieves the highest average received power when passing through the railway. Assume that this $\theta^{\star}_{\mathrm{FB}}$ has been acquired. We fix the angle of the RX beam to $\theta^{\star}_{\mathrm{FB}}$.
2) RT (real-time tracking) [3]: The RX beam is always aligned with the BS during the HST passing through the railway to obtain the maximum beam directional gain. This is the idealized beam management scheme with the optimal performance, requiring continuous beam switching.

TABLE I
Simulation System Parameters

Parameters	Values
System bandwidth: W	2160 MHz
Background noise: N_0	-134 dBm/MHz
Carrier frequency: f_c	30 GHz
Cell radius: R	500 m
Velocity of HST: v	350 km/h
Number of paths: L	1
Distance between RRH and rail: $D_{\min}$	150 m
RRH height: h_R	15 m
MR height (top of train): h_T	5 m
Transmit power: P_T	31 dBm
Threshold of received power: P_{target}	-130 dBm
Radius of a location bin: σ_D	0.5 m
Number of BS antennas: N_t	32
Number of MR antennas: N_r	8
maximum directional gain (BS): $A^{\mathrm{TX}}_{\max}$	8 dBi
maximum directional gain (MR): $A^{\mathrm{RX}}_{\max}$	5 dBi
maximum number of epochs: T	5000
Learning rate: L_R	0.5
Probability of exploration: ε	0.8
Size of replay memory buffer: $m_{\mathcal{B}}$	32

3) BS-P (beam switching according to the position) [37]: Given the number of RX beam switching modes N_s, the length d of the railway is divided into N_s sections, and the beam switching is executed when the HST enters a section. The direction in which the midpoint of each section points to the BS is taken as the direction of the RX beam on this section of railway.
4) BS-A (beam switching according to the angle) [3] [38]: The range of the RX beam angle in the length d of the railway is divided into N_s intervals. When the azimuth of the line connecting MR and BS enters an angle interval, the median value of the interval serves as the direction of the RX beam in this interval.

According to the ICI model (13), for the HTS velocity of 350 km/h, the ICI power is approximately -97 dBm. We fix the beamwidth to $w = 10^{\circ}$ for our DQN-based scheme.

B. Online Computational Complexity Comparison

Among the four comparison schemes, only the FB scheme also needs offline training to determine the beam direction corresponding to the optimal average received power. Its offline training complexity is on the order of $\mathsf{O}\left(L^2 \cdot C_P + C_{DS}\right)$, where C_{DS} denotes the complexity of a direct sort, while again L is the number of rail location bins and C_P denotes the complexity of calculating the downlink SNR once given the RX beam angle.

What really matters is the online computational complexity during the operation of HST. Table II compares the online complexity of the proposed DQN-based algorithm with those of the four benchmark schemes. The FB scheme does not adjust the beam angle during the operation of HST, and it imposes no online complexity. The idealized RT algorithm needs to continuously obtain the optimal absolute beam direction during the operation of HST, which is impossible to realize. Therefore, we assume that it only tracks the N_r optimal absolute beam directions in each position bin, where N_r is the number of MR antennas, and hence its online computational complexity is $\mathsf{O}\left(N_r \cdot L \cdot C_P\right)$. Let the number of switching modes for the BS-P and BS-A schemes be N_s. Then the online computational complexity of these two beam switching schemes is on the order of $\mathsf{O}\left(N_s \cdot C_P\right)$.

C. Training of DQN-ased Beam Management Model

The training process of our DQN-based beam management model is shown in Fig. 6. In Fig. 6, each point represents a learning process, and the depth of the area color represents the

TABLE II
Online Computational Complexity Comparison

Name	Online computational complexity
Proposed-DQN	$\mathsf{O}\left(L \cdot C_P\right)$
FB	None
RT	$\mathsf{O}\left(N_r \cdot L \cdot C_P\right)$
BS-P	$\mathsf{O}\left(N_s \cdot C_P\right)$
BS-A	$\mathsf{O}\left(N_s \cdot C_P\right)$

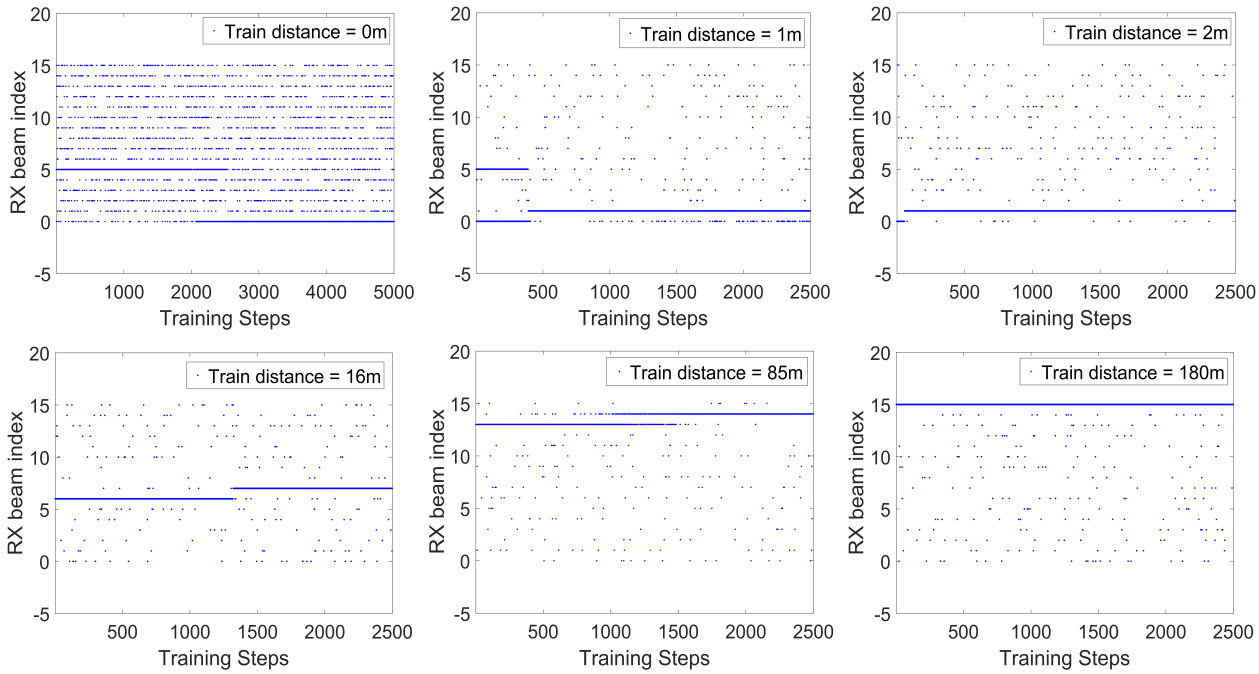


Fig. 6. Training process of the proposed DQN-based beam management model.

degree to execute actions in that area. When the HST is at the starting point of the railway, the DQN-based algorithm builds the model from the empty state, and finds the optimal RX beam index after about 2000 steps. When the HST advances to the second location bin, i.e., the travel distance is 1 m, the optimal RX beam index is found after about 470 steps. Since the DQN-based algorithm executes absolute beam adjustment, the model training between different location bins should be independent. But because the instantaneous received power of the MR is defined as the system state and it reflects the current position of the HST, the decision-making process of our algorithm in the previous location bin can become the subsequent experience. Also we approximate the downlink received power, i.e., similar values of received power are marked as the same system state, which deepens the correlation between the instantaneous received power and the position information of HST. Wen the HST reaches the third location bin, a relatively stable strategy is formed after about 50 steps, which benifts from the experience after learning the decision-making process of the first two episodes.

Theoretically, according to the mm-wave path loss model of Section II-A, its value generally has a logarithmic relationship with the travel distance of the HST. Therefore, as the HST moves away from the BS, the rate of increase in the path loss gradually becomes slower. Then the correlation between the downlink received power and the position information of HST is enhanced, and the experience of the decision-making process from the previous location bins becomes more relevant for the current bin. However, in a few positions, such as 16 m and 85 m, it takes about 1500 steps to form a stable strategy. This may be caused by more rapid changes of the instantaneous received power near these location bins, leading to higher changing rate of the path loss.

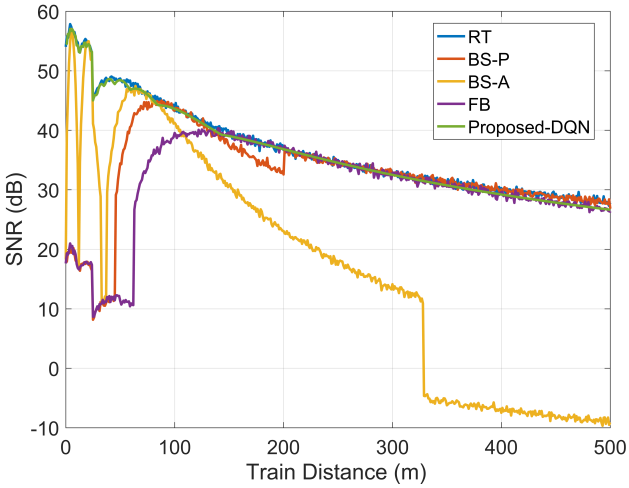


Fig. 7. RX SNR comparison of different beamforming strategies. $N_s = 3$ for BS-P and BS-A.

D. Simulation Results and Discussions

Fig. 7 compares the downlink SNR obtained by the proposed DQN-based algorithm with those of the four benchmark schemes, where the number of switching modes for the BS-P and BS-A schemes is set to $N_s = 3$. Since the path loss increases with the distance, the RX SNR decreases as the HST moves away from the BS. The idealized RT algorithm attains the best performance as expected, followed by our DQN-based algorithm, while the third best is the BS-P. It can be seen that the performance gap between the proposed DQN-based algorithm and the RT scheme is negligible, which indicates that our scheme achieves a near-optimal performance. When the MR is located within 50 m of the BS, the performance of the BS-P scheme is very poor, similar to that of the FB. As the distance increases further, the performance of the BS-P scheme recovers quickly, approaching that of the RT scheme but with a small dip in its SNR in the region of 150 m to 200 m. As

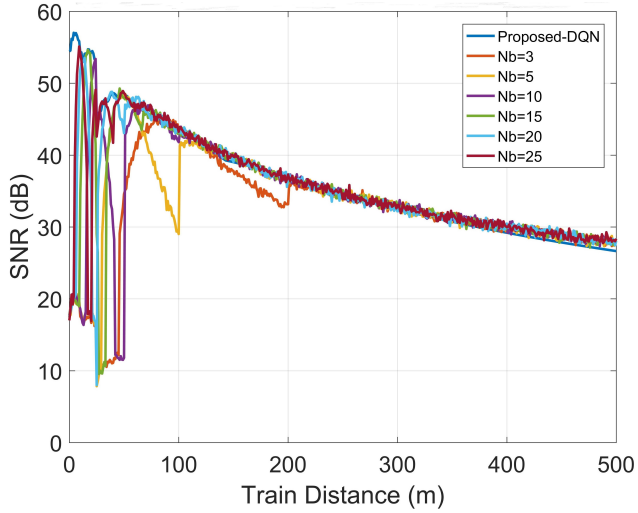


Fig. 8. RX SNR comparison of our proposed scheme and the BS-P with different numbers of switching modes N_s.

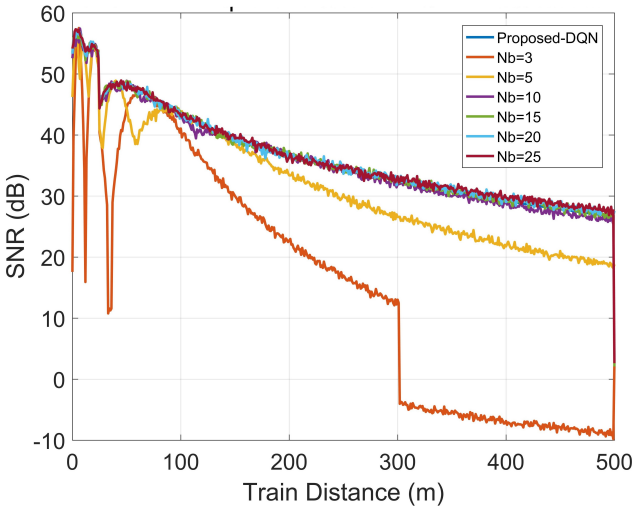


Fig. 10. RX SNR comparison of our proposed scheme and the BS-A with different numbers of switching modes N_s.

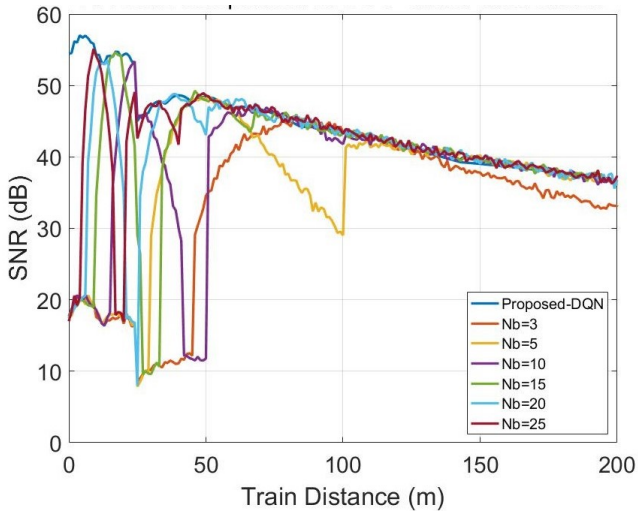


Fig. 9. RX SNR comparison (zoom on the first 200 m of the railway) of our proposed scheme and the BS-P with different numbers of switching modes N_s.

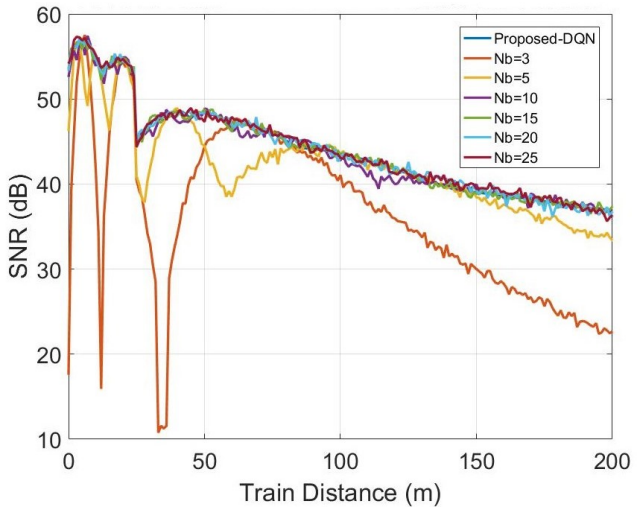


Fig. 11. RX SNR comparison (zoom on the first 200 m of the railway) of our proposed scheme and the BS-A with different numbers of switching modes N_s.

for the BS-A scheme, when the distance is less than 100 m, it executes multiple beam switching operations in a short period of time and its performance varies dramatically but is generally better than the BP-P and FB schemes. However, as the distance increases beyond 100 m, the performance of the BP-A deteriorates sharply. For the FB scheme, its performance is the worst when the distance is less than 100 m. However, for the region beyond 100 m, its performance improves quickly and approaches the optimal performance of the RT scheme.

The results of Fig. 7 confirm that our proposed-DQN scheme can closely approach the optimal performance of the idealized RT scheme. Furthermore, the online computational complexity of our DQN scheme is N_r times lower than that of the RT scheme. These two facts make our scheme particular suitable for the mm-wave train-ground communication system. Next we use our DQN scheme as the performance benchmark to further investigate the impact of N_s on the achievable performance of the BS-P and BS-A schemes. Note that the online computational complexity of our DQN scheme is $\frac{L}{N_s}$ times higher than that of the BS-P and BS-A schemes.

Fig. 8 depicts the RX SNR performance obtained by the BS-P scheme given different numbers of switching modes N_s, in comparison with the RX SNR achieved by our proposed scheme. Since the RX SNR curves of the BS-P with different N_s become very close when the MR is more than 200 m away from the BS, we zoom on the first 200 m of the railway and depict the corresponding results in Fig. 9, in order to clearly show the influence of N_s on the RX SNR performance of the BS-P. It can be seen from Fig. 9 that with $N_s \geq 15$, the SNR performance of the BS-P is approaching the performance of the proposed DQN-based algorithm for the distance over 50 m, but there still exists a significant performance gap within the first 50 m. Obviously, increasing N_s leads to better performance at the cost of higher online computational complexity. On the whole, setting $N_s = 15$ is appropriate for the BS-P scheme in this simulation system. With $N_s = 15$, the online complexity of the BS-P is still much lower than that of our DQN-based scheme.

Similarly, Fig. 10 investigates the impact of N_s on the achievable RX SNR of the BS-A scheme, while Fig. 11 zooms

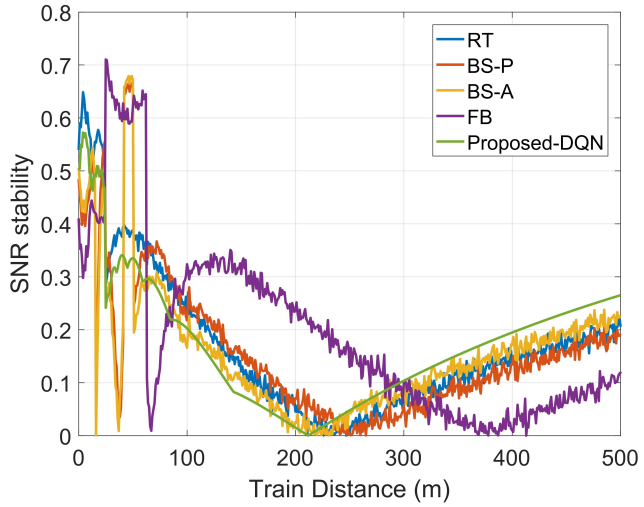


Fig. 12. SNR stability comparison of different beamforming strategies.

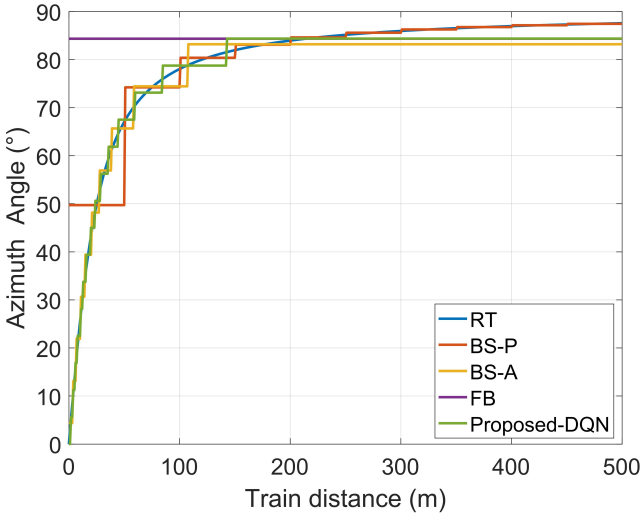


Fig. 13. Beam azimuth angle tracking comparison of different beamforming strategies.

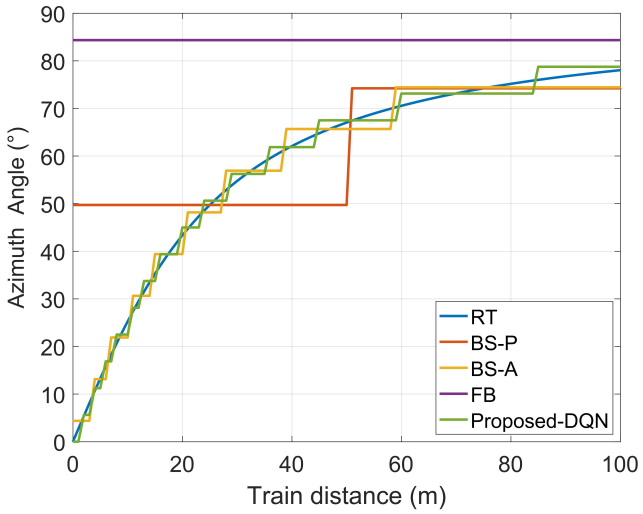


Fig. 14. Beam azimuth angle tracking comparison of different beamforming st

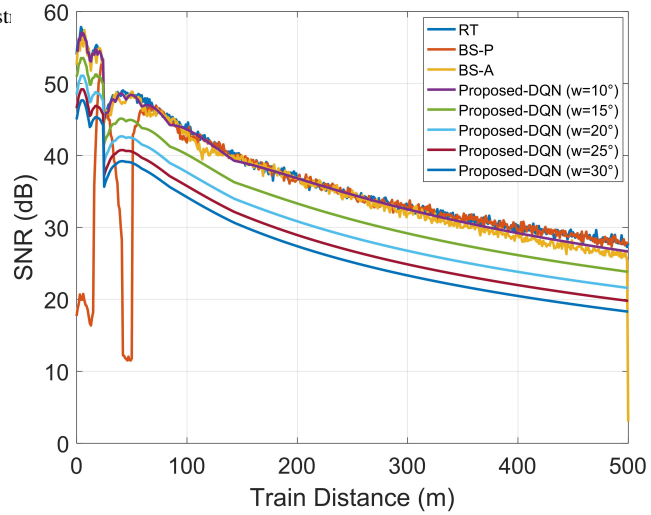


Fig. 15. The RX SNR performance of the proposed DQN-based scheme as the function of beamwidth w, with the idealized RT as the benchmark.

on the results for the first 200 m of the railway, where again we use our DQN scheme as the performance benchmark. It can be seen from Fig. 11 that with $N_s \geq 10$, the RX SNR performance of the BS-A scheme approach that of the proposed DQN-based algorithm. Hence, choosing $N_s = 10$ is cost-effective for the BS-A scheme. In the remaining simulation experiments, we set $N_s = 10$ for the BS-P and BS-A.

The SNR stability results of the five schemes are compared in Fig. 12. In general, the proposed DQN-based algorithm and the RT scheme have the best SNR stability, while the FB has the worst SNR stability. Hence compared with the other three practical schemes, our DQN-based algorithm offers better QoE service.

Fig. 13 depicts the trajectories of RX beam azimuth angle adjustment attained by the five algorithms over the whole section of the railway, while Fig. 14 compares the same trajectories near the BS. The trajectory of the idealized RT scheme represents the optimal beam azimuth angle tracking but it is achieved by infinitely many beam switching operations with infinitely high online computational complexity, which is impractical. Our DQN-based algorithm closely approximates the optimal trajectory of the idealized RT scheme, particularly at the position near the BS. The adjustment frequency required by the proposed algorithm is related to the length of the railway d and the radius of the location bin σ_D. The adjustment frequency of the two beam switching schemes depends on the number of beam modes N_s, which is 10. Obviously, the FB scheme does not adjust the RX beam azimuth angle. It can be seen that the proposed DQN-based algorithm is capable of attaining a near optimal performance while imposing an affordable online complexity.

Lastly, we investigate the impact of the beamwidth w on the achievable RX SNR performance of our proposed DQN-based scheme in Fig. 15. It can be seen from Fig. 15 that as w decreases, the achievable performance of our DQN-based scheme improves. In particular, when the beamwidth is reduced to $w = 10°$, the performance of our DQN-based scheme approaches the optimal performance of the idealized RT scheme. This confirms that the choice of $w = 10°$ is appropriate.

V. Conclusions

In this paper, we have proposed an intelligent beam management scheme based on the DQN to maximize the downlink

SNR of the mm-wave train-ground communication system. Our novel idea has been to deploy a reinforcement learning agent at the MR to perform the beam training by establishing the DQN based mapping between the HST position and the optimal beam direction. During the operation of HST, according to the DNN model obtained in training, the agent can directly predict the optimal beam direction from the HST position that is most likely to obtain the maximum downlink SNR. Extensive simulations have demonstrated that the proposed DQN-based scheme closely approaches the optimal system performance of the extremely-high-complexity idealized RT scheme, while ensuring a certain SNR stability and imposing a low online computational complexity. The results has also shown that our DQN-based scheme outperforms three low-complexity benchmark schemes. This paper therefore has developed the DQN-based beam management scheme which achieves near-optimal performance with affordable online computational complexity, suitable for mm-wave high speed train-ground communications. In the next phase of our research work, we will focus on wideband mm-wave communication, which is one of the hot topics in future communication. Therefore, we aim to optimize the current DRL algorithm presented in this paper by incorporating wideband channel model [39]. However, it is essential to address the challenges posed by beam squint and the difficulty in obtaining instantaneous full-dimensional CSI in wideband mm-wave communication. To mitigate the effects of beam squint, we will employ Hybrid Transmit Precoding (TPC) schemes, using either perfect CSI or low-complexity TPC schemes based on array vectors [34], [35]. Additionally, we will utilize statistical mm-wave CSI for hybrid precoding design to tackling the challenges in acquiring full-dimensional instantaneous CSI [36].

REFERENCES

[1] B. Ai, A. F. Molisch, M. Rupp, and Z.-D. Zhong, "5G key technologies for smart railways," *Proceedings of the IEEE*, vol. 108, no. 6, pp. 856–893, Jun. 2020.

[2] B. Lannoo, D. Colle, M. Pickavet, and P. Demeester, "Radio-over-fiber-based solution to provide broadband internet access to train passengers [topics in optical communications]," *IEEE Communications Magazine*, vol. 45, no. 2, pp. 56–62, Feb. 2007.

[3] M. Gao, B. Ai, Y. Niu, Z. Zhong, Y. Liu, G. Ma, Z. Zhang, and D. Li, "Dynamic mmwave beam tracking for high speed railway communications," in *2018 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, Barcelona, Spain, 2018, pp. 278–283.

[4] D. Fan, Z. Zhong, G. Wang, and F. Gao, "Doppler shift estimation for high-speed railway wireless communication systems with large-scale linear antennas," in *2015 International Workshop on High Mobility Wireless Communications (HMWC)*, Xi'an, China, Oct. 2015, pp. 96–100.

[5] M. Elkashlan, T. Q. Duong, and H.-H. Chen, "Millimeter-wave communications for 5G: fundamentals: Part I [Guest Editorial]," *IEEE Communications Magazine*, vol. 52, no. 9, pp. 52–54, Sep. 2014.

[6] C. Doan, S. Emami, D. Sobel, A. Niknejad, and R. Brodersen, "Design considerations for 60 GHz cmos radios," *IEEE Communications Magazine*, vol. 42, no. 12, pp. 132–140, Dec. 2004.

[7] S. Singh, R. Mudumbai, and U. Madhow, "Interference analysis for highly directional 60 GHz mesh networks: The case for rethinking medium access control," *IEEE/ACM Transactions on Networking*, vol. 19, no. 5, pp. 1513–1527, Oct. 2011.

[8] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave mimo systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[9] Z. Liu, J. Zhang, Z. Liu, H. Du, Z. Wang, D. Niyato, M. Guizani, and B. Ai, "Cell-free XL-MIMO meets multi-agent reinforcement learning: Architectures, challenges, and future directions," *IEEE Wireless Commun.*, to appear 2023.

[10] V. Va, X. Zhang, and R. W. Heath, "Beam switching for millimeter wave communication to support high speed trains," in *2015 IEEE 82nd Vehicular Technology Conference (VTC2015-Fall)*, Boston, MA, USA, Sep. 2015, pp. 1–5.

[11] K. Xiong, B. Wang, C. Jiang, and K. J. R. Liu, "A broad beamforming approach for high-mobility communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 11, pp. 10 546–10 550, Nov. 2017.

[12] Y. Lu, K. Xiong, P. Fan, and Z. Zhong, "Optimal multicell coordinated beamforming for downlink high-speed railway communications," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 10, pp. 9603–9608, Oct. 2017.

[13] L. Yan, X. Fang, X. Wang, and B. Ai, "Ai-enabled Sub-6-GHz and mm-wave hybrid communications: Considerations for use with future hsr wireless systems," *IEEE Vehicular Technology Magazine*, vol. 15, no. 3, pp. 59–67, Sep. 2020.

[14] D. Yu, G. Yue, N. Wei, L. Yang, H. Tan, D. Liang, and Y. Gong, "Empirical study on directional millimeter-wave propagation in railway communications between train and trackside," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 12, pp. 2931–2945, Dec. 2020.

[15] M. G. Kibria, K. Nguyen, G. P. Villardi, O. Zhao, K. Ishizu, and F. Kojima, "Big data analytics, machine learning, and artificial intelligence in next-generation wireless networks," *IEEE Access*, vol. 6, pp. 32 328–32 338, May 2018.

[16] M. Feng and S. Mao, "Dealing with limited backhaul capacity in millimeter-wave systems: A deep reinforcement learning approach," *IEEE Communications Magazine*, vol. 57, no. 3, pp. 50–55, Mar. 2019.

[17] S. Wang, H. Liu, P. H. Gomes, and B. Krishnamachari, "Deep reinforcement learning for dynamic multichannel access in wireless networks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 4, no. 2, pp. 257–265, Jun. 2018.

[18] C. Luo, J. Ji, Q. Wang, L. Yu, and P. Li, "Online power control for 5G wireless communications: A deep q-network approach," in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[19] Z. Wang, L. Li, Y. Xu, H. Tian, and S. Cui, "Handover optimization via asynchronous multi-user deep reinforcement learning," in *2018 IEEE International Conference on Communications (ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[20] O. Naparstek and K. Cohen, "Deep multi-user reinforcement learning for dynamic spectrum access in multichannel wireless networks," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Singapore, Dec. 2017, pp. 1–7.

[21] C. Wang, Q. Wu, Z. Tang, J. Sheng, C. Wu, and Y. Wang, "Spectrum management in high-speed railway cooperative cognitive radio network based on multi-agent reinforcement learning," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, Limassol, Cyprus, Jun. 2020, pp. 702–707.

[22] X. Cai, C. Wu, J. Sheng, J. Zhang, and Y. Wang, "A parameter optimization method for lte-r handover based on reinforcement learning," in *2020 International Wireless Communications and Mobile Computing (IWCMC)*, Limassol, Cyprus, Jun. 2020, pp. 1216–1221.

[23] J. Xu and B. Ai, "Artificial intelligence empowered power allocation for smart railway," *IEEE Communications Magazine*, vol. 59, no. 2, pp. 28–33, Feb. 2021.

[24] Document R1-165484, "WF on evaluation assumptions for high speed train scenario: Macro + relay at 30 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. RAN1#85, May 2016.

[25] D. He, B. Ai, K. Guan, Z. Zhong, B. Hui, J. Kim, H. Chung, and I. Kim, "Channel measurement, simulation, and analysis for high-speed railway communications in 5G millimeter-wave band," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 10, pp. 3144–3158, Oct. 2018.

[26] 3GPP, "Study on channel model for frequencies from 0.5 to 100 GHz. TR 38.901," 2017.

[27] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[28] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[29] IEEE 80215 WPAN Millimeter Wave Alternative PHY Task Group 3c (TG3c), "Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks

(wpans) (amendement 2: Millimeter-wave-based alternative physical layer extension)," Oct. 2009.
[30] Y. Li and L. Cimini, "Bounds on the interchannel interference of ofdm in time-varying impairments," *IEEE Transactions on Communications*, vol. 49, no. 3, pp. 401–404, Mar. 2001.
[31] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[32] H. Brezis and H. Brézis, *Functional analysis, Sobolev spaces and partial differential equations*. Springer, 2011, vol. 2, no. 3.
[33] B. Wang, F. Gao, S. Jin, H. Lin, G. Y. Li, S. Sun, and T. S. Rappaport, "Spatial-wideband effect in massive mimo with application in mmwave systems," *IEEE Communications Magazine*, vol. 56, no. 12, pp. 134–141, Dec. 2018.
[34] Y. Chen, Y. Xiong, D. Chen, T. Jiang, S. X. Ng, and L. Hanzo, "Hybrid precoding for wideband millimeter wave mimo systems in the face of beam squint," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 1847–1860, Mar. 2021.
[35] Y. Chen, D. Chen, and T. Jiang, "Beam-squint mitigating in reconfigurable intelligent surface aided wideband mmwave communications," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*, Nanjing, China, 2021, pp. 1–6.
[36] Y. Chen, D. Chen, T. Jiang, and L. Hanzo, "Channel-covariance and angle-of-departure aided hybrid precoding for wideband multiuser millimeter wave mimo systems," *IEEE Transactions on Communications*, vol. 67, no. 12, pp. 8315–8328, Dec. 2019.
[37] J. Wang, Z. Lan, C. woo Pyo, T. Baykas, C. sean Sum, M. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, "Beam codebook based beamforming protocol for multi-gbps millimeter-wave wpan systems," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 8, pp. 1390–1399, Oct. 2009.
[38] S. I. 802.11ad, "Wireless lan medium access control (MAC) and physical layer (PHY) specifications (amendment 3: Enhancements for very high throughput 60 GHz band)," 2012.
[39] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave mimo systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, Jan. 2014.

Yuanyuan Qiao (Graduate Student Member, IEEE) received the B.S. degree in communication engineering from North China Electric Power University, Hebei, China, in 2019 and the M.Eng. degree in electronics and communication engineering from Beijing Jiaotong University, Beijing, China, in 2021. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, China. His current research interests include wireless resource allocation, ultra-reliable low-latency communications, and high-speed railroad communications.

Yong Niu (Senior Member, IEEE) received the B.E. degree in Electrical Engineering from Beijing Jiaotong University, China, in 2011, and the Ph.D. degree in Electronic Engineering from Tsinghua University, Beijing, China, in 2016.

From 2014 to 2015, he was a Visiting Scholar with the University of Florida, Gainesville, FL, USA. He is currently an Associate Professor with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University. His research interests are in the areas of networking and communications, including millimeter wave communications, device-to-device communication, medium access control, and software-defined networks. He received the Ph.D. National Scholarship of China in 2015, the Outstanding Ph.D. Graduates and Outstanding Doctoral Thesis of Tsinghua University in 2016, the Outstanding Ph.D. Graduates of Beijing in 2016, and the Outstanding Doctorate Dissertation Award from the Chinese Institute of Electronics in 2017. He has served as Technical Program Committee member for IWCMC 2017, VTC2018-Spring, IWCMC 2018, INFOCOM 2018, and ICC 2018. He was the Session Chair for IWCMC 2017. He was the recipient of the 2018 International Union of Radio Science Young Scientist Award.

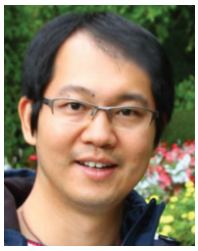
Xiangfei Zhang was born in Anhui, China, in 1998. He received the B.E. degree in communication engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2020. He received the Master degree with the State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, China. His research interests include mmWave wireless communications and wireless resource allocation.

Sheng Chen (Life Fellow, IEEE) received his BEng degree from the East China Petroleum Institute, Dongying, China, in 1982, and his PhD degree from the City University, London, in 1986, both in control engineering. In 2005, he was awarded the higher doctoral degree, Doctor of Sciences (DSc), from the University of Southampton, Southampton, UK. From 1986 to 1999, He held research and academic appointments at the Universities of Sheffield, Edinburgh and Portsmouth, all in UK. Since 1999, he has been with the School of Electronics and Computer Science, the University of Southampton, UK, where he holds the post of Professor in Intelligent Systems and Signal Processing. Dr Chen's research interests include adaptive signal processing, wireless communications, modeling and identification of nonlinear systems, neural network and machine learning, intelligent control system design, evolutionary computation methods and optimization. He has published over 700 research papers. Professor Chen has 19,300+ Web of Science citations with h-index 61 and 37,800+ Google Scholar citations with h-index 82. Dr. Chen is a Fellow of the United Kingdom Royal Academy of Engineering, a Fellow of Asia-Pacific Artificial Intelligence Association and a Fellow of IET. He is one of the original ISI highly cited researchers in engineering (March 2004).

Zhangdui Zhong (Fellow, IEEE) received the B.E. and M.S. degrees from Beijing Jiaotong University, Beijing, China, in 1983 and 1988, respectively.

He is currently a Professor and an Advisor of Ph.D. candidates with Beijing Jiaotong University, where he is also currently a Chief Scientist of the State Key Laboratory of Advanced Rail Autonomous Operation. He is also the Director of the Innovative Research Team, Ministry of Education, Beijing, and a Chief Scientist of the Ministry of Railways, Beijing.

He is also an Executive Council Member of the Radio Association of China, Beijing, and a Deputy Director of the Radio Association, Beijing. His interests include wireless communications for railways, control theory and techniques for railways, and GSM-R systems. His research has been widely used in railway engineering, such as the Qinghai-Xizang railway, DatongQinhuangdao Heavy Haul railway, and many high-speed railway lines in China. He has authored or co-authored seven books, five invention patents, and over 200 scientific research papers in his research area. Prof. Zhong was a recipient of the Mao YiSheng Scientific Award of China, Zhan TianYou Railway Honorary Award of China, and Top 10 Science/Technology Achievements Award of Chinese Universities.



Ning Wang (Member, IEEE) received the B.E. degree in communication engineering from Tianjin University, Tianjin, China, in 2004, the M.A.Sc. degree in electrical engineering from The University of British Columbia, Vancouver, BC, Canada, in 2010, and the Ph.D. degree in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 2013. From 2004 to 2008, he was with the China Information Technology Design and Consulting Institute, as a Mobile Communication System Engineer, specializing in planning and design of commercial mobile communication networks, network traffic analysis, and radio network optimization. From 2013 to 2015, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia. Since 2015, he has been with the School of Information Engineering, Zhengzhou University, Zhengzhou, China, where he is currently an Associate Professor. He also holds adjunct appointments with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada, and the Department of Electrical and Computer Engineering, University of Victoria, Victoria, BC, Canada. His research interests include resource allocation and security designs of future cellular networks, channel modeling for wireless communications, statistical signal processing, and cooperative wireless communications. He has served on the technical program committees of international conferences, including the IEEE GLOBECOM, IEEE ICC, IEEE WCNC, and CyberC. He was on the Finalist of the Governor Generals Gold Medal for Outstanding Graduating Doctoral Student with the University of Victoria in 2013.

Bo Ai (Fellow, IEEE) received the M.S. and Ph.D. degrees from Xidian University, China. He studies as a Post-Doctoral Student at Tsinghua University. He was a Visiting Professor with the Electrical Engineering Department, Stanford University, in 2015. He is currently with Beijing Jiaotong University as a Full Professor and a Ph.D. Candidate Advisor. He is the Deputy Director of the State Key Laboratory of Advanced Rail Autonomous Operation and the Deputy Director of the International Joint Research Center. He is one of the main people responsible for the Beijing Urban Rail Operation Control System, International Science and Technology Cooperation Base. He is also a Member, of the Innovative Engineering Based jointly granted by the Chinese Ministry of Education and the State Administration of Foreign Experts Affairs. He was honored with the Excellent Postdoctoral Research Fellow by Tsinghua University in 2007.

He has authored/co-authored eight books and published over 300 academic research papers in his research area. He holds 26 invention patents. He has been the research team leader for 26 national projects. His interests include the research and applications of channel measurement and channel modeling, dedicated mobile communications for rail traffic systems. He has been notified by the Council of Canadian Academies that, based on Scopus database, he has been listed as one of the Top 1% authors in his field all over the world. He has also been feature interviewed by the IET Electronics Letters. He has received some important scientific research prizes.