Learn to Beamform in Reconfigurable Intelligent Surface Aided MISO Communications with Channel Aging

Zixing Tang*, Ying Wang*, XX*
*Beijing University of Posts and Telecommunications, Beijing, China
Email: tangzx@bupt.edu.cn

***Abstract*—Channel aging induced by Doppler shifts degrades the system performance for the mismatch between the estimated channels and real channels that grows with time. In this paper, a novel frame structure is proposed to modify the active and passive beamforming quickly according to the partial channel state information (CSI) and real-time environment feedback with previous state instead of only outdated estimation CSI, which is capable of utilizing the channel correlation between adjacent instants to adapt to the dynamic random environment. Then a deep reinforcement learning (DRL)-based algorithm is designed, which is trained offline and works online. Simulation results demonstrate that the proposed algorithm mitigates the impacts of channel aging effectively with the low complexity, and outperforms the representative benchmark scheme.**

***Index Terms*—Reconfigurable intelligent surface, channel aging, high mobility, deep reinforcement learning.**

I. Introduction

Reconfigurable intelligent surface (RIS) has the potential of enabling the sixth generation (6G) mobile communication for the reason of the ability to control the propagation environment and to shape electromagnetic waves via numerous subwavelength metallic or dielectric scattering elements [1]. By dynamically adjusting the reflecting coefficients of the RIS elements based on the propagation environment, the desired signals can be enhanced while the interference signals are suppressed [1]–[3]. As such, RIS has been extensively investigated for various wireless systems and applications, such as secure communications [4], non-orthogonal multiple access (NOMA) [5], millimeter-wave (mmWave) [6], and so on. In addition to the above studies relying on the instantaneous channel state information (I-CSI), the study of passive beamforming based on the statistical CSI (S-CSI) is also carried in [3], which avoids the difficulty of obtaining the I-CSI for the RIS-aided cascaded channel.

However, most of the existing works only pay attention to the quasi-static channels, which are not applicable for high-mobility scenarios, such as vehicle-to-vehicle (V2V), high-speed train, unmanned aerial vehicle (UAV) communications [7]. Compared to quasi-static channels, the time-varying characteristics of channels due to Doppler shifts can not be ignored in high-mobility communications, where block-fading model is not reasonable. The CSI evolves with time quickly and the mismatch between the estimated channels and real channels degrades the transmission performance, which is termed as channel aging [8]. In order to address this issue, attention has been increasingly paid to the estimation of time-varying channels [8]–[10], as well as to the design of beamformers that compensate for imperfection due to time-varying effects [11], [12], aiming at improving the robustness of multi-input multi-output (MIMO) systems. Specifically, an outage probability constraint is constructed in [11] to guarantee the minimum signal-to-interference-plus-noise ratio (SINR) of users in the vehicular network. An adaptive construction method for statistical analog beamformer is proposed in [12] to enhance the robustness against angular estimation errors.

Unfortunately, due to the existence of the high dimension of the RIS-aided cascaded channel matrix and the absence of radio frequency (RF) chains, frequent pilot estimation to alleviate the mismatch will squeeze the time for data transmission, which even overwhelms the gain of eliminating channel aging effects. On the other hand, the optimization of reflecting coefficients is coupled with the active beamforming in the BS, which brings the difficulty to the solution. To summarize, the existing works on the MIMO systems are not fully applicable to the RIS-aided communications. In [13], a new transmission protocol is proposed to estimate and then compensate for the Doppler effect for a RIS-aided single-input single-output (SISO) communication, where the cascaded channel is assumed as the single path simply. Hence, it inspires our work on MISO communications, which considers scatter components in both direct and cascaded channels and is able to easily expand to MIMO communications.

In this paper, we propose a novel frame structure and leverage the first-order auto-regressive (AR) model to characterize the impacts of channel aging and describe the channel correlation between adjacent instants. Based on the proposed frame structure, we design a scheme to adjust the active and passive beamforming quickly according to the partial CSI and the real-time environment feedback that is easily acquired without increasing the pilot overhead. Next, a low-complexity deep reinforcement learning (DRL)-based algorithm is designed for the joint optimization of active and passive beamforming, which is capable of adapting to the dynamic environment. Ultimately, numerical results demonstrate the priority of the proposed scheme to mitigate the impacts of channel aging.

The rest of the paper is organized as follows. In Section II, we present the system model, the proposed frame structure, and the corresponding problem formulation. In Section III, a DRL-based solution is proposed with the low complexity. Next, simulation results are presented in Section IV. Finally, we conclude the paper in Section V.

II. System Model

A. Scenario

In this paper, we consider a RIS-aided downlink high-mobility communication system where the BS serving a fast-moving vehicle is equipped with a N_t-antenna uniform linear array (ULA). The vehicle moves at a high speed of v meters/second and is deployed with a RIS that is a uniform planar array (UPA) composed of $N_r = N_r^x \times N_r^y$ reflecting elements. Moreover, we assume that the user in the vehicle is equipped with a single antenna.

The baseband equivalent channels of BS-RIS link, RIS-user link, BS-user link are denoted by $\mathbf{h}_{br} \in \mathbb{C}^{N_t \times N_r}$, $\mathbf{h}_{ru} \in \mathbb{C}^{N_r \times 1}$, $\mathbf{h}_{bu} \in \mathbb{C}^{N_t \times 1}$, respectively. Denote by $\mathbf{\Theta} \in \mathbb{C}^{N_r \times N_r}$ the phase shift matrix of the RIS, specifically, $\mathbf{\Theta} = \operatorname{diag}\left(\beta_1 e^{j\theta_1}, \beta_2 e^{j\theta_2}, \ldots, \beta_{N_r} e^{j\theta_{N_r}}\right)$, where $\theta_n \in [0, 2\pi)$, $\beta_n \in [0,1]$ represent the phase shift and amplitude coefficient of the n-th reflecting element. For simplicity, we set amplitude coefficients to max to one, i.e., $\beta_n = 1$. Denote by $\mathbf{w} \in \mathbb{C}^{N_t \times 1}$ the active beamforming matrix of the BS.

The received signal at the user in instant t can be formulated as

$$y[t] = \left(\mathbf{h}_{ru}^H[t]\mathbf{\Theta}\mathbf{h}_{br}^H + \mathbf{h}_{bu}^H[t]\right)\mathbf{w}x + n, \tag{1}$$

where x denotes the transmitted data stream for the user, and n represents the noise at the receiver. It's assumed that $E\left\{|x|^2\right\} = 1$ and $E\left\{|n|^2\right\} = \sigma^2$. Then, the downlink achievable rate of the user is given by

$$R[t] = \log_2\left(1 + \frac{\left|\left(\mathbf{h}_{ru}^H[t]\mathbf{\Theta}\mathbf{H}_{br}^H + \mathbf{h}_{bu}^H[t]\right)\mathbf{w}\right|^2}{\sigma^2}\right). \tag{2}$$

It's noted that, due to the short distance between the RIS and user as well as the fact that they remain relatively static, the RIS-user channel changes much more slowly as compared to the BS-RIS channel. Thus, it is practically quasi-static and can be characterized by the standard near-field model in [14]. In addition, the time-varying BS-RIS channel and BS-user channel are described in the following subsection.

B. Proposed Frame Structure

The relative high-speed movement between the BS and user leads to temporal variations in the propagation environment which affect the channel coefficients during a transmission frame. The mismatch between the estimated channels and real channels caused by channel aging will increase significantly over time, which forces transmission performances to degrade and achievable rate to fluctuate greatly. In order to mitigate the effects of channel aging, a novel frame structure is proposed, which is shown in Fig. 1.

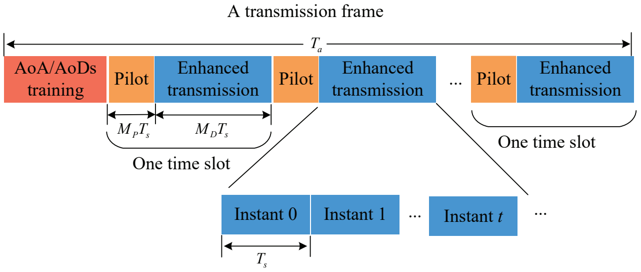

Fig. 1. Illustration of frame structure.

We consider a transmission frame composed of M time slots and each of slots contains $M_P + M_D$ time instants with duration of T_s. In each slot, the first M_P instants are allocated to pilot transmit and the latter M_D instants are allocated to data transmission. Considering the fact that the movement of the vehicle within a few milliseconds is negligible to BS tens of meters away, we assume that the large-scale fading characteristic parameters, such as path-loss, azimuth and elevation determined by the location of the vehicle, ramain approximately constant in a transmission frame. Therefore the large-scale CSI can be achieved with low channel estimation overhead [15].

In addition, the temporal variations in the propagation environment lead to the small-scale CSI evolving with time, thus enhancing the necessity of achieving the corresponding information in each time slot [9], [10]. However, the significant channel aging degrades the channel coherence time even if less than a time slot with the increasement of the movement speed and the communication frequency. The time for data transmission in each slot tends to be greatly squeezed if pilots are more frequently inserted to the slot to probe fast fading information, which even overwhelms the gain of eliminating channel aging effects. Therefore, the small-scale fast fading information is assumed to vary slightly between adjacent instants, where the duration of an instant is much smaller than the channel coherence time. Based on this assumption, a DRL-based beamforming scheme is proposed to mitigate the effects of channel aging without increasing pilot overhead in section III. Specifically, pilots are transmitted to achieve initial small-scale CSI in first M_P instants. The agent initializes the active and passive beamforming and adjusts the scheme according to real-time environment feedback in each instant. Compared to the outdated CSI in the head of each slot, the feedback information of the previous state is more conducive to the beamforming modification.

C. Time Varying Channel Model

In this subsection, a time varying channel model characterized as channel aging is given. Herein we employ the Rician model [16], [17], and as the Doppler effect is considered, the time-varying Rician channel is specified as

$$\mathbf{h}_{br}[t] = \mathrm{PL}_{br}\left(\sqrt{\frac{\kappa_{br}}{1+\kappa_{br}}}\mathbf{h}_{br}^{LoS}[t] + \sqrt{\frac{1}{1+\kappa_{br}}}\mathbf{h}_{br}^{NLoS}[t]\right), \tag{3}$$

$$\mathbf{h}_{bu}[t] = \mathrm{PL}_{bu}\left(\sqrt{\frac{\kappa_{bu}}{1+\kappa_{bu}}}\mathbf{h}_{bu}^{LoS}[t] + \sqrt{\frac{1}{1+\kappa_{bu}}}\mathbf{h}_{bu}^{NLoS}[t]\right), \tag{4}$$

where $\mathrm{PL(dB)} = \mathrm{PL}_0 + 10\alpha\lg(D)$ represents the path-loss chracterizing the large-scale fading determined by the location of the vehicle. PL_0 is the path-loss at the reference distance of one meter, D (in meters) represents the individual link distance, α denotes the path-loss exponent, and κ is the Rician factor, respectively.

Small-scale fading is composed of line of sight (LoS) and non-LoS (NLoS) components. Particularly, the LoS component of the time-varying channel mainly experiences the Doppler-induced phase shifts over different instants. As far as the NLoS component is concerned, it usually fluctuates in both amplitude and phase within a instant owing to the random and multi-path scatters in the environment. The LoS component is written as

$$\mathbf{h}_{br}^{LoS}[t] = \mathbf{a}_L\left(\vartheta_{br}^{AoD}\right)\mathbf{a}_P\left(\vartheta_{br}^{AoA}, \varphi_{br}^{AoA}\right)e^{j2\pi f_{br}^{d} tT_s}, \tag{5}$$

$$\mathbf{h}_{bu}^{LoS}[t] = \mathbf{a}_L\left(\vartheta_{bu}^{AoD}\right)e^{j2\pi f_{bu}^{d} tT_s}, \tag{6}$$

where $\mathbf{a}_L(\vartheta) = \left[1, e^{j\pi\sin\vartheta}, \ldots, e^{j\pi(N_t-1)\sin\vartheta}\right]^T$ denotes the response vector of ULA, with ϑ representing the azimuth angles of departure (AoDs) ϑ_{br}^{AoD} and ϑ_{bu}^{AoD}. The response vector of UPA is denoted by $\mathbf{a}_P(\vartheta, \varphi) = \mathbf{a}_{Px}(\vartheta, \varphi) \otimes \mathbf{a}_{Py}(\vartheta, \varphi)$, with ϑ (φ) representing the azimuth (elevation) angles of arrival (AoAs) ϑ_{br}^{AoA} (φ_{br}^{AoA}). $\mathbf{a}_{Px}(\vartheta, \varphi) = \left[1, \ldots, e^{j\pi(N_r^x-1)\sin\vartheta\cos\varphi}\right]^T$ and $\mathbf{a}_{Py}(\vartheta, \varphi) = \left[1, \ldots, e^{j\pi(N_r^y-1)\sin\vartheta\sin\varphi}\right]^T$ are horizontal array and vertical array response vector respectively. Moreover, $f_{br}^{d} = \frac{v}{\lambda}\sin\vartheta_{br}^{AoA}\cos\varphi_{br}^{AoA}$ is Doppler frequency of the LoS component, with λ denoting the carrier wavelength. Since ϑ_{br}^{AoA} and φ_{br}^{AoA} are unchanged in a transmission frame, f_{br}^{d} also remains constant. The definition of f_{bu}^{d} is similar.

Due to the link between BS and RIS/user experiences time-selective fading, the NLoS component of both links is assumed to evolve according to the first-order AR model [17], [18]:

$$\mathbf{h}^{NLoS}[t] = \rho(T_s)\mathbf{h}^{NLoS}[t-1] + \mathbf{e}_t, \tag{7}$$

where the correlation coefficient $\rho(T_s) \leq 1$ matches with the zeroth-order Bessel function of the first kind, i.e., $\rho(T_s) = J_0(2\pi f_{max}^{d} T_s)$, and $\mathbf{e}_t \sim \mathcal{CN}\left(0, \left(1-\rho(T_s)^2\right)\mathbf{I}\right)$ denotes the uncertainty resulted from channel aging, $f_{max}^{d} = v/\lambda$ is the maximum Doppler frequency of scatterd signals. Morever, the NLoS component of time varying channels in instant t can be rewritten as

$$\mathbf{h}^{NLoS}[t] = \rho^t\mathbf{h}^{NLoS}[0] + \mathbf{z}_t \tag{8}$$

where $\mathbf{h}^{NLoS}[0]$ denotes the initial state in instant 0 that is achieved by pilot training, and $\mathbf{z}_t = \sum_{i=1}^{t}\rho(T_s)^{t-i}\mathbf{e}_i$ denotes an innovation component. Note that the channel correlation degrades over instants, hence, the initial state $\mathbf{h}^{NLoS}[0]$ of each slot is assumed to follow a complex Gaussian distribution with zero mean and unit variance.

D. Problem Formulation

In this paper, we aim to maximize the average achievable rate in one time slot by adjusting the instant active and passive beamforming according to the previous feedback. The instant achievable rate of (2) can be rewritten as

$$R[t] = \log_2\left(1 + \frac{\left|\left(\mathbf{h}_{ru}^H[t]\mathbf{\Theta}[t]\mathbf{h}_{br}^H + \mathbf{h}_{bu}^H[t]\right)\mathbf{w}[t]\right|^2}{\sigma^2}\right), \tag{9}$$

thus yielding the following problem

$$\max_{\{\mathbf{\Theta}[t], \mathbf{w}[t]\}} \frac{1}{T}\sum_{t=1}^{T} R[t] \tag{10a}$$

$$s.t. \qquad \mathbf{w}[t]^H\mathbf{w}[t] \leq P_{\max}, t = \{1, \ldots, T\}, \tag{10b}$$

$$0 \leq \theta_n[t] < 2\pi, n = \{1, \ldots, N_r\}, t = \{1, \ldots, T\}, \tag{10c}$$

where P_{max} indicates the maximum transmitting power of the BS. Promble (10) is intractable for the non-convex constraint in (10c), and the form of the sum of multiple logarithmic functions leads to high orders of the objective function in (10a), which is fairly complex to solve with traditional optimization algorithm. Additionally, the innovation component in (8) fluctuates the instant achievable rate. Next, a DRL-based approach is proposed to overcome these challenges effectively.

III. DRL-based Solution

A. Algorithm Description

Problem (10) desires a solution for jointly optimizing all instant beamforming schemes in one slot. Although beamform independently in each instant can avoid high solution complexity, it ignores the correlation between adjacent instants, which degrades the performance. Hence, in view of the information contained in previous state, we map problem (10) into the Markov decision process (MDP) model and solve it with the deep deterministic policy gradient (DDPG) algorithm. The step of the agent corresponds to the beamforming modification of an instant, and all steps in one slot constitute an episode. The state s_t, the action a_t, and the reward r_t in instant t are defined as follows:

1) *State* s_t: The state in instant t contains the LoS component $\left\{\mathbf{h}_{br}^{LoS}[t], \mathbf{h}_{bu}^{LoS}[t]\right\}$ and large-scale fading information $\{\mathrm{PL}_{br}, \mathrm{PL}_{bu}\}$, which are the partial CSI of time varying channels. The state also contains the CSI of quasi-static channel, i.e., $\mathbf{h}_{ru}$. Furthermore, previous information is one of the important parts of the state, such as the action a_{t-1} excuted in previous instant and corresponding revice SNR, which are the real-time feedback from environment. It's noted that the values of CSI or beamforming are complex, thus splitted into real and imaginary parts. The input dimension n_0 of the agent can be calculated by $n_0 = 2N_tN_r + 4N_t + 3N_r + 1$.
2) *Action* a_t: The action in instant t consists of active and passive beamforming, i.e., $\mathbf{w}_t, \boldsymbol{\theta}_t$. It's worth noting that the initial CSI $\{\mathbf{h}_{br}[0], \mathbf{h}_{bu}[0], \mathbf{h}_{ru}\}$ are obtained by

transmitting in instant 0, and are used to initialize the action a_0 in state s_1 according to [1], which accelerates the convergence of the proposed algorithm.

3) *Reward* r_t: The reward r_t is the achievable rate after excuting the action a_t, which is calculated by (9).

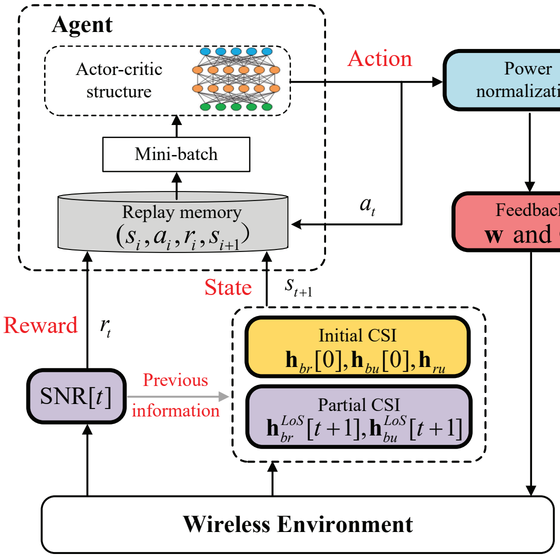

Fig. 2. Structure for training.

In order to reduce the complexity of policy decisions, we train neural networks offline and the trained networks work online and choose action according to real-time feedback of environment. Fig. 2 demonstrates the training process of the agent. To be specific, the agent first obtains the initial CSI and initializes the action in instant 0. Then the current state s_t is observed and input into the actor network to choose action a_t, where a power normalization module is set before performing the action to guarantee the constraint (10b). Next, the BS and RIS in environment modify the active and passive beamforming respectively. Environment feeds back the instant reward r_t to the agent and next state s_{t+1} is generated accordingly. Finally, the transition is stored into the replay memory and the agent samples random mini-batches of experiences to train the networks. The overall algorithm for solving problem (10) is shown in Algorithm 1.

B. Complexity Analysis

The solution for problem 10 is trained offline and works online. Denote by L, n_i, N_b the number of network layers, the neurons number in the i-th layer, the size of mini-batch, respectively. For simplicity, both actor and critic network share the network structure. As a result, the computational complexity for a single network to both evaluate and update in a step is $\mathcal{O}(N_b\left(\sum_{i=0}^{L-1} n_i n_{i+1}\right))$ [6]. Since it takes $N_{ep}N_{step}$ steps to finish training, the total training computational complexity of proposed algorithm is $\mathcal{O}\left(N_{ep}N_{step}N_b\left(\sum_{i=0}^{L-1} n_i n_{i+1}\right)\right)$. For working mode, the computational complexity for choosing an action can be dramatically decreased to $\mathcal{O}\left(\sum_{i=0}^{L-1} n_i n_{i+1}\right)$. Compared to the semidefinite relaxation (SDR)-based solution for a single step, where the worst-case computational complexity is $\mathcal{O}\left(n_0^{3.5}\right)$ [19], DRL-based solution might be more suitable for this problem.

Algorithm 1 DDPG-based algorithm

```
Initialize training actor network parameter $\xi_a^{(\text{train})}$, target actor network parameter $\xi_a^{(\text{target})}$, training critic network parameter $\xi_c^{(\text{train})}$, target critic network parameter $\xi_c^{(\text{target})}$
for $episode = 1, 2, \ldots, N_{ep}$ do
    Collect and preprocess initial CSI $\{\mathbf{h}_{br}[0], \mathbf{h}_{bu}[0], \mathbf{h}_{ru}\}$ to obtain the first state $s_1 = \{\mathbf{h}_{br}^{LoS}[1], \mathbf{h}_{bu}^{LoS}[1], \mathbf{h}_{ru}, \mathbf{w}_0, \boldsymbol{\theta}_0, SNR_0\}$;
    for $t = 1, 2, \ldots, N_{step}$ do
        Select action $a_t = \{\mathbf{w}_t, \boldsymbol{\theta}_t\} = \pi(s_t; \xi_a^{(\text{train})}) + \mathcal{N}$, where $\mathcal{N}$ is a gaussian action noise with variance $\sigma_a$;
        Execute action $a_t$, receive an instant reward $r_t$, generate a new state $s_{t+1}$ according to partial CSI $\left\{\mathbf{h}_{br}^{LoS}[t+1], \mathbf{h}_{bu}^{LoS}[t+1]\right\}$;
        Store the transition $[s_t, a_t, r_t, s_{t+1}]$ into the replay memory;
        Sample random mini-batches of experiences from the replay memory to update $\xi_a^{(\text{train})}, \xi_c^{(\text{train})}$;
        Soft update $\xi_a^{(\text{target})}, \xi_c^{(\text{target})}$
    end for
end for
```

IV. Simulation Results

In this section, we present numerical results to evaluate the performance of the proposed scheme. In the simulation, the coordinates of the RIS and BS are set as (0m, 0m, 0m) and (20m, 20m, 10m), respectively. The user lies in a circle centered at vertical axis with a radius of 0.5 m randomly, and is -0.4 m away from the the horizontal plane. The BS is equipped with $N_t = 4$ antennas, and the RIS is equipped with $N_r = 4 \times 4$ reflecting elements. The carrier frequency, the speed of the vehicle, the duration of a instant are set to $f_c = 30$ GHz, $v = 10$ m/s, $T_s = 0.05$ ms, respectively. Other system parameters are set as follows: $\alpha_{br} = 2.5$, $\alpha_{bu} = 3.5$, $P_{max} = 30$ dBm, $\sigma^2 = -114$ dBm, $\kappa_{br} = 10$, $\kappa_{bu} = 5$, $M_D = 8$.

We first compare the convergence of the agent with different numbers of RIS elements with $N_r = 4 \times 4$, $N_r = 7 \times 7$, $N_r = 10 \times 10$, which is shown in Fig. 3. The results demonstrate that the agent will converge effectively after adequate training episodes, where the learning rates of actor and critic networks are set to 3.33×10^{-4} and 3.33×10^{-3}, the neurons number of 4 hidden layers are set to [1024, 512, 512, 256], the decay factor is set to 0.85, the variance of the action noise is set to 0.15. As expected, the increasement of RIS elements brings higher achievable rate.

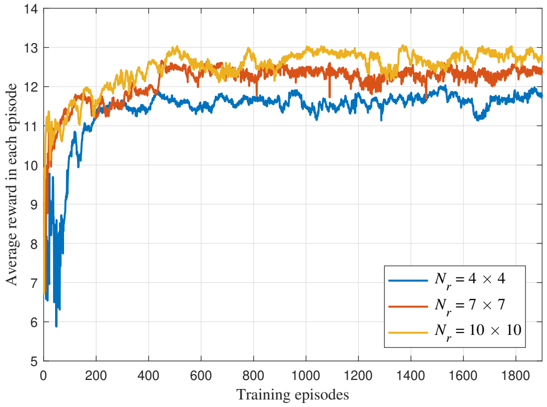

Fig. 3. Structure for training.

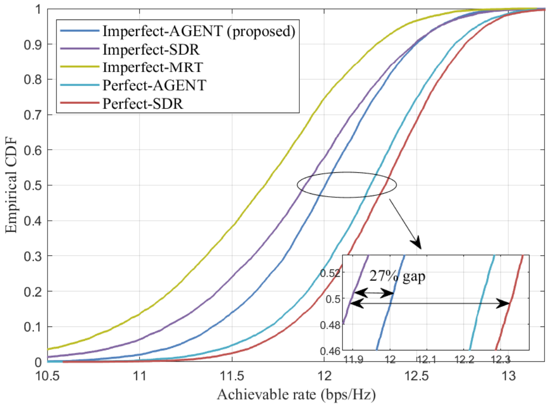

Fig. 4. Average reward performance versus episodes.

Next, the number of RIS elements is fixed to $N_r = 4 \times 4$ and four benchmarks are defined for comparison with the proposed scheme.

1) *Perfect-SDR*: The first benchmark is to leverage SDR algorithm to jointly optimize the BS beamforming matrix and the RIS reflecting phases for each instant with real CSI, which is assumed to be obtained perfectly. Despite the impossibility of practical implementation for expensive pilot overhead and high computational complexity, it can be treated as a system performance upper bound.
2) *Imerfect-SDR*: Considering the unacceptable price of high performance in the first benchmark, the second benchmark is to leverage SDR algorithm to optimize the BS beamforming matrix and the RIS reflecting phases only in instant 0 with initial CSI, and those remain unchanged throughout the time slot, where the pilot overhead equals to this of the proposed scheme.
3) *Imperfect-MRT*: Compared to imperfect-SDR, the third benchmark is to leverage maximum ratio transmission (MRT) to optimize the BS beamforming matrix with the fixed RIS reflecting phases, which is treated as a trivial benchmark.
4) *Perfect-AGENT*: The proposed scheme in section III is performed with initial CSI where the instant CSI is obtained imperfectly. Hence, the fourth benchmark is to assume that CSI in each instant is obtained perfectly. In other words, the real NLoS components are considered in state.

Figure 4 shows the cumulative distribution function of achievable rate in each instant under different schemes, and Fig. 5 presents that of average achievable rate in a time slot. It is seen that the optimization of RIS reflecting phases brings an obvious gain to system performance in despite of the CSI imperfection caused by channel aging. In addition, compared with imperfect-SDR, the proposed scheme with partial CSI decreases the 27% gap to perfect SDR assuming real CSI, which demonstrates that the proposed scheme in this paper is capable to improve the system performance limited by channel aging without increasing the pilot overhead. While considering the complete channel information, the performance of perfect-AGENT with low computational complexity is comparable to that of perfect-SDR, this suggests that our proposed solution can also be extended to general scenarios. Note that value of achievable rate seems larger than value in Fif. 3, which is for the reason that the action noise in training mode degrades the numerical results.

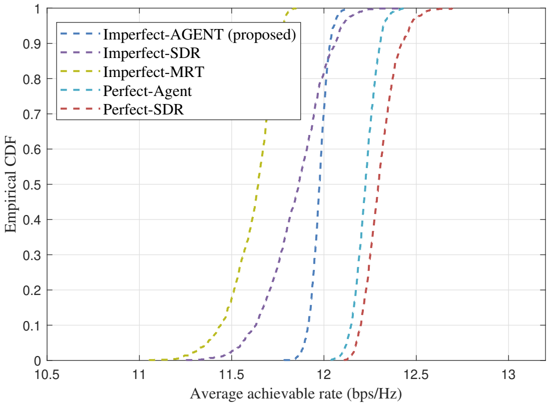

Fig. 5. The cumulative distribution function (CDF) for different methods.

In Fig. 6, we present the average achievable rate of 1000 slot samples, where the shaded area depicts the fluctuation of imperfect-SDR. It can be seen that the optimal active and passive beamforming is not able to compensate for the performance loss caused by channel aging. Alternatively, the proposed scheme fluctuates much smaller than imperfect-SDR for the reason that the modification of active and passive beamforming in each instant in accordance with the environment feedback mitigates the impacts of channel aging significantly. Additionally, let us focus on the fluctuation of perfect-SDR resulted from uncertainty in transmission channel, although our proposed scheme has no ability to predict the uncertainty, it can adapt to the random wireless environment by adjusting the

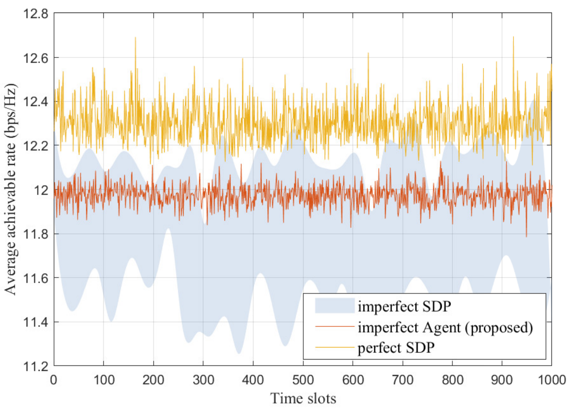

Fig. 6. Average achievable rate in slots.

active and passive beamforming quicky, which demonstrates the robustness of the scheme.

V. Conclusion

In this paper, we propose a novel frame structure suitable for time-varying channels, and a DRL-based solution for jointly optimizing active and passive beamforming. The proposed scheme is capable to mitigate the influence of channel aging by modifying the active BS beamforming and the passive RIS reflecting phases based on the real-time environment feedback. The simulation results demonstrate that the proposed scheme is robust and user achievable rate is improved significantly.

References

[1] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[2] C. Huang, Z. Yang, G. C. Alexandropoulos, K. Xiong, L. Wei, C. Yuen, Z. Zhang, and M. Debbah, "Multi-hop ris-empowered terahertz communications: A drl-based hybrid beamforming design," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 6, pp. 1663–1677, 2021.

[3] M.-M. Zhao, Q. Wu, M.-J. Zhao, and R. Zhang, "Intelligent reflecting surface enhanced wireless networks: Two-timescale beamforming optimization," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 2–17, 2021.

[4] H. Yang, Z. Xiong, J. Zhao, D. Niyato, L. Xiao, and Q. Wu, "Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 375–388, 2021.

[5] B. Zheng, Q. Wu, and R. Zhang, "Intelligent reflecting surface-assisted multiple access with user pairing: Noma or oma?" *IEEE Communications Letters*, vol. 24, no. 4, pp. 753–757, 2020.

[6] X. Guo, Y. Chen, and Y. Wang, "Learning-based robust and secure transmission for reconfigurable intelligent surface aided millimeter wave uav communications," *IEEE Wireless Communications Letters*, vol. 10, no. 8, pp. 1795–1799, 2021.

[7] W. Jiang, B. Han, M. A. Habibi, and H. D. Schotten, "The road towards 6g: A comprehensive survey," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 334–366, 2021.

[8] X. Xia, K. Xu, S. Zhao, and Y. Wang, "Learning the time-varying massive mimo channels: Robust estimation and data-aided prediction," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 8, pp. 8080–8096, 2020.

[9] Q. Qin, L. Gui, P. Cheng, and B. Gong, "Time-varying channel estimation for millimeter wave multiuser mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 10, pp. 9435–9448, 2018.

[10] L. Cheng, G. Yue, X. Xiong, Y. Liang, and S. Li, "Tensor decomposition-aided time-varying channel estimation for millimeter wave mimo systems," *IEEE Wireless Communications Letters*, vol. 8, no. 4, pp. 1216–1219, 2019.

[11] Y. Chen, M. Wen, L. Wang, W. Liu, and L. Hanzo, "Sinr-outage minimization of robust beamforming for the non-orthogonal wireless downlink," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 7247–7257, 2020.

[12] A. Kurt and G. M. Guvensen, "An adaptive hybrid beamforming scheme for time-varying wideband massive mimo channels," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–7.

[13] Z. Huang, B. Zheng, and R. Zhang, "Transforming fading channel from fast to slow: Irs-assisted high-mobility communication," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[14] Z. Abu-Shaban, K. Keykhosravi, M. F. Keskin, G. C. Alexandropoulos, G. Seco-Granados, and H. Wymeersch, "Near-field localization with a reconfigurable intelligent surface acting as lens," in *ICC 2021 - IEEE International Conference on Communications*, 2021, pp. 1–6.

[15] Y. Chen, Y. Wang, and L. Jiao, "Robust transmission for reconfigurable intelligent surface aided millimeter wave vehicular communications with statistical csi," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2021.

[16] Y. Chen, Y. Wang, J. Zhang, and M. D. Renzo, "Qos-driven spectrum sharing for reconfigurable intelligent surfaces (riss) aided vehicular networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 9, pp. 5969–5985, 2021.

[17] I. Zakia, "Impact of doppler shift error on least-squares mimo channel estimation for high-speed railway," *IET Communications*, vol. 14, pp. 206–218(12), January 2020.

[18] J. Zhang, H. Du, P. Zhang, J. Cheng, and L. Yang, "Performance analysis of 5g mobile relay systems for high-speed trains," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 12, pp. 2760–2772, 2020.

[19] Z.-Q. Luo and W. Yu, "An introduction to convex optimization for communications and signal processing," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 8, pp. 1426–1438, 2006.