Deep Reinforcement Learning for Channel Estimation in RIS-Aided Wireless Networks

Kitae Kim, *Student Member, IEEE*, Yan Kyaw Tun, *Member, IEEE*, Md. Shirajum Munir, *Member, IEEE*, Walid Saad, *Fellow, IEEE*, and Choong Seon Hong, *Senior Member, IEEE*

***Abstract*— Accurate channel estimation and allocation are vital in the provision of reconfigurable intelligent surfaces (RIS)-aided wireless network services to mobile users. Typically, channel estimation is carried out using a pilot signal. However, RIS elements cannot transmit or receive pilot signals because they are passive elements. Therefore, to maximize the gain of using the RIS, it is essential to accurately estimate a cascaded channel using a pilot signal between a base station (BS) and a terminal through an RIS. Moreover, although using a large number of pilot signals can guarantee accurate channel estimation performance, this can also drastically lower the wireless communication system's efficiency. Thus, in this letter, a new paradigm for learning-based pilot allocation and channel estimation in RIS systems is proposed. A masked autoencoder (MAE) is trained to achieve high channel estimation accuracy with a limited number of pilots. Then, a deep reinforcement learning(DRL) agent learns pilot allocation policies through MAE. Simulation results show that the MAE channel estimator has almost the same channel estimation performance even though it uses up to 33% fewer pilots than the autoencoder (AE)-based channel estimator. Furthermore, the proposed DRL-based pilot optimization method achieves higher channel estimation performance with 20% fewer pilots than the general autoencoder and other learning algorithms without the proposed RL-based pilot optimization algorithm.**

***Index Terms*— Reconfigurable intelligent surfaces, deep reinforcement learning, autoencoder, channel estimation.**

I. Introduction

The need to support high bandwidth, low latency applications such as extended reality or connected autonomy have strained the capacity and resources of existing cellular networks [1]. Moreover, these services need exceptionally high bandwidth, extremely low latency, and reliable connectivity. As a result, there is a need for new technologies such as massive multiple-input multiple-output (MIMO), millimeter wave (mmWave) communication, ultra-dense networks, and AI-empowered wireless networks, in order to satisfy the needs of tomorrow's wireless services. Intelligent reflecting surface (RIS), a promising technology for next-generation wireless networks, has recently attracted a lot of attention due to its ability to increase the coverage and capacity of wireless communication systems with minimal hardware cost and energy usage [2]. However, the effective operation of RIS systems requires addressing many challenges such as effective channel estimation and optimal phase shift control. In fact, accurate channel state information (CSI) is needed for channel estimation before transmission in an RIS-aided wireless communication system in order to effectively adjust the phases of RIS elements for better transmission performance [3], [4]. Similarly, channel estimation through CSI can also be used to precisely demodulate the transmitted data at the receiver. However, most of the prior works on RIS [5], [6], [7], [8] assume perfect CSI, which is difficult to achieve in reality especially. Therefore, research for channel estimation in an RIS-aided wireless communication system is widely conducted. Particularly, a deep learning-based channel estimation framework is garnering interest and has shown to have significant potential [9]. In [10], the authors proposed a channel estimation scheme based on least square (LS) estimation with partial on-off and deep learning-based super-resolution (SR) networks. Moreover, the work in [11] modelled the channel estimation process as a denoising problem and adopted a deep residual learning (DReL) approach to implicitly learn the residual noise for recovering the channel coefficients from the noisy pilot-based observations. However, these prior works [9], [10], [11], [12] did not consider the pilot pattern and the number of the allocating pilot signal. In [13] pilot patterns are optimized through network pruning after initially allocating equally-spaced pilots. However, this pattern is the optimal pilot pattern for equally-spaced pilots. Also, since pruning is performed every time channel estimation is performed, it is inefficient. Furthermore, in [14], neural network pruning during the training process for the pilot optimization method is proposed. However, pruning is performed during the training process. Thus it is challenging to cope with dynamically changing network conditions.

Manuscript received 28 April 2023; accepted 12 May 2023. Date of publication 29 May 2023; date of current version 12 August 2023. This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (No. RS-2023-00207816), and part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) Grant funded by the Korea Government (MSIT) (Artificial Intelligence Innovation Hub) under Grant 2021-0-02068, (No.RS-2022-00155911, Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University)), and (No.2019-0-01287, Evolvable Deep Learning Model Generation Platform for Edge Computing) and part by the Korea Institute of Energy Technology Evaluation and Planning(KETEP) and the Ministry of Trade, Industry & Energy(MOTIE) of the Republic of Korea (No. 20209810400030). The associate editor coordinating the review of this letter and approving it for publication was W. Jiang. *(Corresponding author: Choong Seon Hong.)*

Kitae Kim and Choong Seon Hong are with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: glideslope@khu.ac.kr; cshong@khu.ac.kr).

Yan Kyaw Tun is with the Division of Network and Systems Engineering, School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, 114 28 Stockholm, Sweden, and also with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: yktun@kth.se).

Md. Shirajum Munir is with VMASC, Department of Electrical and Computer Engineering, Old Dominion University, Suffolk, VA 23435 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: munir@khu.ac.kr).

Walid Saad is with the Wireless@VT, Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA 24061 USA, and also with the Department of Computer Science and Engineering, Kyung Hee University, Yongin-si, Gyeonggi-do 17104, South Korea (e-mail: walids@vt.edu).

Digital Object Identifier 10.1109/LCOMM.2023.3280821

1558-2558 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Instituto Nacional De Telecomunicações (INATEL). Downloaded on April 14,2025 at 11:54:26 UTC from IEEE Xplore. Restrictions apply.

Because the locations of important pilots are different in various network conditions. Moreover, In work [15], [16], pilot pattern optimization based on feature selection and compressive sampled CSI feedback method were studied. However, it doesn't take into account dynamic network changes due to various changes in the surrounding environment, such as user movement. Wireless channels in RIS-Aided or MIMO environments are sensitive to even small changes in their surroundings. In contrast, the main contribution of this letter is a new approach for the optimization of pilot patterns to maximize channel estimation accuracy with the minimum number of pilots for users in various surrounding environments in RIS-aided wireless communication systems. In this regard, we first train an autoencoder (AE) based channel estimator to obtain high channel estimation accuracy with a small number of pilots. Then, we train a DRL agent that determines the optimal pilot pattern for a certain number of pilots under dynamic environments that can be applied to different channel estimators.

II. System Model

The considered system model consists of a single base station equipped with a single antenna, one RIS with N passive elements, and a single UE with a single antenna. Moreover, an OFDM system with K subcarriers is adopted. Therefore, we can define the channel gain from the BS to the RIS and from the RIS to the UE through the subcarrier k as $h_{BR,k}$ and $h_{RU,k}$ respectively. We assume that there is no direct path between BS and UE. For channel modelling, we adopt a wideband geometric channel model from [17]:

$$\boldsymbol{h}_{T,d} = \sqrt{\frac{M}{\rho_T}} \sum_{l=1}^{L} \alpha_l p(dT_s - \tau_l) \mathbf{a}(\theta_l, \phi_l), \tag{1}$$

where L is the total number of clusters and each cluster l contains multipath components with the same delay τ_l. θ_l and ϕ_l denote the azimuth and elevation of the angle of arrival (AoA) for cluster l. Moreover, ρ_T is the pathloss between a transmitter and the receiver, and $p(\tau)$ represents a pulse shaping function for T_s-spaced signaling evaluated at τ seconds. Based on (1) the channel gain $h_{RU,k}$ between the RIS and a UE over subcarrier k:

$$\boldsymbol{h}_{RU,k} = \sum_{d=0}^{D-1} \boldsymbol{h}_{RU,d} e^{-j\frac{2\pi k}{K}d}, \tag{2}$$

where D is the channel tap length. Similarly, we can calculate the channel gain between the RIS and a base station over subcarrier k $h_{BR,k}$ as:

$$\boldsymbol{h}_{BR,k} = \sum_{d=0}^{D-1} \boldsymbol{h}_{BR,d} e^{-j\frac{2\pi k}{K}d}, \tag{3}$$

Therefore, the received signal at the BS, y_k will be:

$$y_k = (\boldsymbol{h}_{RU,k} \odot \boldsymbol{h}_{BR,k})^T \boldsymbol{\Gamma}_k s_k + n_k, \tag{4}$$

where $\odot$ is Hadamard product, $(\cdot)^T$ denotes the transpose and $\boldsymbol{\Gamma} = \operatorname{diag}(\beta_1 e^{-j\theta_1}, \ldots, \beta_n e^{-j\theta_n})$ is the phase shit matrix of the m-th element in RIS. β_n and θ_n are the amplitude and phase shift caused by the n-th element where $\beta_m \in \{0, 1\}$ and $\theta_m \in [0, 2\pi)$. Moreover, s_k is the transmitted signal over the k-th subcarrier, and n_k is the Gaussian white noise with zero mean and variance. Therefore, the cascaded channel between the BS and the user through the RIS through the subcarrier k can be briefly expressed as $\boldsymbol{H}_k = (\boldsymbol{h}_{RU,k} \odot \boldsymbol{h}_{BR,k})^T \boldsymbol{\Gamma}_{\boldsymbol{k}}$. In addition, the channel through subcarrier k at the symbol time subcarrier t can be expressed as:

$$\boldsymbol{H}_{k,t} = (\boldsymbol{h}_{RU,k,t} \odot \boldsymbol{h}_{BR,k,t})^T \boldsymbol{\Gamma}_{k,t}, \tag{5}$$

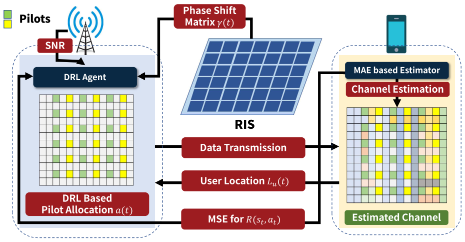


Fig. 1. Proposed DRL agent and MAE channel estimator.

Accordingly, we can rewrite the received signal $\boldsymbol{s}_{k,t}$ through the subcarrier k at time t is as:

$$y_{k,t} = \boldsymbol{H}_{k,t} s_{k,t} + n_{k,t}, \tag{6}$$

Lastly, the following mean-square error (MSE) is used as a metric for measuring channel estimation accuracy and loss function of the autoencoder-based channel estimator:

$$\epsilon = \frac{1}{S} \sum_{s=1}^{S} \left\| \hat{H} - H \right\|_F^2, \tag{7}$$

where S is the total number of channel samples. $\hat{H}$ and H denote the estimated channel and true channel. Further, $\|\cdot\|_F$ is the Frobenius norm.

III. Channel Estimator and Deep Reinforcement Learning Based Pilot Allocation

Although autoencoders are trained for high channel estimation accuracy with fewer pilots, it is not easy to achieve or exceed similar values for channel estimation accuracy with a large number of pilots. Therefore, to achieve high channel estimation accuracy through the autoencoder in this environment, a DRL agent is trained to determine the number and location of pilots. The channel state changes according to the positions of the UE and RIS, the RIS phase shift matrix of RIS, and the BS position. Therefore, the DRL agent for optimal pilot pattern allocation must consider the surrounding environments of various users. In addition, a reward function is required to minimize channel estimation accuracy and the number of pilot allocations. This reward is calculated through the number of pilots according to the pilot pattern allocated by the agent and the channel estimation performance by the MAE-based channel estimator. Through this learning process, the agent learns the optimal pilot pattern according to the user's location and wireless environment. For the DRL algorithm, we adopt the Deep Q-Network(DQN). The reason for using DQN is that it has a simple structure but can achieve stable learning and high performance [18].

A. Masked Autoencoder-Based Channel Estimator

In the typical autoencoder, the encoder maps the features extracted by reducing the dimension of the entire input data to the latent space. The decoder restores the original input data

Authorized licensed use limited to: Instituto Nacional De Telecomunicações (INATEL). Downloaded on April 14,2025 at 11:54:26 UTC from IEEE Xplore. Restrictions apply.

in latent space as opposed to the encoder. The encoder and decoder learn the training parameters θ and Φ, respectively, and for the bottleneck effect, the input data $x \in \mathbb{R}^m$ and the latent vector $z \in \mathbb{R}^n$ satisfies $n < m$. Therefore, the autoencoder operation and the loss function of AE can be expressed as below and the MSE in (10) is calculated through (7).

$$\text{Encoder} : p_{\theta}(x) = z, \tag{8}$$

$$\text{Decoder} : q_{\Phi}(z) = \hat{x}, \tag{9}$$

$$\text{argmin}_{\theta,\Phi}\text{MSE}(x - q_{\Phi}(p_{p}hi(x))), \tag{10}$$

In our approach, we adopt the masked autoencoder [19] to improve the channel estimation performance. According to [19], prediction performance improves when the pretrained weights learned through the pre-training process with 75% of the masked image is used for training for downstream tasks. An image is a natural signal, and its information is sparse. Therefore, even when masking many patches, it is not difficult to learn to infer masked patches through the surrounding patches. In other words, it is possible to reconstruct an original image with a tiny part of the image (pilot signals). Wireless communication channels can also be expressed as a two-dimensional matrix like an image and have spatial and temporal relationships. Therefore, MAE is useful to apply to channel estimation in which the value of an adjacent channel must be estimated through a pilot. The channel state, according to time and frequency, is represented as a two-dimensional image to apply the MAE-based model to channel estimation. Since the channel gain is complex, this image is divided into two channel images: Real and Complex. In Channel gain, the real value is mapped to the R channel in an RGB image, and the complex value is mapped to the G channel. Finally, all B channel values are set to 0. Also, if the values that make up the channel gain are very small or negative, the values are converted to images using normalization techniques. The normalized values are inversely transformed back to the normalized values during channel estimation.

Accordingly, if the number of subcarriers is N_f and the number of symbols is N_t, the size of image $\boldsymbol{I}$ is $n_f \times n_t \times 2$. MAE generates pre-trained weights through the pre-training process, and the pre-training sequence is as follows. First, pilots are allocated through a uniform distribution and it is assumed that the channel state value by the pilot is perfectly known. Then, we train the MAE over the ground-truth channel images. Next, the channel image is divided into multiple patches of size $p \times p$. The image patches that do not contain pilots are masked. Accordingly, only image patches containing pilots are used as input to the decoder to reconstruct the target image $\hat{\boldsymbol{I}}$ (ground-truth channel image). Note that positional encoding is also provided when inputting patches to the encoder and decoder. The encoder in MAE is trained to predict the rest of the channel image patch well, even if a part of it is lost through the pre-training process. In order to apply the pre-trained weight to the downstream task, channel estimation, we remove the encoder learned from pre-training. By attaching a new decoder to the encoder of the pre-training process, MAE is retrained to reconstruct the target channel matrix image. In this case, the encoder's input is a channel matrix interpolated through the allocated pilot $\boldsymbol{I}_{int}$. During this training, the encoder's pre-trained weights are fine-tuned for channel estimation, a downstream task. Accordingly, we can rewrite equations (8), (9), and (10) for the MAE channel estimator as below.

$$\text{Encoder} : p_{\theta}(\boldsymbol{I}_{int}) = z, \tag{11}$$

$$\text{Decoder} : q_{\Phi}(z) = \hat{\boldsymbol{I}}, \tag{12}$$

$$\text{argmin}_{\theta,\Phi}\text{MSE}(\boldsymbol{I}_{int} - q_{\Phi}(p_{\phi}(\boldsymbol{I}_{int}))), \tag{13}$$

B. DRL-Based Pilot Optimization for MAE Channel Estimator

The pilot pattern optimization is essential for optimizing the pilot allocation position to obtain the highest channel estimation performance through the MAE. Hence, in this section, we introduce a DRL pilot allocation optimization method for the MAE channel estimators according to user trajectory. The DRL-based pilot allocation algorithm aims to minimize MSE by allocating the minimum number of pilots to the optimal positions on the resource grid. Typically, the DRL model is defined as state space $\mathcal{S}$, action space $\mathcal{A}$, and a reward function $\mathcal{R}(t)$. The DRL agent finds the optimal policy by receiving a reward $R(t)$ for taking action $a(t) \in \mathcal{A}$ in each state $s(t) \in \mathcal{S}$. The following is the definition of state space and action space reward for DRL-based pilot allocation optimization for the MAE channel estimator.

1) *State Space* $\mathcal{S}$: We consider the state space with tuples defining the state of the RIS and the location of the users. i.e., the phase shift matrix of RIS, $\Gamma(t)$ and user location at time t, $L_u(t)$. Moreover, the average of the signal-to-noise ratio(SNR) for a user SNR_u is also considered. Therefore, the state at time slot t can be defined as $s(t) = \{\Gamma(t), L_u(t), SNR_u(t)\}$. Also, we assume that the user receives a resource block every time step t during total time duration T.
2) *Action Space* $\mathcal{A}$: In order to allocate P_n pilots on the resource grid, the agent determines the inter-pilot intervals n_f from the first subcarrier and n_t from $n_i(t)$-th symbol on the frequency and the time axis, respectively. Therefore, the action space is defined as $a(t) = \{n_i(t), n_f(t), n_t(t)\}$. Note that, we define $n_i(t) = \{2, 3\}$, $n_f(t) = \{4, 8, 12, 16, 20\}$ and $n_t(t) = \{2, 3, 4\}$ which means the $a(t)$ is discrete action space.
3) *Reward* $\mathcal{R}$: As mentioned previously, the goal of the DRL agent is to obtain the minimum channel estimation error with the minimum number of pilots. Therefore, the wider the pilot allocation interval and the lower the corresponding MSE, the higher the reward. Accordingly, the reward at each time step t is calculated as:

$$R(s_t, a_t) = \begin{cases} \dfrac{P_n}{\alpha \cdot \text{MSE}}, & \text{if } P_n \leq N_{\max} \\ -\alpha \cdot \text{MSE}, & \text{otherwise.} \end{cases} \tag{14}$$

Here, α is a coefficient multiplied because the value of the MSE is significantly smaller than the value of $n_f + n_t$. Therefore, the purpose of the learning agent across T time slots is to maximize the future reward $\hat{R}$, which is defined as:

$$\hat{R} = \sum_{t_0=0}^{T} \gamma \times R_{t-t_0}(a_t), \tag{15}$$

Authorized licensed use limited to: Instituto Nacional De Telecomunicações (INATEL). Downloaded on April 14,2025 at 11:54:26 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1 Learning Process for Pilot Optimization

```
1: Initialize: the initial network $Q(s, a; \theta_d)$ and target DQN parameters $\theta^-$
2: for episode= $1, 2, \ldots, E$ do
3:     Initialize randomly each UE's position and calculate SNR for each user.
4:     for time slot= $1, 2, \ldots, T$ do
5:         Observe the state $s(t)$ at base station
6:         Select action $a(t)$ and convert the channel matrix into image $\boldsymbol{I}_{int}$
7:         Compute instant reward using 11, 12, 13 and 14
8:         Save $s(t)$, $a(t)$, $r(t)$, $s(t+1)$ in replay memory
9:         Sample minibatch of experiences from replay memory randomly
10:        Train the DQN using stochastic gradient descent
11:        Update $\theta_d$, $\theta^-$, $Q(s, a; \theta_d)$, target DQN
12:    end for
13: end for
14: Output: Optimal network $Q^{\pi_{opt}}$
```

where γ and R_{t-t_0} are the discount factor the difference between rewards in two adjacent timeslots.

The reward function $R(t)$ changes according to the pilot allocation by the agent during the time duration T. Hence, $R_{t-t_0}(a_t)$ represents the difference in rewards between two adjacent time slots. Furthermore, Q-function according to policy π is defined as follows:

$$Q^\pi(s,a) = \hat{R}(s,a) + \gamma \sum_{s' \in S} P^a_{s,s'} \sum_{a' \in A} \pi(a' \mid s') Q^\pi(s', a'), \tag{16}$$

where π is pilot allocation policy and $P^a_{s,s'}$ is the transition probability from s to the next state s' according to the action. Here, the state and reward update is dependent on data received from the MAE channel estimator, user and RIS. As for the optimal policy π^*, it can be obtained by maximizing the value function represented by the Bellman Equation, as shown below:

$$V(s) = \mathbb{E}[R_{t'} + \gamma V(s_{t'} \mid s_t = s). \tag{17}$$

The proposed DRL process for pilot pattern optimization is summarized as Algorithm 1. Note that when the agent computes instant reward, the MAE estimator is used to minimize the MSE (line 7). Therefore, the time complexity of Algorithm 1 is mainly related to the network structure of the deep Q-network (DQN) and the MAE estimator. Assuming that the DQN and MAE estimator has I and J layers respectively, the time complexity of DQN can be calculated as $ET\sum_{i=0}^{I} n_{Q,i} n_{Q,i+1}$. Similarly, the time complexity of the MAE estimator will be $\sum_{j=0}^{J} n_{M,j} n_{M,j+1}$. $n_{Q,i}$ and n_M, j represent the number of neurons in i-th and j-th layer in the DQN and MAE estimator. Accordingly, the total complexity of Algorithm 1 is calculated as:

$$\mathcal{O}\left(ET(\sum_{i=0}^{I} n_{Q,i} n_{Q,i+1} + \sum_{j=0}^{J} n_{M,j} n_{M,j+1} n_{Q,i}) \right). \tag{18}$$

The radio channel image has much more information redundancy than a natural image. Thus a high percentage of masking is required to learn high-level estimation ability. In other words, if the masking ratio increases, redundancy information in the image decreases. Therefore, the encoder part of the MAE estimator learns to extract the latent to reconstruct the entire radio channel image through less information. As a result, self-supervised learning is possible to learn high-level semantics for inferring the masked area. Moreover, improving the inference ability by increasing the masking ratio means high channel estimation accuracy can be obtained even with a few pilots. To exploit these advantages, we maximize the performance of the MAE channel estimator by integrating the DRL agent for optimizing the pilot pattern and the MAE channel estimator.

TABLE I

Simulation Parameters

Parameter	Value	Parameter	Value
Frequency	28GHz	Number of users	50
RIS Elements	64	Number of symbols	28
Number of subcarriers	240	Subcarrier spacing	15kHz

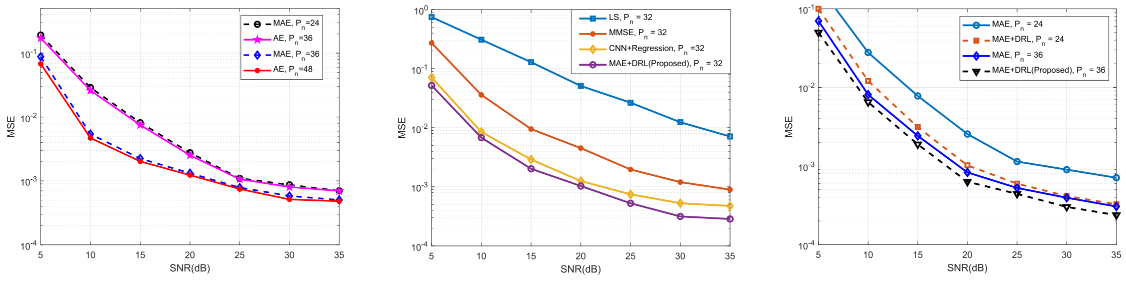
IV. Simulation Results

The simulation parameters provided in Table I. In addition, for the comparison of channel estimation performance with MAE, the traditional channel estimation methods, LS and MMSE algorithms, and CNN + regression Layer-based channel estimator are adopted as baseline approaches. To generate channel samples, we used the DeepMIMO [20] dataset. We generate a total of 115,840 channel sample data by collecting channel parameters from the fixed BS and RIS and 1,810 different UE locations through the scenario "O128." For user mobility, we use the random walk model. Furthermore, for each experiment, we randomly select 50 users from different locations and calculate the average MSE.

Fig. 2a shows the MSE performance of the MAE estimator with different SNRs and the number of pilots. channel estimation performance of general AE and MAE are compared by assigning the same pilot number. As a result of the experiment, the MAE channel estimator uses 25% and 33% less than the general AE and shows similar channel estimation performance. This means the pre-trained encoder that predicts the masked patch is also effective for channel estimation learning. However, the MAE channel estimator still cannot exceed the MSE performance when more pilots are allocated.

Fig.2b shows the experimental results of MSE performance of the MAE-based estimator versus other channel estimation methods under different SNRs. At the same pilot number $P_n = 32$, the MAE-based channel estimator outperforms other baseline channel estimators. This figure also shows that the neural network-based method has much better channel estimation performance than the existing traditional LS or MMSE-based channel estimators. For experiments in Fig. 2a and 2b, in all cases, we use the same pilot pattern allocated randomly. Fig. 2c shows the MSE performance of the MAE channel estimator and the MAE channel estimator to which the proposed DRL-based pilot pattern optimization is applied. DRL in the legend represents the DRL-based pilot pattern optimization. In both cases ($P_n = 24$, $P_n = 36$), the performance of the MAE channel estimator is better when the proposed DRL-based pilot pattern optimization technique is applied. In addition, we confirm that the difference in MSE performance was more significant depending on whether pilot pattern optimization is applied in a small number of pilots. This means that pilot allocation positions are more critical when using fewer pilots. In addition, as the SNR increases, it can be confirmed that the MSE of the case with

Authorized licensed use limited to: Instituto Nacional De Telecomunicações (INATEL). Downloaded on April 14,2025 at 11:54:26 UTC from IEEE Xplore. Restrictions apply.



(a) (b) (c)

Fig. 2. Performance of channel estimators: (a) Proposed versus AE, (b) Proposed versus other channel estimation methods, (c) Proposed MAE estimator with pilot optimization.

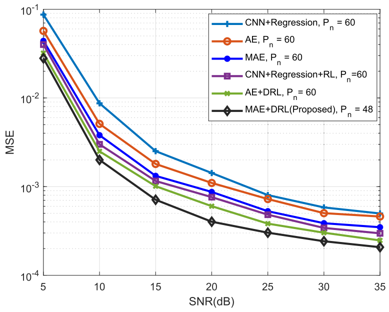


Fig. 3. Proposed versus learning-based channel estimators with DRL.

MAE, $P_n = 36$ and the case with $P_n = 24$ to which pilot pattern optimization is applied are almost the same. Therefore, if the SNR is sufficiently high through the proposed channel estimator and pilot pattern optimization, the performance of the wireless network system can be maximized by using fewer pilots. Finally, in Fig. 3, we apply the DRL-based pilot pattern optimization method to the CNN+Regression and AE-based channel estimators under the same number of pilots. We can see that the proposed MAE-DRL achieves the lowest MSE with 20% fewer pilots than other baseline algorithms in every case. In addition, when we apply the DRL-based pilot pattern optimization on the CNN and AE-based channel estimators, it achieves higher channel estimation performance than when the DRL-based method is not applied. Through this, the necessity of pilot optimization for channel estimation is emphasized once again.

V. Conclusion

In this letter, we have proposed a framework for pilot pattern optimization based on DRL and an MAE-based channel estimator. The proposed framework aimed to achieve excellent MSE performance even with a fixed or limited number of pilot signals. For MAE-based channel estimator learning, pretrained weights for MAE were first obtained. Based on this, a new decoder was trained, and the channel estimation was then updated. Then, we trained a DRL-based pilot assignment optimization agent for the MAE channel estimator. Simulation results show that the proposed method performed better than previous baseline algorithms. As a result, even when the same number of pilots were assigned, the MSE's performance varied. This indicated that the pilot pattern may impact the wireless communication system's effectiveness.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020.

[2] E. Basar et al., "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[3] Q. Zhang et al., "Millimeter wave communications with an intelligent reflector: Performance optimization and distributional reinforcement learning," *IEEE Trans. Wireless Commun.*, vol. 21, no. 3, pp. 1836–1850, Mar. 2022.

[4] M. Jung et al., "On the optimality of reconfigurable intelligent surfaces (RISs): Passive beamforming, modulation, and resource allocation," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4347–4363, Jul. 2021.

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

[6] Q. Wu and R. Zhang, "Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1838–1851, Mar. 2020.

[7] C. Huang et al., "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[8] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, Aug. 2020.

[9] H. Ye et al., "Power of deep learning for channel estimation and signal detection in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 1, pp. 114–117, Feb. 2018.

[10] Y. Wang et al., "Channel estimation in IRS-enhanced mmWave system with super-resolution network," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2599–2603, Aug. 2021.

[11] C. Liu et al., "Deep residual learning for channel estimation in intelligent reflecting surface-assisted multi-user communications," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 898–912, Feb. 2022.

[12] M. Ye et al., "Channel estimation for intelligent reflecting surface aided wireless communications using conditional GAN," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2340–2344, Oct. 2022.

[13] K. Kim and C. S. Hong, "Pruned autoencoder based mmWave channel estimation in RIS-assisted wireless networks," in *Proc. 23rd Asia–Pacific Netw. Oper. Manage. Symp. (APNOMS)*, Sep. 2022, pp. 1–4.

[14] M. B. Mashhadi and D. Gündüz, "Pruning the pilots: Deep learning-based pilot design and channel estimation for MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6315–6328, Oct. 2021.

[15] M. Soltani et al., "Pilot pattern design for deep learning-based channel estimation in OFDM systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 12, pp. 2173–2176, Dec. 2020.

[16] J. Wang et al., "Compressive sampled CSI feedback method based on deep learning for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 69, no. 9, pp. 5873–5885, Sep. 2021.

[17] A. Taha et al., "Deep learning for large intelligent surfaces in millimeter wave and massive MIMO systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[18] Z. Wang et al., "Dueling network architectures for deep reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2016, pp. 1995–2003.

[19] K. He et al., "Masked autoencoders are scalable vision learners," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 15979–15988.

[20] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.

Authorized licensed use limited to: Instituto Nacional De Telecomunicações (INATEL). Downloaded on April 14,2025 at 11:54:26 UTC from IEEE Xplore. Restrictions apply.