Deep Reinforcement Learning Based Relay Selection in Intelligent Reflecting Surface Assisted Cooperative Networks

Chong Huang, *Student Member, IEEE*, Gaojie Chen, *Senior Member, IEEE*, Yu Gong,
Miaowen Wen, *Senior Member, IEEE* and Jonathon A. Chambers, *Fellow, IEEE*

***Abstract*—This paper proposes a deep reinforcement learning (DRL) based relay selection scheme for cooperative networks with the intelligent reflecting surface (IRS). We consider a practical phase-dependent amplitude model in which the IRS reflection amplitudes vary with the discrete phase-shifts. Furthermore, we apply the relay selection to reduce the signal loss over distance in IRS-assisted networks. To solve the complicated problem of joint relay selection and IRS reflection coefficient optimization, we introduce DRL to learn from the environment to obtain the solution and reduce the computational complexity. Simulation results show that the throughput is significantly improved with the proposed DRL-based algorithm compared to random relay selection and random reflection coefficients methods.**

***Index Terms*—Intelligent reflecting surface (IRS), relay selection, throughput, deep reinforcement learning**

I. Introduction

WITH the development of 5th generation (5G) communications, the intelligent reflecting surface (IRS) has attracted much attention in current research due to its efficiency for wireless transmission as a cost-effective solution [1], [2]. An IRS is an array containing a vast number of passive reflecting elements, each of which can control the amplitude and phase shift of the incident signal to boost the reception quality in IRS-assisted communications. Since compared with many other technologies, IRS requires much less energy to forward the signal, it is widely acknowledged as a low-cost and efficient solution for future wireless networks [3], [4].

More recently, various related work has been investigated for improving the spectral efficiency and achievable rate [5], [6]. In [5], the phase shifts of the IRS were optimized to maximize the spectral efficiency, and 2 bit quantization was verified to guarantee high spectral efficiency. A joint optimization of the IRS reflection coefficients and transmit power allocation was proposed to maximize the orthogonal frequency division multiplexing (OFDM) achievable rate in [6].

However, the high computational complexity for optimizing the phase shifts of IRS is a complicated problem for practical implementation [7]. Fortunately, the deep reinforcement learning (DRL) algorithm can be used to solve complicated problems and reduce the computational complexity for wireless communications without the training data set [8]. Therefore, the phase shifts were optimized via the DRL algorithm to enhance the received signal-to-noise ratio SNR and reduce the computational complexity in [9]. In [10], a DRL-based joint design of the transmit beamforming matrix and phase shifts was proposed to improve the sum rate in IRS-assisted networks. Most related works, however, assume the reflection amplitude is fixed, which is not practical as the reflection amplitude varies with the phase shift, according to the practical phase shift model in [11]. Furthermore, the above DRL-based schemes only consider the continuous phase shift design. Therefore, in this paper, we will consider the discrete phase shift variables and the practical phase shift model to design our system.

On the other hand, a cooperative relay network is an attractive technology to improve the outage performance in wireless communications [12]. To amalgamate the benefits of the IRS-assisted and relay-assisted networks, a hybrid half-duplex (HD) decode-and-forward (DF) relay and IRS network with continuous phase shifts and fixed reflection amplitude was investigated to improve the achievable rate in [13]. To further enhance achievable rate, [14] proposed optimization of continuous phase shifts with fixed reflection amplitude for a hybrid IRS with full-duplex (FD) DF relay networks. Moreover, relay selection is an efficient way to harvest the diversity gain in cooperative communications [15]. Motivated by this, [16] utilized a DRL-based relay selection scheme to enhance the outage performance.

In this paper, therefore, we propose DRL-based relay selection in IRS-assisted cooperative networks (DRL-RI) to maximize the throughput with the discrete phase shifts and practical phase-dependent amplitude model. The main contributions of this paper are listed as follows:

- We propose joint relay selection and optimization of IRS reflection coefficients for cooperative networks, whilst considering the discrete phase shifts and the practical phase shift model.
- We introduce the DRL algorithm to solve the complicated non-convex optimization problem and thereby reduce the computational complexity of optimization in wireless networks.
- Simulation results show that the proposed DRL-based

This work was supported by EPSRC grant number EP/R006377/1 (“M3NETs”).

C. Huang, G. Chen and J. A. Chambers are with School of Engineering, Leicester, University, UK, Email: {ch481, gaojie.chen, jonathon.chambers}@leicester.ac.uk.

Y. Gong is with Wolfson School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, UK, Email: y.gong@lboro.ac.uk.

M. Wen is with the National Engineering Technology Research Center for Mobile Ultrasonic Detection, South China University of Technology, Guangzhou 510640, China (e-mail: eemwwen@scut.edu.cn).

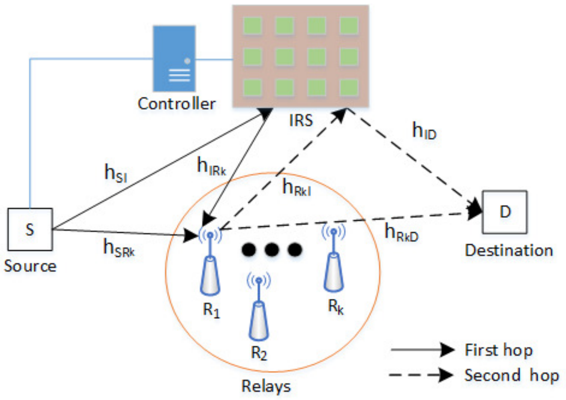

Fig. 1. System model of the hybrid relay and IRS network.

scheme can achieve a higher throughput than the random relay-selection/reflection-coefficients methods.

The rest of the paper is organized as follows. Section II introduces the system model and the problem formulation. The DRL-based algorithm is proposed in Section III. Simulation results verify the proposed scheme in Section IV. Finally, Section V concludes the paper.

II. System Model and Problem Formulation

As shown in Fig. 1, we consider a two-hop IRS assisted cooperative network, which is composed of one source S, one destination D, K HD DF relays R_k ($k \in \{1, ..., K\}$), and one IRS I with M reflecting elements. Each of S, D and R_k nodes is equipped with a single omnidirectional antenna. The IRS is equipped with a controller to determine the phase shift for each reflecting element at a given time slot. We assume there is no direct link between S and D by considering the signal loss over distance. Moreover, we assume the channels of $S \to R_k$ and $R_k \to D$ links are assumed to be Non-Line-of-Sight (NLoS) Rayleigh fading channels. On the other hand, we assume the channels from and to I are assumed to be Rician fading with pure Line-of-Sight (LoS) components [14], [17]. Therefore, we can obtain the channel coefficients h_{ij} between node i and node j as

$$h_{ij} = \begin{cases} \sqrt{\frac{K_{ij}}{K_{ij}+1}}\hat{h}_{ij}, & \text{Rician (LoS)} \\ \bar{h}_{ij}, & \text{Rayleigh (NLoS)} \end{cases}, \tag{1}$$

where K_{ij} denotes the Rician factor between node i and j. In NLOS Rayleigh fading channels, $\bar{h}_{ij} = \bar{g}_{ij} d_{ij}^{-\bar{\alpha}/2}$, $ij \in \{SR_k, R_kD\}$, where $\bar{g}_{ij}$ is modeled by complex-Gaussian small-scale fading with zero mean and unit variance, d_{ij} denotes the distance between nodes i and j, $\bar{\alpha}$ denotes the path loss exponent for NLoS Rayleigh fading channels, and all channels are assumed to remain unchanged during the two hops. On the other hand, in LoS Rician fading channels, $\hat{h}_{ij} = \hat{g}_{ij} d_{ij}^{-\hat{\alpha}/2}$, where $\hat{\alpha}$ denotes the path loss exponent for a LoS Rician fading channel, and $\hat{g}_{ij}$ can be expressed as

$$\hat{g}_{ij} = \sqrt{\beta_0}[1, e^{-j\pi \sin\psi_{ij}}, ..., e^{-j\pi(M-1)\sin\psi_{ij}}]^T, \tag{2}$$

where β_0 is the path loss at the reference distance $D_0 = 1$ m [18], $\psi_{ij} \in [0, 2\pi]$ is the angle of departure (AoD) or angle of arrival (AoA) for the signal between nodes i and j[1].

At the first hop $S \to R_k$, the source S transmits the signal x_S to both I and relay R_k, and I can reflect the incident signal to R_k. Thus, the received signal at relay R_k is given by

$$y_{R_k} = \sqrt{P_S}(h_{SR_k} + \mathbf{h}_{IR_k}^H \boldsymbol{\Theta} \mathbf{h}_{SI}) x_S + n_{R_k}, \tag{3}$$

where P_S denotes the transmit power at S, n_{R_k} denotes the additive-white-Gaussian-noise (AWGN) with variance σ_n^2 at R_k, $\boldsymbol{\Theta} = \text{diag}(\eta_1 e^{j\theta_1}, \eta_2 e^{j\theta_2}, ..., \eta_M e^{j\theta_M})$ denotes the diagonal reflection matrix for the IRS, with $\eta_m \in [0, 1]$ and $\theta_m \in [0, 2\pi]$ denoting the reflection amplitude and phase-shift for the mth reflecting element of I, respectively. Without loss of generality, we assume $\mathbf{v} = [v_1, ..., v_M]$ denotes the reflection coefficient vector for the IRS, such that $\eta_m = |v_m|$ and $\theta_m = \arg(v_m)$ for the mth IRS element [11]. Notice that the reflection amplitude varies with the phase shift. Therefore, in this paper we apply the practical model to obtain the amplitude and phase shift based on the reflection coefficient as in Fig. 3(b) of [11] with the effective resistance $R = 2\ \Omega$. Moreover, we assume that the phase shifts are discrete variables for implementing the IRS in practice as in [19], and the range of the phase shift for each IRS element $\mathcal{F}$ can be given as

$$\mathcal{F} \triangleq \left\{0, \frac{2\pi}{L}, ..., \frac{(L-1)2\pi}{L}\right\}, \tag{4}$$

where L denotes the number of phase quantization levels.

Based on (3), the received SNR at R_k for the first hop transmission can be given as

$$\gamma_{R_k} = \frac{P_S \left| h_{SR_k} + \mathbf{h}_{IR_k}^H \boldsymbol{\Theta} \mathbf{h}_{SI} \right|^2}{\sigma_n^2}. \tag{5}$$

Therefore, the channel capacity for the first hop transmission is $C_{SR_k} = \log_2(1 + \gamma_{R_k})$.

At the second hop $R_k \to D$, relay R_k transmits the decoded signal x_{R_k} to both I and D, and I can reflect the incident signal to D. Thus, the received signal at D is given by

$$y_D = \sqrt{P_R}(h_{R_kD} + \mathbf{h}_{ID}^H \boldsymbol{\Theta} \mathbf{h}_{R_kI}) x_{R_k} + n_D, \tag{6}$$

where P_R denotes the transmit power for node R_k and n_D denotes the AWGN with variance σ_n^2 at D. Therefore, the received SNR at D can be given as

$$\gamma_D = \frac{P_R \left| h_{R_kD} + \mathbf{h}_{ID}^H \boldsymbol{\Theta} \mathbf{h}_{R_kI} \right|^2}{\sigma_n^2}. \tag{7}$$

Thus, the channel capacity for the second hop transmission is $C_{R_kD} = \log_2(1 + \gamma_D)$. Moreover, we assume that the transmission for each hop is available when the corresponding channel capacity satisfies

$$C_{ij} \geq \vartheta, \tag{8}$$

where ϑ denotes the target rate. This means that if C_{ij} satisfies (8), the corresponding link can support the single packet

[1] The angle ψ_{ij} was randomly generated between $[0, 2\pi]$ for each channel in this paper as in [5], [13].

transmission from nodes i to j at a given time slot.

Due to considering the DF relay, a packet can be transmitted from S to D successfully when $\min\{C_{SR_k}(t), C_{R_kD}(t+1)\} \geq \vartheta$ at time slot t and $t+1$. To investigate the maximum throughput in IRS-assisted cooperative networks with practice phase shift model, the joint relay selection and reflection coefficients optimization can be formulated as

$$O = \max_{k(t),\mathbf{v}(\mathrm{t})} \frac{1}{T} \sum_{t=1}^{T-1} \mu\big(\min\{C_{SR_k}(t), C_{R_kD}(t+1)\} \geq \vartheta\big), \tag{9}$$

$$\text{s.t. } t = 1, 3, 5, ..., T-1, \tag{9a}$$

$$\mathbf{v}(\mathrm{t}) = [v_1(t), ..., v_M(t)], \tag{9b}$$

$$\theta_m(t) = \arg(v_m(t)) \in \mathcal{F}, \forall m, \tag{9c}$$

$$\eta_m(t) = |v_m(t)|, \tag{9d}$$

where T denotes the number of time slots observed at the destination, $\mu(.) = 1$ if the enclosed holds and $\mu(.) = 0$ if otherwise. With the relay selection, the discrete phase shifts variables, and the relation between the phase shifts and reflection amplitudes, (9) is a complicated non-convex optimization problem [9] and hard to solve. The exhaustive search algorithm to maximize the throughput has a high complexity of $O(KL^M)$. In addition, the existing IRS optimization schemes usually require high computational complexity to find the solution [20]. To solve the optimization problem in (9) with low complexity, we introduce DRL in the following section.

III. Deep Reinforcement Learning Based Optimization Scheme

To avoid the overestimation problem, the double deep Q-Learning network (DDQN) is applied in this paper. Firstly, there is an agent in the DDQN algorithm to make decisions to optimize the relay selection and IRS reflection coefficients for the proposed network. The agent can apply the ϵ-greedy strategy to explore the network and make decisions randomly, and then learn the decision policy from its exploration experience. Secondly, when the agent selects the exploitation mode, it makes decisions from its stored experience. We can model the proposed system as a Markov Decision Process (MDP) [16]. In DDQN, the algorithm has two different Q-tables, A and B, to store its experience. We assume $s(t) = \{t, h_{SR_k}(t), h_{R_kD}(t)\}$ denotes the system state of the MDP at time slot t, and $a(t) = \{k, \mathbf{\Theta}(t)\}$ denotes the action of the MDP (decision) at time slot t, where $k \in \{1, ..., K\}$. Then the function of updating Q-values in the Q-table A at time slot t is given by

$$\begin{aligned} Q^A(s(t), a(t)) = Q^A(s(t), a(t)) + \rho(r_{s(t),a(t)} + \delta \cdot Q^B \\ (s(t+1), \mathrm{argmax}_a\{Q^A(s(t+1), a)\}) \\ - Q^A(s(t), a(t))), \end{aligned} \tag{10}$$

where $r_{s(t),a(t)}$ is the reward of the MDP to evaluate the corresponding state $s(t)$ and action $a(t)$, $\rho \in (0,1)$ denotes the learning rate for Q-tables in the DDQN, $\delta \in (0,1)$ denotes the discount rate in the DDQN, and $\mathrm{argmax}_a\{Q^A(s(t+1), a)\}$ denotes the action with the maximum Q-value for the next state $s(t+1)$ in Q-table A. In the proposed scheme, the reward is given to the agent when a packet arrives at the destination successfully. To reduce the impact of the overestimation problem, Q-table A provides the policy to make the decision for the next action, but the updating value $Q^B(s(t+1), \mathrm{argmax}_a\{Q^A(s(t+1), a)\})$ is selected from another policy in Q-table B. Then we can form the function of updating Q-table B at time slot t as

$$\begin{aligned} Q^B(s(t), a(t)) = \; & Q^B(s(t), a(t)) + \rho(r_{s(t),a(t)} \\ & + \delta \cdot Q^A(s(t+1), \mathrm{argmax}_a\{Q^B(s(t+1), a)\}) \\ & - Q^B(s(t), a(t))). \end{aligned} \tag{11}$$

Since the dimension of the action-state space is high in the proposed MDP, it is difficult to form and update Q-tables for the DDQN. To solve this problem, the deep neural network (DNN) is introduced in the DDQN as the function approximator instead of Q-tables. Similar to Q-tables, the DNN can receive the state as the input and output the actions as the decisions for the proposed network. It significantly reduces the computational complexity of estimating the optimal decision for IRS assisted communication. Moreover, the DDN can use the gradient descent algorithm to update the neural network for high performance with high-dimensional environment. In this paper, we apply Adam [21] as the adaptive learning rate iterative optimization algorithm to calculate the gradients for the DDN.

In the proposed scheme, every T time slots the agent can generate samples for each time slot as $\{s(t), a(t), r_{s(t),a(t)}, s(t+1)\}$, and then selects W samples randomly for the training in the DNNs to avoid the overfitting problem. Two neural networks are designed for the proposed scheme as the prediction network and the target network, and provide the estimation value $Q^P(s(t), a(t))$ and the target value $Q^T\big(s(t+1), \mathrm{argmax}_a Q^P(s(t+1), a)\big)$, respectively. Thus, we can calculate the loss between the prediction network and the target network, and then obtain the gradients via the Adam algorithm to update the prediction network. The loss function in the proposed algorithm is given by

$$\begin{aligned} \mathbb{L}_M = \sum_{t=1}^{W} \bigg(& r_{s(t),a(t)} + \delta \cdot Q^T\Big(s(t+1), \mathrm{argmax}_a Q^P(s(t+1), a)\Big) \\ & - Q^P(s(t), a(t))\bigg)^2. \end{aligned} \tag{12}$$

After updating the prediction networks V times, we can copy the weights from the prediction networks to update the target network. The Pseudo code of the proposed DRL-RI scheme is shown in **Algorithm 1**. The computational complexity of the proposed algorithm is $V(T+W)$ for each iteration during training. After training, the computational complexity of the DRL-based algorithm for making decisions is much smaller than that in training, because it only depends on the structure of the neural network without any more learning. Thus, the

Algorithm 1 DRL-RI:

```
Initialize the environment.
Repeat:
for $v = 1, \cdots, V$ do
    for $t = 1, \cdots, T$ do
        Use the $\varepsilon$-greedy strategy to obtain the action $a(t)$.
        Obtain the reward $r_{s(t),a(t)}$ and state $s(t+1)$.
        Generate a sample $\{s(t), a(t), r_{s(t),a(t)}, s(t+1)\}$.
    end for
    for $w = 1, \cdots, W$ do
        Get value $Q^P(s(t), a(t))^w$ from the prediction network.
        Get value $Q^T(s(t+1), \text{argmax}_a Q^P(s(t+1), a))^w$ from the target network based on $s(t+1)$.
    end for
    Use the loss function (12) to update the prediction network.
end for
Update the target network.
```

proposed algorithm can reduce the complexity significantly, compared with conventional methods such as SDR with complexity of $O(K(M+1)^6)$ [9].

IV. Simulation Results

Simulation results of the proposed DRL-based schemes are shown in this section. Unless otherwise stated, we set the parameters for the system as follows: the number of relays $K = 5$, the transmit power to noise ratio $P/\sigma_n^2 = P_S/\sigma_n^2 = P_R/\sigma_n^2 = 35$ dB, the number of IRS elements $M = 16$, the path loss exponent $\hat{\alpha} = 2$, $\bar{\alpha} = 2.5$, the Rician factor $K = 10$ dB for links with Rician fading, the target rate $\vartheta = 0.5$ bps/Hz, the discount coefficients $\delta = 0.9$, the number of time slots for updating the prediction network $T = 500$, the training sample number $W = 32$, and the iteration number of updating the target network $V = 100$. Moreover, the suggested quantization for the IRS $\log_2(L)$ is two bits based on [5], [19]. Thus, we consider the number of phase quantization level $L = 4$ in this paper. The locations of S, I, D, R_1, R_2, R_3, R_4 and R_5 are (0, 0) m, (5, 25) m, (0, 50) m, (1, 25) m, (-4.1, 23.4) m, (4.5, 26.1) m, (3.8, 24.2) m and (1.5, 29.1) m, respectively.

Fig. 2 shows the throughput versus training iterations for the proposed scheme. It is shown that the DRL-RI scheme can achieve approximately 0.1 packets/time slot at the beginning, and converges to about 0.4 packets/time slot after 13,000 training iterations. This result indicates that due to the high-dimensional space of the hybrid relay and IRS networks, the DRL-based scheme needs many iterations to explore the environment and the convergence is not very stable during training. However, finally the DRL algorithm can converge and obtain a solution because it can learn from the exploration experience. Moreover, after training, the proposed scheme can obtain a low complexity DNN for making decisions [9], which can be implemented to reduce the computational complexity in hybrid relay and IRS networks.

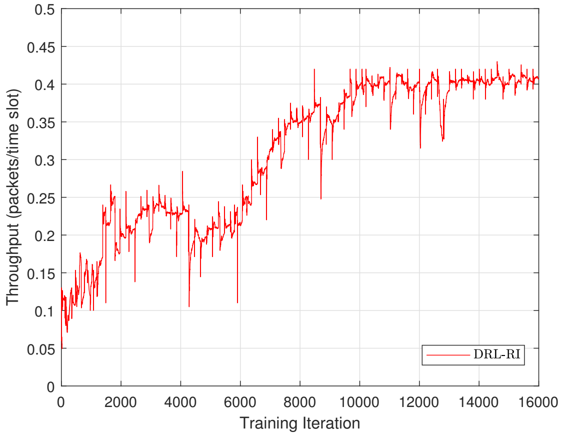

Fig. 2. Throughput vs. training iterations.

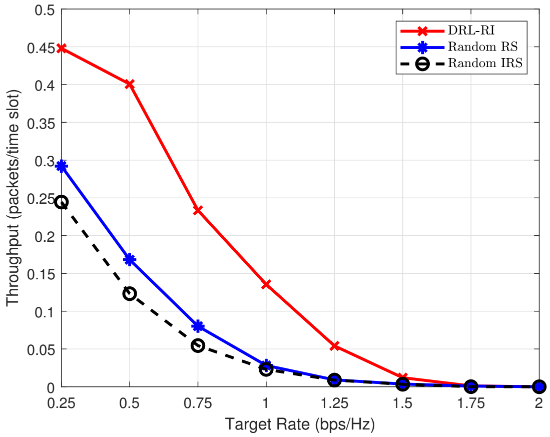

Fig. 3. Throughput vs. target rate.

Fig. 3 shows the comparison of throughput versus different target rates between the proposed scheme, IRS reflection coefficient optimization scheme with random relay selection (Random RS), and relay selection scheme with random IRS reflection coefficient (Random IRS). It is shown that the proposed scheme outperforms the other two schemes significantly. The DRL-RI scheme achieves about 0.4 packets/time slot when the target rate $\vartheta = 0.5$ bps/Hz, while Random RS and Random IRS achieve 0.17 and 0.12 packets/time slot, respectively. The proposed DRL-RI scheme can not only optimize the reflection coefficients for the IRS, but also optimize the relay selection to reduce the outage probability. Thus, the proposed scheme can amalgamate the benefits of relay selection and IRS to achieve a high throughput.

Fig. 4 shows the comparison of throughput versus different transmit power to noise ratios between the proposed scheme, Random RS, and Random IRS. It is shown that the proposed DRL-RI scheme achieves approximately 0.45 packets/time slot when the transmit power to noise ratio $P/\sigma_n^2 = 40$ dB, while

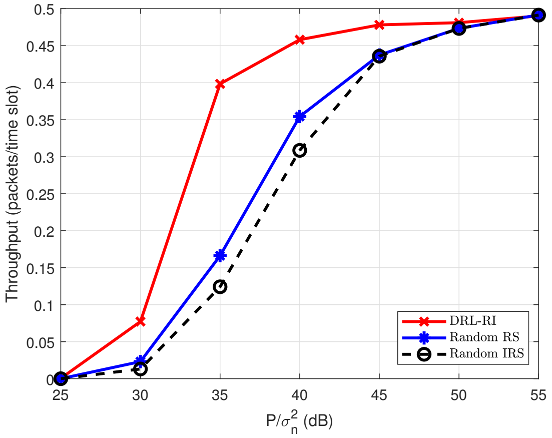

Fig. 4. Throughput vs. P/σ_n^2.

Random RS and Random IRS only achieve 0.35 and 0.31 packets/time slot, respectively. It is clearly shown that the performance of all algorithms get better with the increase of the transmit power to noise ratio. This is because the SNR varies directly proportionally to the transmit power to noise ratio, based on (5) and (7).

V. Conclusion

This paper investigated the throughput maximization problem in cooperative networks with IRS joint relay selection and discrete IRS reflection coefficients optimization. We apply the DRL algorithm to learn from the environment to map the relation between the optimization variables and throughput, solve the non-convex optimization problem in which the IRS reflection amplitudes vary with the discrete phase-shifts. Compared with the random relay selection algorithm and the random IRS reflection coefficient optimization algorithm, the proposed scheme can obtain significant performance gain. This result shows the benefits of joint relay selection and IRS reflection coefficients to reduce the signal loss over distance, and provide a potential way to solve complicated optimization problems in wireless communications with low computational-complexity.

References

[1] M. Di Renzo, M. Debbah, D. Phan-Huy, A. Zappone, M. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin, *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[2] C. Pan, H. Ren, K. Wang, M. Elkashlan, A. Nallanathan, J. Wang, and L. Hanzo, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[3] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Communications Magazine*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[4] C. Pan, H. Ren, K. Wang, W. Xu, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Multicell MIMO communications relying on intelligent reflecting surfaces," *IEEE Transactions on Wireless Communications*, vol. 19, no. 8, pp. 5218–5233, Aug. 2020.

[5] Y. Han, W. Tang, S. Jin, C. Wen, and X. Ma, "Large intelligent surface-assisted wireless communication exploiting statistical CSI," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 8, pp. 8238–8242, Aug. 2019.

[6] Y. Yang, S. Zhang, and R. Zhang, "IRS-enhanced OFDM: Power allocation and passive array optimization," in *IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, PP. 1-6.

[7] Y. Song, M. R. Khandaker, F. Tariq, and K.-K. Wong, "Truly intelligent reflecting surface-aided secure communication using deep learning," *[Online] arXiv preprint arXiv:2004.03056*, 2020.

[8] C. Zhang, P. Patras, and H. Haddadi, "Deep learning in mobile and wireless networking: A survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, ThirdQuarter, 2019.

[9] K. Feng, Q. Wang, X. Li, and C. Wen, "Deep reinforcement learning based intelligent reflecting surface optimization for MISO communication systems," *IEEE Wireless Communications Letters*, vol. 9, no. 5, pp. 745–749, May. 2020.

[10] C. Huang, R. Mo, and C. Yuen, "Reconfigurable intelligent surface assisted multiuser MISO systems exploiting deep reinforcement learning," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 8, pp. 1839–1850, Aug. 2020.

[11] S. Abeywickrama, R. Zhang, Q. Wu, and C. Yuen, "Intelligent reflecting surface: Practical phase shift model and beamforming optimization," *IEEE Transactions on Communications*, vol. 68, no. 9, pp. 5849–5863, Sep. 2020.

[12] Z. Tian, G. Chen, Y. Gong, Z. Chen, and J. A. Chambers, "Buffer-aided max-link relay selection in amplify-and-forward cooperative networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 2, pp. 553–565, 2015.

[13] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "A hybrid relay and intelligent reflecting surface network and its ergodic performance analysis," *IEEE Wireless Communications Letters*, vol. 9, no. 10, pp. 1653–1657, Oct. 2020.

[14] Z. Abdullah, G. Chen, S. Lambotharan, and J. A. Chambers, "Optimization of intelligent reflecting surface assisted full-duplex relay networks," *IEEE Wireless Communications Letters*, (Early Access), 2020.

[15] M. Sami, N. K. Noordin, M. Khabazian, F. Hashim, and S. Subramaniam, "A survey and taxonomy on medium access control strategies for cooperative communication in wireless networks: Research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2493–2521, Fourthquarter, 2016.

[16] Y. Su, X. Lu, Y. Zhao, L. Huang, and X. Du, "Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks," *IEEE Sensors Journal*, vol. 19, no. 20, pp. 9561–9569, Oct. 2019.

[17] K. Kobayashi, T. Ohtsuki, and T. Kaneko, "Precoding for MIMO systems in line-of-sight (los) environment," in *IEEE Global Telecommunications Conference(GLOBECOM)*, Washington, DC, USA, Nov. 2007, pp. 4370-4374.

[18] S. Fang, G. Chen, and Y. Li, "Joint optimization for secure intelligent reflecting surface assisted UAV networks," *IEEE Wireless Communications Letters*, (Early Access), 2020.

[19] P. Xu, G. Chen, Z. Yang, and M. Di Renzo, "Reconfigurable intelligent surfaces assisted communications with discrete phase shifts: How many quantization levels are required to achieve full diversity?," *IEEE Wireless Communications Letters*, (Early Access), 2020.

[20] 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, Jan. 2021.

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