Neural Episodic Control-based Adaptive Modulation and Coding Scheme for Inter-satellite Communication Link

Donggu Lee1, Young Ghyu Sun1, Isaac Sim1, Student Member, IEEE, Jae-Hyun Kim2, Member, IEEE, Yoan Shin3, Senior Member, IEEE, Dong In Kim4, Fellow, IEEE, and Jin Young Kim1, Senior Member, IEEE

1Department of Electronic Convergence Engineering, Kwangwoon University, Seoul 01897, South Korea
2Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea
3School of Electronic Engineering, Soongsil University, Seoul 06978, South Korea
4Department of Electrical and Computer Engineering, Sungkyunkwan University, Suwon 16419, South Korea

Corresponding author: Jin Young Kim (e-mail: jinyoung@kw.ac.kr).

This work was supported in part by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2021-0-00892), Research on advanced physical layer technologies of LEO (low-earth orbit) satellite communication systems for ubiquitous intelligence in space and in part by the MSIT, Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-2018-0-01424) supervised by the IITP and in part by the Kwangwoon University in 2021.

ABSTRACT Inter-satellite links (ISLs) play an important role in the global navigation satellite system (GNSS), which is known as one of the key technologies for the next generation of navigation satellite systems. Deep reinforcement learning algorithms have achieved significant improvement over various wireless communications systems. However, it has been reported that deep Q network (DQN) algorithm requires an enormous number of trials. To resolve this problem, in this paper we propose an adaptive modulation and coding scheme based on a neural episodic control (NEC) algorithm, which is one of deep reinforcement learning algorithms. The proposed scheme adjusts the modulation and coding scheme region boundaries with a differentiable neural dictionary of the NEC agent, which enables the effective integration of the previous experience. In addition, we propose a step-size varying algorithm to encourage the NEC agent to efficiently approach the suboptimal state. We confirm that the proposed scheme can reduce the number of trials to 1/8, compared to the DQN-based adaptive modulation scheme. To further evaluate the proposed scheme, we employ an online learning loss evaluation algorithm that calculates the loss in time-step based on interaction records of the reinforcement learning agent and the derived modulation and coding scheme region boundaries.

INDEX TERMS Adaptive modulation and coding, deep learning, inter-satellite communications, neural episodic control, reinforcement learning.

I. INTRODUCTION

Inter-satellite links (ISLs) play a significant role in the global navigation satellite system (GNSS), which enables autonomous operating inter-satellite systems and accurate orbit determination [1]. ISL technologies are widely known as one of the key technologies for the future generation of navigation satellite systems [2]. Also, there is an increasing interest in small-size satellites and their cooperative mission for remote sensing networks [3]. With the remarkable advances in commercialized reusable launch vehicle technology of SpaceX, and OneWeb project [4], [5], it is expected that the demands of the ISLs will be increasing, and it is also expected that the ISLs form the dense and massive networks. From these trends, it is important to improve communication performance for inter-satellite communication links.

Adaptive modulation and coding (AMC) schemes have been an effective method to improve spectral efficiency of a wireless communication link within a given quality [6]. The promising feature of the AMC scheme is a real-time balancing variation over symbol transmission rate, transmit power level, coding scheme or rate, constellation size, etc. [7]-[9]. The AMC technologies can be a strong candidate to improve the spectral efficiency of the satellite communications systems with some distinctive limitations, such as limited power source, propagation characteristics, on-board power constraints, etc. [10]. Moreover, the maximum power loss more than 20dB may occur. It is caused by the distance between the satellites, which varies according to the constellation of the satellites [11]. Therefore, the AMC scheme can be a way to efficiently utilize spectrum resources within varied channel conditions.

In [11], an AMC scheme for the inter-satellite communication system is proposed. Here, the AMC scheme is formulated with the distance between the satellites. The proposed scheme has two strategies which are target bit error rate (BER) algorithm and maximum throughput algorithm. In the target BER algorithm, the modulation and coding scheme (MCS) is determined according to the point that achieves the given target BER. Contrary to the target BER algorithm, the
maximum throughput algorithm chooses the MCS region boundaries following the envelopes of the maximum throughput. Those MCS region boundaries in the form of signal-to-noise ratio (SNR) are recalculated into the distance between satellites. In addition, a continuous phase modulation-based AMC scheme is proposed in [12] with the framework of [11]. It has been reported that the proposed scheme in [12] can improve the spectral efficiency for the inter-satellite communication link subject to the same transmit power level.

A. BACKGROUND AND RELATED WORKS

Reinforcement learning algorithm, which is one of the machine learning algorithms, is based on interactions between environment and a learning model [13], which is called an agent. Reinforcement learning algorithms have been effectively employed in various decision-making areas of communication systems [14]-[20]. In [14], a fuzzy Q-learning-based MCS selection and MIMO configuration scheme for high-speed access evolution systems is proposed. Here, the proposed scheme observes channel quality indication and block error rate between a base station and user equipment. By combining the fuzzy logics and Q-learning algorithm, an appropriate MCS and MIMO configuration are selected via a closed-loop iteration of the reinforcement learning agent. It is confirmed that the proposed scheme outperforms the conventional adaptive threshold selection scheme and Q-learning-based hybrid automatic repeat request scheme [15] under the given block error rate requirement.

However, due to the enormous dimension of state-action space in a practical environment, deep reinforcement learning algorithms have been proposed, bringing in significant advances in the decision-making process [16], [17]. The deep reinforcement learning algorithms employ the deep neural network and reinforcement framework to improve the automotive decision-making process in an online manner. Comparing with conventional heuristic approaches, the deep reinforcement learning algorithms easily adapted to environmental changes by learning from new experiences of the agent [18]. Moreover, it has been known as the reinforcement learning-based approaches can provide optimal decision at each time step without any prior knowledge of the environments [13].

In [19], a deep Q network (DQN) algorithm-based adaptive modulation scheme with outdated channel state information is proposed. To overcome a limitation of outdated channel state information that changes non-linearly, the DQN agent selects a proper modulation scheme from received signal strength. Here, it has been reported that the proposed DQN-based adaptive modulation scheme outperforms conventional outdated channel state information measurement schemes. Moreover, in [20], a DQN algorithm-based adaptive modulation scheme with a trial efficient strategy is proposed. The proposed scheme divides the SNR range into 4 different regions with different bits per symbol rate, which achieves a reduction in computational complexity compared to the scheme in [21]. It has been known that the DQN algorithm has the following advantages [22].

Firstly, deep neural networks are employed in the DQN algorithm to encourage the agent to efficiently explore among its action candidates by setting the output of the networks as the action options, which can be viewed as the concept of virtual exploration [23], [24]. Secondly, the correlation between input data is removed to stabilize the learning stage with a replay memory scheme [22], [25]. Lastly, the main and target networks are separated to stabilize the update process of the neural network. With those advantageous properties, the effect of the dilemma between computational complexity and throughput loss can be relieved, which is incurred by the quantized states in [21].

It has been known that the DQN algorithm has various improved forms with advantageous properties. The double DQN (DDQN) [26] and dueling DDQN algorithms [27] are proposed to improve the DQN algorithm. The DDQN algorithm attempts to mitigate the effect of overestimation in the DQN algorithm. Here, over-optimistic selection and evaluation on actions can be effectively reduced by decomposing the max operation of the target value of the DQN algorithm into the action selection process. Meanwhile, the dueling DDQN algorithm adopts a distinctive neural network structure, termed dueling architecture [27]. In the dueling architecture, the last hidden layer is split into two layers to generate the state value and the action advantage value, which is added into the Q-values [27, Figure. 1]. It is reported that the agent can learn which states are more valuable to learn, without observing the effects of each action at each state, via the dueling structured network.

B. MAIN CONTRIBUTIONS

It has been reported that the DQN algorithm requires a large number of trials and training processes of the reinforcement learning agent [28]. Motivated by this fact, in this paper we propose a neural episodic control (NEC) algorithm-based AMC scheme for ISLs. The NEC algorithm, which is one of the deep reinforcement learning algorithms, provides the advantageous properties of faster integration of the previous experience compared to the DQN algorithm. By employing a distinctive memory structure of the NEC algorithm, it is able to make optimal decisions among the action candidates. Further, it has been reported that the adaptive modulation scheme in [20] can be viewed as a gradient-based approach. From this sense, a variable step-size algorithm [29] can be applied for our scheme to minimize the number of trials of the agent and the convergence time of the algorithm.

The main contributions of this paper are listed in the followings.

- The NEC-based AMC scheme has been proposed. By employing the NEC algorithm, the number of trials of the agent to the suboptimal state can be significantly reduced, while the NEC algorithm requires less volume of memory than the DQN algorithm.
- To minimize the required number of trials to the suboptimal state of the agent, a step-size varying algorithm has been employed to encourage the agent to reach the suboptimal point faster, in conjunction with the NEC-based AMC scheme.
- We employ an online learning loss evaluation algorithm to calculate spectral utilization loss based on interaction
records of the reinforcement learning agent and the derived MCS region boundaries.

The remainder of this paper is organized as follows. Section II presents the system models of inter-satellite communication and reinforcement learning. In Section III, the NEC-based AMC scheme and the online learning loss evaluation algorithm are described. Simulation settings and results of the proposed scheme are shown in Section IV. Lastly, concluding remarks are given in Section V.

II. SYSTEM MODEL
A. INTER-SATELLITE COMMUNICATION MODEL

It has been reported that the loss for the satellite-to-satellite communication links contains various loss factors such as directivity of antenna, polarization loss of antenna, etc. [11]. However, free-space loss is the primary factor of the ISL, which is dominantly dependent on the distance between satellites [12]. SNR per bit of the ISL at the receiver can be obtained as [11]

\[ \frac{E_b}{N_0} = EIRP + \frac{G}{T} + T_r - 20 \log \left( \frac{4\pi d}{\lambda} \right) - N_0 - D - L_0, \]

where \( E_b \) denotes the energy per bit, \( N_0 \) is the one-sided power spectral density, \( EIRP \) is the effective isotropic radiated power, \( G/T \) is the factor of quality of receiving antenna, \( T_r \) represents the noise of the system, \( d \) is the distance between satellites, \( \lambda \) is the wavelength of the carrier, \( D \) denotes the data rate, and \( L_0 \) is the system margin, respectively.

From the perspective of the constellation of the satellites, Standard Walker 24/3/1 constellation [11] has been considered in this paper. It has consisted of 24 medium earth orbit (MEO) satellites, 3 geostationary orbit (GEO) satellites, and 3 inclined GEO (IGEO). The maximum and minimum distances between the satellites on the constellation have been provided in [11, Table 1]. The GNSS satellites have the information of distance between other satellites by using ephemeris. Thus, a control channel between transmitter and receiver is not required for the AMC schemes of inter-satellite communication systems. The architecture of the adaptive communication model for the inter-satellite communication link is shown in Figure 1 [11], [12]. As shown in the Figure 1, the transmitter selects a proper MCS by calculating the distance between the transmitter and the receiver. Then, the receiver demodulates and decodes the transmitted signals via pilot signal and the distance information.

B. REINFORCEMENT LEARNING MODEL

Reinforcement learning algorithm aims to make the agent learn a task with the observation of environments and its interaction [13], which are called a state and an action, respectively. The agent evaluates its action by a reward, which is given by the environments and the primary goal of the agent is to maximize the cumulative reward. In this paper, the reinforcement learning agent has the following components:

- \( S = \{s_1, s_2, s_3\} \) represents the state values, which are the rate region boundaries of the SNR range [20]. In this paper the 4 different modulation schemes are allocated with the 3 state values. Moreover, convolutional coding scheme [30] with the rate of 1/2, 2/3, 3/4, and 5/6 has been employed in each rate region, respectively.

- \( A = \{a_1, \ldots, a_6\} \) denotes the candidates of the action of the agent. Each pair of the action increases or decreases the corresponding state value [20]. Considering an example of a pair of \( a_3 \) and \( a_5 \), decreasing the value of \( s_3 \) by a step size and \( a_6 \) increases the \( s_3 \) by the same amount.

- \( R = \{r_i\} \) is reward function, where \( r_i \) denotes the reward unit earned by the agent, which is assumed to be [21]

\[ r_i = \sum_{i=1}^{M} r_i C_i (1 - \overline{BER}_i), \]

where \( r_i \) denotes the modulation order of the \( i \)-th MCS region, \( C_i \) is the code rate of \( i \)-th MCS region, and \( \overline{BER}_i \) is the average BER of each MCS region, and \( M \) is the number of MCS regions, respectively. The reward unit is set to represent the average number of bits that successfully transmitted for each MCS region.

C. NEURAL EPISODIC CONTROL MODEL

Deep reinforcement learning algorithms have shown significant performance improvements over the various environments with neural networks. However, it has been reported that the DQN-based approaches take an enormous number of trials of the agent [28], and the NEC algorithm has been proposed to resolve this problem. The NEC algorithm has a distinctive memory unit called differentiable neural dictionary (DND), which can be represented as

\[ M_a = (K_i, V_i), \text{ for all } a \in A, \]

where \( K_i \) and \( V_i \) are the \( i \)-th elements of the key and the corresponding value, respectively. The NEC agent will have a DND for each action that the agent can take. The DNDs are utilized for two operations of the agent: 1) write and 2) lookup [28].

In the operation of writing, a new key-value pair will be added to the DND of the corresponding action that the agent takes. The key vector can be obtained from embedding, which is the output of the neural network. The corresponding value \( V_i \) can
The rate of the proposed scheme has not been $M = Q$. 

\[ \alpha \{ Q(s_{t+1}, a_{t+1}) \} \]

where $\delta$ denotes the discounting factor, $N$ is the number of trials, and the subscript of the $r$ is time step, respectively. The value of $Q_N(s_t, a_t)$ is stored as the corresponding value $V_t$. However, if the embedded key vector already exists in the DND, the corresponding value is updated with the classic Q-learning value as [28]

\[ Q_t \leftarrow Q_t + \alpha \{ Q(s_t, a_t) - Q_t \}, \]

where $\alpha$ is the learning rate of the Q-learning algorithm. The agent will append or update the key-value pairs through its interaction to the corresponding DND.

The agent will estimate the best action choice with the process of the lookup. After the embedding is made, the agent compares the Q-function value of each action DND, which can be obtained as [28]

\[ Q(s, a) = \sum_k k(e, e_i) Q_i, \]

where $k(x, y)$ denotes the inverse kernel function [28, Eq. 5], $e_i$ is the $i$-th element of $K_e$, and $e$ is the embedding, respectively. The output of each DND can be viewed as a form of weighted sum of values, while proper $e_i$ vectors can be selected as the most similar sets of $e$. In this paper, $k$-nearest neighbor algorithm has been considered to pick up the proper $e_i$ vectors.

In Figure 2, an architecture of DND of each action is shown. The NEC agent selects the optimal action at each time step via the lookup operation of each action DND. To encourage the agent to explore various action space, $\varepsilon$-greedy algorithm-based exploration and exploitation strategy has been considered.

\[ P_b(E) \approx \frac{2}{H} \sum_{i=1}^{\max(4/M,1)} Q_i \left( \sqrt{\frac{E_b \log_2 M}{N_0} \sin \left( \frac{(2i-1)\pi}{M} \right) } \right), \]

where $H = \max(\log_2 M,2)$, and $Q_i$ denotes the Gaussian $Q$ function [31, Eq. 4.2]. The BER approximation for M-QAM can be expressed as [31, Eq. 8.15]

\[ P_b(E) \approx 4 \left( \frac{\sqrt{M} - 1}{\sqrt{M \log_2 M}} \right)^{\frac{\Pi/2 - 1}{1}} \sum_{i=3}^{\infty} Q_i \left( 2i + 1 \right) \frac{3E_b \log_2 M}{N_0(M - 1)}, \]

We also consider the convolutional coding scheme for the proposed AMC scheme. The coding gain of the convolutional coding scheme can be obtained as [32, Eq. 7.23]

\[ G_c \leq 10 \log_{10} (C d_f), \]

where $C$ denotes the code rate and $d_f$ represents the free distance of the coding scheme, respectively.

The spectral utilization rate of each modulation and coding scheme can be represented as [12]

\[ SU_i = R_i C_i \text{ [bps/Hz]} \]

In this paper, weighted average of the spectral utilization rate is considered to reflect the range of each MCS region, which can be derived as

\[ SU_w = \sum_{i=0}^{N} \frac{SU_i L_i}{L_{\text{total}}} \]

where $L_{\text{total}}$ denotes the range of SNR region in dB scale and it is assumed to be 25dB in this paper, and $L_i$ is the range of the $i$-th MCS region. However, it has been observed that the overall spectral utilization rate of the proposed scheme has not been significantly varied because the term of the range of each MCS region is dominant in (11). To compensate this, the overall SNR range has been normalized into the range of [0, 1]. The spectral utilization rate with the normalized range can be obtained as
where \(L_{i,n}\) denotes the normalized range of each MCS region.

### B. NEC-BASED AMC SCHEME

The proposed NEC-based AMC algorithm is shown in Algorithm 1. It is assumed that the DNDs for all action are initialized by the random actions of the agent before the interaction of the agent starts. Then, the agent will get an embedding vector from \(S\), and observe the spectral utilization rate and the BER of each region before and after taking an action to calculate the reward. The agent will get a penalty when at least one of the maximum BER of each region exceeds the target BER, which is assumed to be \(10^{-3}\) in this paper. The agent will load the previous optimal state at the beginning of the episode and when the BER of each region is violated, to encourage the agent to efficiently converge to the suboptimal state [20]. Then, if the spectral utilization rate is increased, the agent will get the reward via the reward function. The agent will also get the penalty if the \(S\) is invalid. The invalid conditions are 1) when \(S\) exceeds the bound of simulation range, and 2) each state value gets closer within a margin of 3dB. The penalty will cause decrease in the cumulative reward, and it is expected that the agent will follow the valid condition. After the reward is given to the agent, the key-value pair is written to the corresponding DND, which can be expressed by \((e, Q_{\theta}(s_t, a_t))\). Replay memory has been employed in the NEC algorithm, like the DQN algorithm [20]. The information of transition is saved in the replay memory to use in the training process of the neural network. It has been assumed to minimize the \(L2\) loss between the predicted output value of an action and the \(Q\)-value from mini-batches that randomly sampled in the replay memory [28], which can be obtained as

\[
L(\theta) = \left\{ Q_{\theta}(s_t, a_t) - Q_{\theta}(s_{\text{opt}}, a_{\text{opt}}) \right\}^2,
\]

where \(\theta\) denotes the trainable variables in the neural network, and the terms with the subscript \(M_R\) represent the data sampled from the replay memory. Back propagation-based update on \(\theta\) of the neural network has been considered, which can be viewed as the neural network is trained to make effective representation of embeddings to reflect context and feature of environments.

The proposed scheme can be viewed as a gradient-based approach, which is based on the steps to the suboptimal state of the agent. Inspired by this fact, the step-size varying algorithm can be applied. The step-size of the algorithm is set to be \(1/(N_{\text{ep}} + 1)\) in dB scale, which is inversely proportional to the number of episodes. The required number of trials to reach the suboptimal state of the proposed algorithm can be described as

\[
SU_{w,n} = \sum_{i=0}^{N} SU_{i,n},
\]

where \(SU_{i,n}\) denotes the normalized range of each MCS region.

#### Algorithm 1 NEC-based AMC algorithm

1: Initialize replay memory \(M_R\) to capacity \(P\)
2: Initialize \(M_A\) for all \(A\)
3: Initialize state-value buffer \(B\) to capacity \(T\)
4: Initialize environment of \(S, A, R,\) and \(r_t\)
5: For episode = 0, \(T\) do
6: if episode > 0 then Load the arg max \(S\) from \(B\)
7: end if
8: For step = 0, \(J\) do
9: Get an embedding vector \(e\) from \(S\)
10: Evaluate spectral utilization rate \(SU_{w,n}\) via (12)
11: Evaluate BER of each rate region via (7)-(9)
12: Take action \(a_t\) with \(\varepsilon\)-greedy algorithm and update \(S\)
13: Evaluate spectral utilization rate \(SU'_{w,n}\) via (12)
14: Evaluate BER of each rate region via (7)-(9)
15: Calculate \(r_t\) via (2)
16: if max(BER) > \(10^{-3}\) then
17: \(r_t \leftarrow r_t - 1.5r_u\)
18: \(\text{Load the arg max } S\) from \(B\)
19: end if
20: if \(SU'_{w,n} > SU_{w,n}\) then \(r_t \leftarrow r_t + r_u\) end if
21: if \(S\) is invalid then \(r_t \leftarrow r_t - 1.5r_u\) end if
22: Write \((e, Q_{\theta}(s_t, a_t))\) to \(M_R\)
23: Store transition \((s_t, a_t, Q_{\theta}(s_t, a_t))\) in \(M_R\)
24: if step = \(J\) then \(\text{Store arg max } S\) in \(B\) end if
25: Sample random mini-batches from \(M_R\)
26: Train network with mini-batches via (13)
27: end for
28: end for

\[
\frac{S_{\text{subopt}} - S_{\text{mit}}}{\Delta} \geq N_{\text{req}}.
\]

where \(S_{\text{subopt}}\) denotes the suboptimal state, \(S_{\text{mit}}\) refers to the initial state of the agent, \(\Delta\) is the step-size per action, and \(N_{\text{req}}\) is the required number of trials. Because \(S_{\text{subopt}}\) and \(S_{\text{mit}}\) are assumed to be same in the step-size varying algorithm and the fixed step-size case, \(N_{\text{req}}\) is only dependent on the step-size. It can be interpreted that efficient convergence to the suboptimal state can be achieved with the step-size varying algorithm because of its advantage of faster approach in the range of lower number of episodes and precise tuning in the range of higher number of episodes.

The error between the optimal \(Q\)-function value and the \(Q\)-function value of the \(i\)-th iteration can be obtained as

\[
E_i = Q(s, a; \theta_i) - Q'(s, a) = E_{TAE} + E_{OE} + E_{OD},
\]

where \(Q'(\cdot)\) denotes the optimal \(Q\)-function value, \(E_{TAE}\) is the target approximation error, \(E_{OE}\) refers to the overestimation error, and \(E_{OD}\) denotes the optimality difference. The three components of the overall error can be described as

\[
E_{TAE} = Q(s, a; \theta_i) - y^d(s, a),
\]

\[
E_{OE} = y^d(s, a) - \hat{y}(s, a),
\]

\[
E_{OD} = \hat{y}(s, a) - Q'(s, a).
\]
\[ E_{Q \theta} = \hat{y}^i(s, a) - Q^*(s, a), \]  

(18)

where \( y^i(s, a) \) is the target value of the network and \( \hat{y}^i(s, a) \) denotes true target value, which can be expressed as [33]

\[ y^i(s, a) = \mathbb{E} \left[ \eta_i + \delta \max_{a'} Q(s', a'; \theta_{t-1}) \mid s, a \right], \]  

(19)

\[ \hat{y}^i(s, a) = \mathbb{E} \left[ \eta_i + \delta \max_{a'} y^i(s', a') \mid s, a \right]. \]  

(20)

Because the \( N \)-step \( Q \)-function can be viewed as the target value of the DQN algorithm [28], it can be expected that the target value of the NEC network can be derived by substituting (4) into (19).

The overestimation error represents a bias of the overall error term, which has its upper bound that is proportional to the number of actions in a state [34]. It can be assumed that the overestimation error is not dominant because the selectable actions in all states are assumed to be same in the NEC and DQN algorithms. Besides, the optimality difference represents difference between the optimal \( Q \)-function value and the true target value. Because the state and the action space of the true target value is under its boundaries of the environment, it can be expected that the optimality difference is not a dominant term.

On the other hand, the target estimation error, which can be a dominant within the error terms above, is comprised of the following components [33]: 1) suboptimality of the network trainable parameters because of inexact minimization of optimizer, 2) modeling error of neural network, and 3) generalization error of unexperienced state-action pairs. The first and second factors of the target estimation error can be assumed to be equal in both cases of the NEC and DQN algorithms because they are assumed to employ the same neural network and optimizer. Thus, it can be interpreted that the NEC algorithm can effectively reduce the error between the optimal \( Q \)-function value and its \( Q \)-function value with the integration of previous experiences by employing the DNDs and the \( N \)-step \( Q \)-function.

In addition, it can be expected that the NEC algorithm has the following benefits compared to the DQN algorithm. Firstly, the NEC algorithm can integrate the previous experience, which is the record of adjusting MCS region boundaries, with a form of a dictionary. By using the architecture of DND of each action candidate, it allows the agent to take optimal action much faster. Secondly, the \( N \)-step \( Q \)-function in (4) that can be viewed as the sum of on-policy and off-policy back-ups, and it can act as the target network of the DQN algorithm [28], which stabilizes the training processes of the DQN agent by separating from the main network. Further, due to this advantageous property, it has been reported that the replay memory for the NEC agent can be reduced compared to the DQN algorithm. Lastly, the slowly varying embedding vectors produced by the neural network can stabilize the DND. Contrary to the DND, some of the memory architecture-based approaches empty out its memory at the beginning of each episode [35], and it can be expected that the memory is unstable over the episodes.

C. ONLINE LEARNING

The online learning loss evaluation algorithm, which can evaluate real-time loss of the AMC model, is shown in Algorithm 2. In this algorithm, the MCS region boundaries and the interaction records of the NEC and DQN models will be utilized. Firstly, the interaction records of the AMC model are loaded to reflect exact interactions of the reinforcement learning model. Then, the SNR value, which is denoted as \( \gamma \), and the probability of \( p_{sam} \) are randomly sampled. \( p_t \) denotes the threshold probability of current link being maintained. If \( p_t > p_{sam} \), the SNR between satellites is maintained in a range of 1dB. The spectral utilization rate loss due to misplaced MCS is occurred in the case of \( s_t^* < SNR < s_t \) for all MCS regions, where \( s_t^* \) refers to the optimal state. For example, if the SNR value falls into the region of \( s_t^* < SNR < s_t \), the 16-QAM with code rate of 2/3 should be employed. However, the QPSK with code rate of 1/2 will be assigned in the area until the agent reaches the suboptimal state. The spectral utilization rate loss is evaluated with \( L_I \) loss, and it can be obtained as

\[ L_o = |SU^*_t - SU_{I,t}|. \]  

(21)

where \( SU^*_t \) is the spectral utilization rate with the optimal MCS region boundaries and \( SU_{I,t} \) denotes the spectral utilization rate with the MCS region boundaries at time-step \( t \). The time-step \( t \) is set to be unit time step in the Algorithm 2, where unit time-step is assumed to include the time of a single adjusting step of the agent, generation of random SNR and link maintain probability, and the loss evaluation process.

IV. SIMULATION RESULTS

A. NEC-BASED AMC SCHEME

The simulation parameter settings are listed in Table 1, and further assumptions on the simulation have been made as follows: 1) a 4-layered fully connected network has been employed [20], 2) the dimensions of the input and output of neural network are set to be 3 for NEC algorithm, which are the dimension of \( S \) and embedding vector, 3) the output dimension
is set to be 6 for the DQN algorithm, 4) RMSprop optimizer [36] has been considered for the training process with the learning rate of 0.4, 5) DND of each action has the least recently used (LRU) property [28], 6) the k-nearest algorithm of the lookup operation has been approximated by using kd-trees [37], and 7) the initial state is set to be the lowest point of the spectral utilization rate with the equally divided the simulation range, except the highest MCS region [20].

In Figure 3, the simulation results on the maximum spectral utilization rate, which indicates the maximum spectral utilization rate achieved by the agent within one episode, are compared among the AMC schemes with the four algorithms: the NEC, dueling DDQN, DDQN, and DQN algorithms. It is assumed that the maximum volume of replay memory and the mini-batch size of the dueling DDQN and DDQN algorithms are set to be the same as DQN algorithm. As seen in Figure 3, the proposed scheme can approach the suboptimal state with the smallest number of episodes among the four algorithms, and the average difference in the maximum spectral utilization rate over the episodes between the NEC-based scheme and the DQN-based scheme is 0.4879 bps/Hz. The proposed scheme reaches the suboptimal state within 24 episodes, while the other algorithm-based schemes require the number of episodes to reach the same point by 48 episodes for dueling DDQN, 68 episodes for DDQN, and 74 episodes for DQN, respectively. It has been observed that the time needed to achieve the suboptimal state can be reduced to a half compared with the DQN-based scheme, while using less memory. In addition, the steep improvement of the proposed scheme can be viewed as due to the optimal decision on actions with the lookup process of the DNDs. Due to the trade-off relationship between the spectral utilization rate and the BER of each MCS region, all algorithms approach the suboptimal point by adjusting MCS region boundaries while sacrificing the BER. Thus, the four different base algorithm-based AMC schemes in Figure 3 achieve the same performance under the BER constraint after the sufficient number of adjusts, which is significantly varied with the base algorithms.

In addition to the simulation results in Figure 3, we also investigate a dynamic programming approach. In the proposed scheme, the NEC agent controls the three state values to approach the suboptimal state. From this sense, the formulated problem of the proposed scheme can be viewed as finding shortest path over three dimensional points. Dijkstra’s shortest path algorithm [39] is employed to find the shortest path from the initial point to the suboptimal point, while each point is set to be \((s_1, s_2, s_3)\). In Table 2, the simulation parameters and the evaluation result are given. It is assumed to have the weight of the path only for the path between adjacent nodes with a single step. Due to the enormous volume of the search space with \(\Delta=0.05\)dB, it is set to be \(\Delta=0.5\)dB. It requires 9,992 node explorations for deriving the shortest path. If a node exploration process is considered as an action of the reinforcement learning agent, it takes 99 episodes to approach the suboptimal state, which is significantly slower than the four different reinforcement learning-based approaches in Figure 3. It is also expected that it takes more episodes to approach the final point with the same step-size of the reinforcement learning models.

The simulation results for various \(k\) values are shown in Figure 4. The \(k\) value is the parameter of the \(k\)-nearest neighbor algorithm for the lookup operation of the proposed NEC scheme. For instance, if the \(k\) value is set to be 10, the agent will calculate the \(Q\)-function value based on the 10 elements of the DND for each action candidate. In the simulation results of

| Parameter | Value | Description |
|-----------|-------|-------------|
| \(M_k\)  | 1000 for NEC | Maximum volume of replay memory |
|          | 3000 for DQN |
| \(M_n\)  | 200 for all \(a \in A\) | Maximum volume of each DND |
| \(T\)    | 100 | Maximum number of trials per episode |
| \(J\)    | 100 | Maximum number of episodes |
| \(\epsilon\) | \(1/(N_{ep} + 1)\) | The parameter of \(\epsilon\)-greedy algorithm |
| \(\delta\) | 0.9 | Discounting factor |
| \(V_{\text{mini-batch}}\) | 20 for NEC, 40 for DQN | Volume of mini-batch |
| \(L_{\text{total}}\) | \([0.25\text{ dB}]\) | SNR range of simulation |
| \(\Delta\) | 0.05dB | Step-size per action |
| \(d_f\) | 10, 6, 5, 4 | Free distance of each region, respectively |
| \(\alpha\) | 0.1 | Learning rate in (5) |
| \(EIRP\) | 39dBW | 
| \(G/T\) | 2dB/K | 
| \(L_q\) | 5dB | 
| \(T_r\) | 513K | 
| \(k\) | 30 | The parameter of \(k\)-nearest neighbor algorithm of DNDs |

| Parameter | Value | Description |
|-----------|-------|-------------|
| \(\Delta\) | 0.5dB | Step-size |
| \(N_{\text{node}}\) | 2,400 | The number of nodes |
| \(N_{\text{path}}\) | 146,056 | The number of paths |
| \(W_{\text{path},i,p}\) | \(1/\left[SU_{\text{max}}(i) - SU_{\text{max}}(p)\right]\) | Weight of the path from node \(i\) to \(p\) |
| The number of node exploration | 9,992 |
NEC models, all cases of the NEC models yield rapid improvement on the spectral utilization rate in the range of lower number of episodes. It has been observed that as the \( k \) value increases, the required episode to reach the suboptimal state is decreased. However, the significant performance degradation on convergence has been occurred with the model of \( k = 40 \). It can be interpreted that if the parameter of \( k \) is set to be unnecessarily high, it makes the NEC agent to look up irrelevant embeddings in the DNDs, which can be a primary factor of degradation.

In Figure 5, the simulation results are compared between fixed step-size approaches and the step-size varying algorithm. It is clearly seen that the step-size varying algorithm yields the fastest convergence to the suboptimal state. The required number of episodes to reach the suboptimal state for the case of \( \Delta = 0.025 \) dB has been observed to be 47. Contrary to this case, the step-size varying algorithm has reached the suboptimal state with 9 episodes. Besides, it is observed that the NEC model with the step-size varying algorithm requires less than 1/8 of the DQN model in Figure 3. It can be interpreted that the proposed step-size varying algorithm can efficiently approach the suboptimal state, and it can effectively reduce the number of trials of the agent. Moreover, it can be expected that the step-size varying algorithm can relieve the effect of a dilemma on selecting a proper step-size, which is slow convergence with small step-size and leaving room for improvement with unfilled target BER with the large step-size.

In Table 3, a comparison of the MCS boundaries provided by the target BER algorithm in [11, Figure 3] and the proposed scheme is shown. Due to the target BER in [11] being set to be \( 10^{-4} \), approximations on the MCS region boundaries by their simulation results are assumed to compensate the difference. It is confirmed that the proposed scheme yields improvement in the spectral utilization rate by 0.7916 bps/Hz, and it can be interpreted that the proposed scheme more properly allocates the MCS and adjusts the MCS region boundaries than the scheme in [11]. Moreover, the region boundaries in distance of the proposed AMC scheme are derived in Table 4, where \( d \) denotes the distance between the satellites. The distances have been calculated via (1). The carrier frequency is assumed to be 30GHz [11].

From the simulation results, it was confirmed that the NEC algorithm can achieve faster convergence to the suboptimal state than the DQN-based algorithms. It can be interpreted that the improvement on convergence time is due to the following advantages of the NEC algorithm: 1) efficient integration of the previous experience of the agent through DND structure, 2) stabilizing on training process by employing \( N \)-step Q-function, and 3) stable DNDs with slow varying embedding vectors.

TABLE 3. Comparison of MCS region boundaries.

| Region Number [11] | Modulation | Code Rate | \( SU \) | Region [dB] |
|--------------------|------------|-----------|-----------|-------------|
| 1                  | QPSK       | 1/2       | 1         | \( \gamma \leq 7.6 \) |
| 2                  | 16QAM      | 1/2       | 2         | \( 7.6 < \gamma \leq 9.2 \) |
| 3                  | 16QAM      | 3/4       | 3         | \( 9.2 < \gamma \leq 12.4 \) |
| 4                  | 64QAM      | 2/3       | 4         | \( 12.4 < \gamma \leq 14.2 \) |
| 5                  | 64QAM      | 5/6       | 5         | \( 14.2 < \gamma \leq 17.9 \) |
| 6                  | 256QAM     | 3/4       | 6         | \( 17.9 < \gamma \leq 25 \) |

| Proposed          | Modulation | Code Rate | \( SU \) | Region [dB] |
|-------------------|------------|-----------|-----------|-------------|
| 1                  | QPSK       | 1/2       | 1         | \( \gamma \leq 5.08 \) |
| 2                  | 16QAM      | 2/3       | 8/3       | \( 5.08 < \gamma \leq 9.63 \) |
| 3                  | 64QAM      | 3/4       | 9/2       | \( 9.63 < \gamma \leq 14.8 \) |
| 4                  | 256QAM     | 5/6       | 20/3      | \( 14.8 < \gamma \leq 25 \) |

TABLE 4. Region Boundaries of the proposed scheme in Distance.

| Region Number | Region [km] |
|---------------|-------------|
| 1             | \( d > 38343.57 \) |
| 2             | \( 38343.57 < d \leq 17995.94 \) |
| 3             | \( 17995.94 < d \leq 10239.44 \) |
| 4             | \( d < 10239.44 \) |
Moreover, it was shown that the NEC algorithm requires 2,502 unit times, respectively. It can be confirmed that the NEC with the step-size varying algorithm can reduce the cumulative loss to about 1/4 compared to the DQN case [20]. From the simulation results, it has been observed that the maximum difference between the cases of \( p_l = 0.1 \) and \( p_l = 0.3 \) is about 8.04 bps/Hz, and it can be interpreted that the online learning loss is more dependent on the interaction records of the AMC agent than the threshold probability. Moreover, it was confirmed that the proposed scheme of the NEC agent with step-size varying algorithm can effectively reduce the online learning loss compared to the DQN agent with the fixed step-size scheme in [20]. Thus, we expect that the proposed scheme

**B. ONLINE LEARNING ANALYSIS**

The simulation results on the online learning loss of the proposed AMC scheme and the other DQN-based AMC schemes are shown in Figures 6 and 7. Each simulation result includes the interaction records of each agent and the derived optimal MCS region boundaries. Moreover, due to the randomness of probability, the simulation results were obtained with the averaged results of 1,000 times of the loss evaluation process. As seen in Figure 6, the AMC scheme with the fixed step-size NEC-based AMC scheme yields significant performance improvement on online loss compared to the DQN algorithm. It is observed that the NEC algorithm requires 2,502 unit time to reach the suboptimal state, and the cumulative loss is 2,190.17 bps/Hz for \( p_l = 0.1 \) and 2,190.9 bps/Hz for \( p_l = 0.3 \). Meanwhile, the dueling DDQN algorithm yields 3,184 bps/Hz for \( p_l = 0.1 \) and 3,183.27 bps/Hz for \( p_l = 0.3 \) within 4,874 unit times; the DDQN algorithm yields 4,919.15 bps/Hz for \( p_l = 0.1 \) and 4,911.11 bps/Hz for \( p_l = 0.3 \) within 6,855 unit times; the DQN algorithm yields 6,947.15 bps/Hz for \( p_l = 0.1 \) and 6,941.84 bps/Hz for \( p_l = 0.3 \) within 7,470 unit times, respectively. It is clearly seen that the online loss can be reduced to less than 1/3 compared to the DQN algorithm case [20] by employing advantageous properties of the NEC algorithm.

In Figure 7, the simulation results of the online learning loss evaluation are shown. It is observed that the NEC-based AMC scheme with the step-size varying algorithm reaches the suboptimal state at 1,900 unit time step, and its cumulative online loss is 410.08 bps/Hz for \( p_l = 0.1 \) and 408.56 bps/Hz for \( p_l = 0.3 \). On the other hand, the dueling DDQN algorithm yields 972.86 bps/Hz for \( p_l = 0.04 \) and 971.97 bps/Hz for \( p_l = 0.3 \) within 2,340 unit times; the DDQN algorithm yields 968.05 bps/Hz for \( p_l = 0.1 \) and 966.67 bps/Hz for \( p_l = 0.3 \) within 3,016 unit times; the DQN algorithm yields 1,623.1 bps/Hz for \( p_l = 0.1 \), and 1,623.45 bps/Hz for \( p_l = 0.3 \) within 4,000 unit times, respectively. It can be confirmed that the NEC with the step-size varying algorithm can reduce the cumulative loss to about 1/4 compared to the DQN case [20]. From the simulation results, it has been observed that the maximum difference between the cases of \( p_l = 0.1 \) and \( p_l = 0.3 \) is about 8.04 bps/Hz, and it can be interpreted that the online learning loss is more dependent on the interaction records of the AMC agent than the threshold probability. Moreover, it was confirmed that the proposed scheme of the NEC agent with step-size varying algorithm can effectively reduce the online learning loss compared to the DQN agent with the fixed step-size scheme in [20]. Thus, we expect that the proposed scheme
will provide the implemental advantages in an inter-satellite communication environment with limited energy and communication resources.

V. CONCLUSION

In this paper, we have proposed the NEC-based AMC scheme with the step-size varying algorithm. The proposed scheme adjusted the MCS region boundaries with the interactions of the NEC agent. The lookup and write operations of each DND enabled the agent to make the optimal decisions on actions by employing the previous experience of the agent. Compared to the work in [20], the proposed scheme adopted an advanced deep reinforcement learning algorithm in the practical inter-satellite communication systems to significantly reduce the number of trials of the agent deriving the optimal MCS region boundaries while using the less volume of memory. Further, the proposed scheme effectively accelerated this process with the step-size varying algorithm. It was confirmed that the proposed scheme can reduce the number of trials to less than 1/8 compared to the DQN-based adaptive modulation scheme in [20], and also the proposed scheme can achieve the fastest convergence to the suboptimal state, compared to the dueling DDQN, DDQN, and DQN algorithm-based schemes. Moreover, it was observed that the proposed AMC scheme yields the higher spectral utilization rate by assigning proper MCS. We also employed the online learning loss evaluation algorithm, which can be viewed as kind of the loss evaluation process in a practical environment, to calculate the spectral utilization rate loss during the convergence to the suboptimal state. From the simulation results, it was confirmed that the online learning loss can be effectively reduced by employing the NEC-based AMC scheme with the step-size varying algorithm. Thus, it can be expected that the proposed scheme has advantages in resource-limited environment of satellite communications, considering the smaller number of trials of the agent, the less volume of memory, and the less amount of online loss. Especially, the proposed scheme can be effective in 6G mobile communication systems and next generation wireless communication systems that are expected to form a massive and dense network, where the link is changed in real time and an optimal MCS region is required.

As for the future work, it can be viewed as an extension of the proposed deep reinforcement learning framework into another domain of the adaptive communication systems, such as adaptive transmitting power level control and spectrum resource allocation schemes. Moreover, to optimize the training process and to transfer the knowledge domain of the reinforcement learning agent, a transfer learning algorithm can be considered for the future work.

References

[1] H. Wang, Q. Chang, Y. Xu, and X. Li, “Adaptive narrow-band interference suppression and performance evaluation based on code-aided in GNSS inter-satellite links,” IEEE Syst. J., vol. 14, no. 1, pp. 538-547, Mar. 2020.
[2] Y. Bai, Y. Guo, X. Wang, and X. Lu, “Satellite-ground two-way measuring method and performance evaluation of BDS-3 inter-satellite link system,” IEEE Access, vol. 8, pp. 157530-157540, Aug. 2020.
[3] R. Radhakrishnan, W. W. Edmonson, F. Aghf, R. M. R.-Osorio, F. Pinto, Jr., and S. C. Burleigh, “Survey of inter-satellite communication for small satellite systems,” in Proc. 11th Int. Conf. on Satellite Commun. and Netw., July 2018.
[4] V. S. Reddy, “The SpaceX effect,” New Space, vol. 6, no. 2, pp. 125-134, June 2018.
[5] B. Kelly, L. Nevile, E. Draffan, and S. Fanou, “One world, one web but great diversity,” in Proc. W4A Conf., Beijing, China, Apr. 2008, pp. 141-147.
[6] A. Svensson, “An introduction to adaptive QAM modulation schemes for known and predicted channels,” Proc. IEEE, vol. 95, no. 12, pp. 2322-2336, Dec. 2007.
[7] S. T. Chung and A. J. Goldsmith, “Degrees of freedom in adaptive modulation: A unified view,” IEEE Trans. Commun., vol. 49, no. 9, pp. 1561-1571, Sept. 2001.
[8] M.-S. Alouini and A. J. Goldsmith, “Adaptive modulation over Nakagami fading channels,” Wireless Pers. Commun., vol. 13, nos. 1–2, pp. 119-143, May 2000.
[9] A. J. Goldsmith and S.-G. Chua, “Variable-rate variable-power MQAM for fading channels,” IEEE Trans. Commun., vol. 45, no. 10, pp. 1218-1230, Oct. 1997.
[10] O. Kodheli et al., “Satellite communications in the New Space era: A survey and future challenges,” IEEE Commun. Surveys Tut., vol. 23, no. 1, First Quarter. 2021.
[11] J. Huang, Y. Su, W. Liu, and F. Wang, “Adaptive modulation and coding techniques for global navigation satellite system inter-satellite communication based on the channel condition,” IET Commun., vol.10, no. 16, pp. 2091-2095, Oct. 2016.
[12] R. Xue, H. Yu, and Q. Cheng, “Adaptive coded modulation based on continuous phase modulation for inter-satellite links of global navigation satellite systems,” IEEE Access, vol. 6, pp. 20652-20662, May 2018.
[13] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
[14] W.-C. Chung, C.-J. Chang, K.-T. Feng, and Y.-Y. Chen, “An MIMO configuration mode and MCS level selection scheme by fuzzy Q-learning for HSPA systems,” IEEE Trans. Mobile Comput., vol. 11, no. 7, pp. 1151-1162, July 2012.
[15] C.-J. Chang, C.-Y. Chang, and F.-C. Ren, “Q-learning-based hybrid ARQ for high speed downlink packet access in UMTS,” in Proc. 2007 IEEE 65th Veh. Tech. Conf., Dublin, Ireland, Apr. 2007, pp. 2610-2615.
[16] J. Ge, Y.-C. Liang, J. Joung, and S. Sun, “Deep reinforcement learning for distributed dynamic MISO downlink-beamforming coordination,” IEEE Trans. Commun., vol. 68, no. 10, Oct. 2020.
[17] R. Sarikhani and F. Keynia, “Cooperative spectrum sensing meets machine learning: Deep reinforcement learning approach,” IEEE Commun. Lett., vol. 24, no. 7, July 2020.
[18] W. Chen, X. Qiu, H.-N. Dai, Z. Zheng, and Y. Zhang, “Deep reinforcement learning for internet of things: A comprehensive survey,” IEEE Commun. Surveys Tut., Early Access, Apr. 2021.
[19] S. Mashhadi, N. Ghiasi, S. Farahmand, and S. M. Razavizade, “Deep reinforcement learning based adaptive modulation with outdated CSI,” IEEE Commun. Lett., vol. 25, no. 10, pp. 3291-3295, July 2021.
[20] D. Lee, Y. G. Sun, S. H. Kim, I. Sim, Y. M. Hwang, Y. Shin, D. I. Kim, and J. Y. Kim, “DQN-based adaptive modulation scheme over wireless communication channels,” IEEE Commun. Lett., vol. 24, no. 6, pp. 1289-1293, June 2020.
[21] J. P. Leite, P. H. P. de Carvalho, and R. D. Vieira, “A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems,” in Proc. IEEE Wireless Commun. Net. Conf. (WCNC), Shanghai, China, Apr. 2012, pp. 809-814.
[22] V. Minh et al., “Human-level control through deep reinforcement learning,” Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[23] P. V. R. Ferreira et al., “Multi-objective reinforcement learning for cognitive satellite communications using deep neural network ensembles,” IEEE J. Sel. Areas Commun., vol. 36, no. 5, pp. 1030–1041, May 2018.
[24] P. V. R. Ferreira et al., “Reinforcement learning for satellite communications: From LEO to deep space operations,” IEEE Commun. Mag., vol. 57, no. 5, pp. 70–75, May 2019.
[25] S. Valkic and O. Winther, “The effect of correlated input data on the dynamics of learning,” in Proc. Adv. Neural Inf. Process. Syst.(NIPS), Denver, CO, USA, Dec. 1996, vol. 9, pp. 169-175.
[26] H. van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double Q-learning,” in Proc. AAAI, Phoenix, AZ, USA, Feb. 2016, pp. 2094-2100.
[27] Z. Wang, T. Schaul, M. Hessel, H. van Hasselt, M. Lanctot, and N. de Freitas, “Dueling network architectures for deep reinforcement learning,” in Proc. Int. Conf. Mach. Learn., NY, USA, June 2016, vol. 48, pp. 1995-2003.

[28] A. Pritzel, B. Uria, S. Srinivasan, A. Puigdomenech, O. vinyals, D. Hassabis, D. Wierstra, and C. Blundell, “Neural episodic control,” in Proc. Int. Conf. Mach. Learn., Sydney, Australia, Aug. 2017, vol. 70, pp. 2827-2836.

[29] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, “Applications of deep reinforcement learning in communications and networking: A survey,” IEEE Commun. Surveys Tut., vol. 21, no. 4, Fourth Quarter, 2019.

[30] W. W. Wu, D. Haccoun, R. Peile, and Y. Hirata, “Coding for satellite communication,” IEEE J. Sel. Areas Commun., vol. 5, no. 4, pp.724-748, May 1987.

[31] M. K. Simon and M.-S. Alouini, Digital Communication Over Fading Channels: A Unified Approach to Performance Analysis, Hoboken, NJ, USA: Wiley, 2000.

[32] B. Sklar, Digital Communications, vol. 2, Upper Saddle River, NJ, USA: Prentice Hall, 2001.

[33] O. Anschel, N. Baram, and N. Shimkin, “Averaged-DQN: Variance reduction and stabilization for deep reinforcement learning,” in Proc. Int. Conf. Mach. Learn., Sydney, Australia, Aug. 2017, vol. 70, pp. 176-185.

[34] S. Thrun and A. Schwartz, “Issues in using function approximation for reinforcement learning,” in Proc. Connectionist Models Summer, Hillsdale, NJ, USA, Nov. 1993, pp. 1-9.

[35] A. Graves et al., “Hybrid computing using a neural network with dynamic external memory,” Nature, vol. 538, pp. 471-476, Oct. 2016.

[36] S. Ruder, “An overview of gradient descent optimization algorithms,” arXiv preprint arXiv:1609.04747, 2016.

[37] J. L. Bentley, “Multidimensional binary search trees used for associative searching,” Commun. ACM, vol. 18, no. 9, pp. 509-517, Sept. 1975.

[38] B. Muhammad, V. Alberti, A. Marassi, E. Cianca, and M. Messerotti, “Performance assessment of GPS receivers during the September 24, 2011 solar radio burst event,” J. Space Weather Space Clim., vol. 5, no. 32, Oct. 2015.

[39] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, Introduction to Algorithms 3rd Ed. Cambridge, MA, USA: MIT Press, 2009.

**Donggu Lee** (S’19) received the B.Sc. and M.Sc. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Korea, in 2019 and 2021, respectively. His research interests include machine learning application for wireless communication systems, satellite communications, adaptive communication systems, etc.

**Young Ghyu Sun** (S’18) received the B.Sc. (summa cum laude) and M.Sc. degree from the Department of Electronic Engineering, Kwangwoon University, Seoul, Korea, in 2018 and 2020, respectively, where he is currently pursuing the Ph.D. degree. He was a recipient of the IEEE Student Paper Gold Award in 2020. His research interests include wireless communications, wireless energy harvesting, deep learning, Internet of energy, etc.

**Isaac Sim** (S’16) received the B.Sc. and M.Sc. degrees (magna cum laude) from the Department of Electronic Convergence Engineering, Kwangwoon University, Seoul, South Korea, in 2016 and 2018, respectively, where he is currently pursuing the Ph.D. degree. His research interests include wireless energy harvesting, the internet of energy, wireless sensor networks, and anti-jamming techniques.

**Jae-Hyun Kim** (S’89-M’99) received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Hanyang University, Ansan, South Korea, in 1991, 1993, and 1996, respectively. In 1996, he was with the Communication Research Laboratory, Tokyo, Japan, as a Visiting Scholar. From April 1997 to October 1998, he was a Postdoctoral Fellow at the Department of Electrical Engineering, University of California at Los Angeles. From November 1998 to February 2003, he worked as a member of technical staff at the Performance Modeling and QoS Management Department, Bell laboratories, Lucent Technologies, Holmdel, NJ, USA. Since 2003, he has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, as a Professor. His research interests include medium access control protocols, QoS issues, cross layer optimization for wireless communication, satellite communication, and mobile data offloading. He is the Center Chief of Satellite Information Convergence Application Services Research Center (SICAS) sponsored by the Institute for Information and Communications Technology Promotion in Korea. He has been the Chairman of the Smart City Committee of 5G Forum in Korea, since 2018. He is the Executive Director of the Korea Institute of Communication and Information Sciences (KICS). He is a member of the KICS, the Institute of Electronics and Information Engineers, and the Korea Information Science Society.

**Yoan Shin** (M’90-SM’07) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, South Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from The University of Texas at Austin in 1992. From 1992 to 1994, he was a Member of Technical Staff with Microelectronics and Computer Technology Corporation (MCC), Austin, TX, USA. Since 1994, he has been with the School of Electronic Engineering, Soongsil University, Seoul, where he is currently a Professor. From 2015 to 2017, he has served as the Dean of the Office of Research of Soongsil University. From 2009 to 2010, he was a visiting professor with The University of British Columbia, Vancouver, Canada. He is the Executive Vice-president of the Korea Institute of Communication and Information Sciences (KICS) for the year 2021. He has been...
serving as an organizing/technical committee member for various prominent international conferences. In particular, he was the Technical Program Committee Chair of APWCS 2013, the General Co-chair of APWCS 2014 and APWCS 2015, and the Keynote/Panel Co-chair of WCNC 2020. He is also the Executive Vice-chair of ICC 2022. His area of interests includes wireless communication and communication signal processing.

**Dong In Kim** (S’89-M’91-SM’02-F’19) received the Ph.D. degree in electrical engineering from the University of Southern California, Los Angeles, CA, USA, in 1990. He was a Tenured Professor with the School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada. Since 2007, he has been with Sungkyunkwan University, Suwon, South Korea, where he is currently a Professor with the College of Information and Communication Engineering. He was a first recipient of the NRF of Korea Engineering Research Center in Wireless Communications for RF Energy Harvesting from 2014 to 2021. From 2002 to 2011, he served as an Editor and a Founding Area Editor of Cross-Layer Design and Optimization for the IEEE Transactions on Wireless Communications. From 2008 to 2011, he served as the Co-Editor-in-Chief of the IEEE/KICS Journal of Communications and Networks. He served as the Founding Editor-in-Chief of the IEEE Wireless Communications Letters from 2012 to 2015. From 2001 to 2014, he served as an Editor of Spread Spectrum Transmission and Access for the IEEE Transactions on Communications. Since 2015, he is serving as an Editor-at-Large of Wireless Communication I for the IEEE Transactions on Communications. He has been elevated to the grade of Fellow of the IEEE for contributions to cross-layer design of wireless communications systems, Fellow of the Korean Academy of Science and Technology (KAST) and Fellow of the National Academy of Engineering of Korea (NAEK).

**Jin Young Kim** (M’95-SM’08) received the B. S., M. S., and Ph. D. degrees from the School of Electrical Engineering, Seoul National University (SNU), Seoul, Korea, in 1991, 1993, and 1998, respectively. He was Member of Research Staff at the Institute of New Media and Communications (INMC) and at the Inter-university Semiconductor Research Center (ISRC) of the SNU from 1994 to 1998. He was Postdoctoral Research Fellow at the Department of Electrical Engineering, Princeton University, NJ, U.S.A, from 1998 to 2000. He was Principal Member of Technical Staff at the Central Research and Development Center, SK Telecom, Korea, from 2000 to 2001. He is currently Full Professor at the School of Electronics Engineering, Kwangwoon University, Seoul, Korea. He had his sabbatical leave as Visiting Scientist at the LIDS (Laboratory of Information and Decision Systems), Massachusetts Institute of Technology (M.I.T), MA, U.S.A from 2009 to 2010. His research interests include design and implementation of wireline/wireless multimedia communication systems for applications to spread-spectrum, cognitive radio, ultrawideband (UWB), powerline communication with basis on modulation/demodulation, synchronization, and detection and estimation theory. He received the Best Paper Awards from several academic conferences and societies including Jack Neubauer Best Systems Paper Award from IEEE VT Society (2001), the Award of Prime Minister of Korea Government (2011). He is now Senior Member of IEEE, Regular Member of IET and IEICE.