Deep Deterministic Policy Gradient-Based Spectral Efficiency for Massive MIMO Communication System

Nasaruddin Nasaruddin*, Afzal Risky, Yunida Yunida, and Ramzi Adriman
Department of Electrical and Computer Engineering, Universitas Syiah Kuala, Banda Aceh, Indonesia
Email: nasaruddin@usk.ac.id (N.N.), afzal.r@mhs.usk.ac.id (A.R.), yunida@usk.ac.id (Y.Y.),
ramzi.adriman@usk.ac.id (R.A.)

Manuscript received July 25, 2024; revised November 6, 2024; accepted December 5, 2024
*Corresponding author

***Abstract*—Massive MIMO is a new configuration of MIMO technology that uses many antennas up to the order of hundreds to serve tens or hundreds of User Equipment (UE) at the same time and frequency. Massive MIMO technology is one of the spatial diversity techniques used to increase the Spectral Efficiency (SE) of current Fifth-Generation (5G) communication systems. Massive MIMO has a high complexity in signal processing because it serves a large amount of user traffic at the same time. Therefore, this paper proposes using the Deep Deterministic Policy Gradient (DDPG), a deep learning method that significantly improves both SE and runtime performance, making it highly effective for large-scale wireless communication systems. In the simulation, we are modeling a massive MIMO system with multiple Access Points (APs) and User Equipment (UEs). We are training the channel using the proposed DDPG model. Then, we analyze each end-user path's Signal-to-Interference plus Noise Ratio (SINR) and compare it with conventional massive MIMO (without deep learning). In addition, the complexity of the proposed DDPG model in terms of runtime is analyzed and compared with the Convex Optimization Algorithm (CVX). The simulation results indicate that the performance of the massive MIMO system is improved with the proposed DDPG model. It achieves an optimal and higher spectral efficiency (SE) of 85% compared to not using the DDPG method. Additionally, it achieves an average Signal-to-Interference-plus-Noise Ratio (SINR) of 19.54 dB, while the conventional method only provides an average SINR of 15.32 dB. Furthermore, the proposed DDPG model has a lower complexity with a runtime ratio of 1:8000 compared to the CVX algorithm for the same number of epochs.**

***Index Terms*—Convex optimization, Deep Deterministic Policy Gradient (DDPG), deep learning, massive MIMO, spectral efficiency**

I. INTRODUCTION

5G technology is the current generation of cellular communications networks that are used to transmit data at very high speeds, reaching 20 Gigabits per second (Gbps), very low latency of 1 ms, wide bandwidth availability, and supported by the use of antennas on a large scale [1, 2]. The 5G cellular network utilizes a radio spectrum, enabling the technology to connect with multiple devices simultaneously. Apart from that, 5G technology also has a big impact on the Mobile Broadband (MBB) internet network, which can be connected to machine-to-machine (M2M) and Internet-of-Things (IoT) networks [3]. Unlike previous generations of communication technology, 5G utilizes the New Radio (NR) spectrum, offering adaptable access speeds based on the spectrum band used. Based on the 5G Public Private Partnership (5G PPP) in 2015, the vision of this fifth-generation technology is to become a key technology in the digital world with the support of ultra-high band infrastructure [4].

Several key technologies support 5G technology to improve wireless communication network performance. These include beamforming, millimeter wave (mmWave), full-duplex, small cell, and massive Multiple Input Multiple Output (MIMO) techniques [5–7]. Beamforming enables directional transmission of signals, improving signal strength and reducing interference, particularly in dense environments. The mmWave technology, operating in the 24 GHz to 100 GHz range, provides high data rates and bandwidth, although it has a limited range and is susceptible to obstructions. Full-duplex communication, allowing simultaneous transmission and reception on the same frequency, effectively doubles spectral efficiency compared to traditional half-duplex systems. Small cells, which are low-power base stations, improve network coverage and capacity in specific areas, especially in urban or indoor settings, and complement larger macro-cell networks [6]. The MIMO system is a spatial diversity technique that simultaneously uses multiple antennas to transmit information from numerous user equipment (UEs). On the other hand, a massive MIMO system is a new configuration form of traditional MIMO in which multiple antennas, on the order of hundreds, operate simultaneously to serve tens to hundreds of UEs on the same frequency at the same time [7].

In recent years, deep learning has started to be used to facilitate performance analysis of 5G communication systems, especially massive MIMO, which has high system complexity [8, 9]. It can provide processing results that are faster, more consistent, more reliable, and easier to configure. Therefore, this method is very suitable for use in complex communication technologies such as massive MIMO. This is also a new paradigm in machine learning that can function like the human brain and build multi-layered neural networks. Generally, deep learning is trained on multiple similar examples, allowing the

doi: 10.18178/ijeetc.14.1.13-22

machine to learn from previously available data to optimize the machine's error bounds and have better generalization capabilities [10, 11].

Research on massive MIMO over the last twenty years has become a basic approach to improving Spectral Efficiency (SE) [12–15]. So massive MIMO is one of the spatial diversity techniques used to increase SE and energy efficiency in 5G communication systems [16–19]. SE is an important performance metric in massive MIMO systems, which is related to the suitability of the amount of information contained in a particular channel. In other words, SE is the real data speed in bits per second (bps), which is directly proportional to the amount of bandwidth. Massive MIMO has high complexity in signal processing due to the use of many antennas. The use of many antennas in massive MIMO corresponds to a significant increase in data traffic at the same time and frequency because the performance of Massive MIMO systems decreases due to the power of data processing. Therefore, Deep Learning (DL) is used to overcome complexity problems in massive MIMO.

One of the DL methods for analyzing SE in massive MIMO communication systems is Deep Reinforcement Learning (DRL). DRL is a branch of machine learning, where machines can continue to learn from their environment to obtain more optimal performance parameters [20, 21]. In research conducted by Amjad Iqbal *et al.* [22] in the form of optimizing energy and spectral efficiency in the Cloud Radio Access Network (CRAN) with the DRL method based on the dueling Double Deep q-Network (D3QN) approach as a control policy to achieve maximum energy and spectral efficiency. The D3QN-based DRL method significantly improves CRAN system performance with dynamic channel access, mobile offloading, and optimal access management.

Furthermore, the DRL method has also been used as a power allocation algorithm in MIMO-free cells [23]. In the research, two DRL methods were utilized: Deep q-Network (DQN) and Deep Deterministic Policy Gradient (DDPG). These methods aim to maximize the sum Spectrum Efficiency (SE) in massive MIMO-free cells. The numerical simulation results show that the sum SE value is 33% higher than The weighted Minimum Mean Square Error (WMMSE) precoding method, and the execution time is 0.1% higher than the WMMSE method. Unlike traditional supervised learning, both approaches have low computational complexity, which requires a large data set with a complex computational algorithm.

DDPG is particularly adept at managing continuous action spaces, a critical requirement for power control and beamforming tasks where actions (e.g., adjusting transmission power or antenna weights) vary continuously. The actor-critic framework employed by DDPG is well-suited for optimizing such high-dimensional, complex systems, allowing for efficient exploration and learning of optimal policies. As described by Lillicrap *et al.* [24], this framework facilitates precise decision-making and minimizes computational complex-ity compared to traditional optimization methods. Consequently, the application of DDPG yields significant improvements in both spectral efficiency and runtime performance, making it a highly effective approach for large-scale wireless communication systems. Furthermore, other studies, such as those by Zhao *et al.* [23], have demonstrated the superiority of DDPG in enhancing the performance of massive MIMO systems.

Massive MIMO systems, which utilize many antennas to simultaneously serve multiple UEs, introduce significant signal processing and resource allocation complexity. As the scale of these systems increases, maintaining high SE becomes a critical challenge. Traditional optimization techniques, such as Convex Optimization (CVX), though effective in smaller systems, struggle to cope with the scalability demands of large-scale networks. Specifically, as the number of Access Points (APs) and UEs grows, the computational overhead of these methods increases substantially, making them less efficient for real-time applications in massive MIMO.

Convex optimization, as outlined in [25], is known for its ability to provide optimal solutions to various engineering problems. However, its applicability diminishes in large-scale systems like massive MIMO, due to the increased dimensionality and the time-consuming nature of solving high-complexity optimization problems in real-time. This limitation is further supported by Zheng *et al.* [26], who demonstrate that convex optimization-based precoding techniques become computationally prohibitive as the number of antennas and users increases. These findings underscore the need for more scalable and computationally efficient methods to address the challenges of massive MIMO systems.

Based on the facts, this paper proposes a DDPG-based DRL method to analyze the SE performance of 5G massive MIMO communication systems using the Minimum Mean-Square Error (MMSE) precoder technique. This is intended to align the UE's need for high data rates with their spectrum. Meanwhile, spectrum availability in the 5G massive MIMO communication system is limited, so it is necessary to increase the SE to optimize the performance of massive MIMO.

The main contributions of this paper can be summarized as follows:

1) We analyze the close form of SE using a DDPG-based DRL for a 5G massive MIMO communication system.
2) We provide the DDPG algorithm to obtain the SE of a massive MIMO system model.
3) We analyze the complexity of the proposed DDPG model and compare it with the Convex Optimization (CVX) algorithm.

The structure of this paper is as follows: Section II gives related works, Section III provides the research method, and Section IV gives results and discussion. Section V presents the conclusions.

II. Related Works

A. Deep Learning

The MIMO communication system is a form of evolution from the conventional one-antenna system, which requires at least two antennas on both the sending and receiving sides. MIMO systems have advantages in terms of increasing signal reliability and reducing communication interruptions when disturbances such as multipath fading and interference occur [27]. Currently, MIMO has developed further into massive MIMO with the emergence of high-frequency communication systems, which can be seen from the use of antennas in the order of hundreds to serve hundreds to thousands of users at the same time [28]. Massive MIMO can increase the capacity of the communication system without requiring additional spectrum. This is because the greater the number of transmitter and receiver antennas, the greater the possible path that the signal can take to reach its destination so that data speed and system reliability can increase. However, the use of a large number of antennas makes signal processing more complex due to the large number of information signals being processed simultaneously. So, this causes a decrease in overall system performance, and the signal processing time required is also longer. Therefore, a system is needed that can reduce the level of complexity in massive MIMO systems, one of which is by applying the Deep Learning (DL) method [8].

B. Deep Reinforcement Learning

Recently, DL methods have been proposed to reduce the complexity of MIMO systems and achieve optimal performance [29–33]. Commonly used DL-based techniques include supervised learning and reinforcement learning. Supervised learning-based DL methods use Deep Neural Networks (DNNs) to predict system outputs with highly complex computational algorithms [34]. Furthermore, Deep Reinforcement Learning (DRL) can also be used to determine the mapping of large-scale fading coefficients and increase the sum rate through channel amplification [24]. On the other hand, DL techniques based on reinforcement learning focus on what actions the agent needs to perform when viewing the environment, such as the need to increase the cumulative reward. Reinforcement learning also known as DRL is a DL technique that does not require a training dataset, making it suitable for dynamic wireless networks.

C. Deep Deterministic Policy Gradient

Reinforcement learning methods are used to solve performance optimization problems in cellular networks [35, 36]. However, none of these studies consider massive MIMO systems. In recent years, several of his DRL techniques for large-scale networks have been proposed that apply the Deep Q-Learning Network (DQN) algorithm to real-world scenarios and are expected to improve performance. However, since this approach is limited only to discrete-form data control problems, continuous performance optimization in the downlink stage can reduce the cumulative rate of the system. Tiong *et al.* subsequently [37] proposed a new DRL algorithm called DDPG. This can be implemented in massive self-free MIMO systems, especially in the continuous power control process during the downlink phase. In addition, Zhao *et al.* [23] also proposed two DRL methods based on dynamic power allocation, DQN, and DDPG, for massive cell-free MIMO systems with moving users to maximize the downlink sum rate.

Massive MIMO systems are highly dynamic and involve continuous interactions with multiple user devices, antennas, and varying channel conditions. The DDPG algorithm is particularly well-suited because:

1) It can adapt to the continuous and high-dimensional action space required for controlling power allocation and beamforming in real time.
2) Its actor-critic structure allows for more precise optimization of performance metrics like SE in a computationally efficient manner.
3) Its sample efficiency and exploration strategy make it feasible to train the model without excessive data requirements, which is advantageous when real-world interactions are limited or costly.

In this paper, we focused on DDPG because it is particularly well-suited for the continuous action space inherent in massive MIMO systems. Tasks like power allocation and beamforming require continuous optimization [23]. CNNs, while powerful in feature extraction tasks (such as image processing), are less commonly applied to such continuous control problems [24]. Similarly, alternative RL methods like DQN are better suited for discrete action spaces, which is less applicable to our use case [38].

III. METHOD

A. System Model

The considered system model in this paper is a downlink cell-free massive MIMO network with an operating frequency of 1.9 GHz and a bandwidth of 20 MHz, consisting of a Base Station (BS), K UEs, and M Access Points (APs), as shown in Fig. 1. Each UE and AP have one antenna, and all UEs are served simultaneously by all APs within a coverage area of 1×1 km^2.

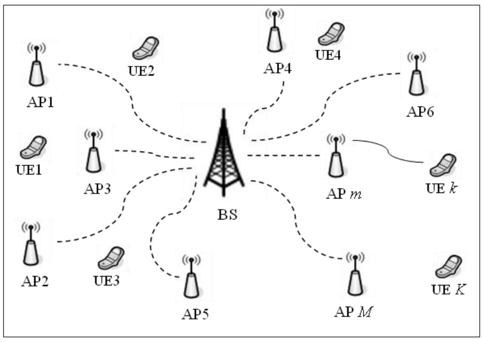


Fig. 1. Return training of the proposed DDPG model in the downlink phase.

The channel model in the system is Rayleigh fading. The different channel models do not influence the development of the proposed model. The channel coefficient between the m-th AP and the kth UE can be written as

$$g_{m,k} = \sqrt{\beta_{m,k}}h_{m,k} \tag{1}$$

where $\beta_{m,k}$ and $h_{m,k}$ respectively are large and small-scale fading coefficients which are distributed independently and identically (i.i.d) with a random variable $\mathcal{CN}(0, 1)$ with $m = 1, 2, \dots, M$ and $k = 1, 2, \dots, K$.

The information transmission process is divided into two parts, the uplink and downlink phases (payload data transmission), as follows:

1) Uplink phase

At this stage, all UEs first transmit the pilot sequence $\boldsymbol{\varphi}_k$ to all APs during the training phase with a sample size τ_p used by the *k*th UE which is denoted as $\sqrt{\tau_p}\boldsymbol{\varphi}_k$. So, the pilot signal vector received at the *m*-th AP can be written as

$$\mathbf{y}_{p,m} = \sqrt{\tau_p \rho_u} \sum_{k=1}^{K} g_{m,k}\, \boldsymbol{\varphi}_k + \mathbf{n}_{p,m} \tag{2}$$

where ρ_u is the normalized transmission power in the uplink phase for each symbol and $\mathbf{n}_{p,m} \backsim \mathcal{CN}$ (0,1) is additive white Gaussian noise (AWGN) on the m-th AP.

2) Downlink phase

At this stage, all APs transmit signals to all K UEs simultaneously. This can be explained as follows:

$$u_m = \sqrt{\rho_d} \sum_{k=1}^{K} \sqrt{\eta_{m,k}} \hat{g}^*_{m,k}\, q_k \tag{3}$$

where ρ_d is the normalized transmission power in the downlink phase for each symbol, q_k is the intended symbol for the *k*th UE with a squared expected value $\mathrm{E}\{|\mathrm{q_k}|^2\} = 1$, and $\eta_{m,k}$ is the power control coefficient with $\mathbb{E}\left\{\left|\eta_{m,k}\right|^2\right\} \leq \rho_d$. Meanwhile, the signal received on the *k*th UE can be written as

$$r_k = \sqrt{\rho_d} \sum_{m=1}^{M} \sqrt{\eta_{m,k}} g_{m,k} \hat{g}^*_{m,k}\, q_{\boldsymbol{k}} + \sqrt{\rho_d} \sum_{m=1}^{M} \sum_{\acute{k} \neq k}^{K} \sqrt{\eta_{m,\acute{k}}} g_{m,k} \hat{g}^*_{m,\acute{k}}\, q_{\acute{k}} + n_k \tag{4}$$

where n_k denotes the noise in the *k*th UE with a random variable $\mathcal{CN}$ (0,1).

B. Spectral Efficiency

Spectral Efficiency (SE) is the average number of bits of information per complex-valued sample transmitted on a channel (bits/s/Hz). SE must be considered in the channel between the UE and the BS, measured in bits/s/Hz/cell. The channel between BS and UE at a location certainly has different SE depending on the encoding/decoding scheme applied, but the maximum achievable SE value is important in transmission system design. To get a value of SE performance, it is necessary to analyze the downlink signal-to-interference-plus-noise ratio (SINR) equation for each *k*th UE as follows.

$$\mathrm{SINR}_k = \rho_d \left(\sum_{m=1}^{M} \sqrt{\eta_{m,k}}\, \gamma_{m,k}\right)^2 \times \left(\frac{1}{\rho_d \left(\sum_{\acute{k} \neq 1}^{M} \sqrt{\eta_{m,\acute{k}}} \gamma_{m,\acute{k}} \frac{\beta_{m,k}}{\beta_{m,\acute{k}}}\right)^2 \left|\boldsymbol{\varphi}_{\acute{k}}^H \boldsymbol{\varphi}_k\right|^2} + \frac{1}{\rho_d \sum_{m=1}^{M} \sum_{\acute{k} \neq k}^{K} \eta_{m,\acute{k}} \gamma_{m,\acute{k}} \beta_{m,k} + 1} \right) \tag{5}$$

Meanwhile the SE of each *k*th UE [38] can be written as

$$SE_k = log_{10}(1 + SINR_k). \tag{6}$$

C. DDPG Model

DDPG is a model of DRL that uses a Deep Neural Network (DNN) as a policy network and produces output in the form of an action DDPG uses an actor method $A_C(\delta_a)$ to take action by looking at a state s, where the critical actor $A_C(\delta_a)$ is a network policy with δ_a as a network parameter. Criticism $C_r(\delta_c)$ is a parameter for evaluating action a, where this criticism is a different network from the critical actor. The optimal policy or policy from DDPG is formulated into Eq. (7) [36].

$$\mathrm{op_{DDPG}} = \arg \max_{Ac(\delta_a^{op})} Cr\left(\delta_c^{op}\right) \tag{7}$$

where $\mathrm{op_{DDPG}}$ is the optimal policy of DDPG, $A_C(\delta_a)$ and $C_r(\delta_c)$ are an actor and critic who collaborate with each other, respectively, to obtain optimal parameters δ_a^{op} and δ_c^{op} are as follows [36].

$$\delta_a^{op} = \arg \max_{\delta_a} L(\delta_a) \tag{8}$$

$$\delta_c^{op} = \arg \max_{\delta_c} L(\delta_c) \tag{9}$$

This DDPG system uses a multi-agent system for training $A_C\ (\delta_a)$ and $C_r(\delta_c)$. In contrast to other system models, DDPG has output from actors $A_C\ (\delta_a)$ in the form of a continuous value, which is shown in (10) [23].

$$a_{m,k}^t = \left[Ac(\delta_a)|_{s_{m,k}^t}\right]_0^{\sqrt{\rho_u}} \tag{10}$$

D. DDPG Algorithm

DDPG algorithm is oriented toward or recognizes the environment by interacting with it. DDPG agents can then learn from the feedback obtained to develop optimal policies. In general, DRL can learn about their environment based on Markov's decisions. The DDPG algorithm can be divided into three implementations: value-based, policy-based, and critical agent-based approaches. In this paper, an algorithm based on a policy approach is used $\mu(s|\theta_u)$ and value networks $Q\left(s, a\middle|\theta_Q\right)$ which is the actor value according to the following Bellman function [36]:

$$Q(s,a)^u = \eta_{(s,a,r,s')\in B}\left[r(s,a) + \varsigma_{a'}^{max} Q^u(s',a')\right] \tag{11}$$

Based on Eq. (11), the value of $Q(s,a)$ is the actor's value which is obtained variably as the system model updates the parameters. The DDPG system uses a critical network function to collect every change in parameter values. Critic uses the $\theta_Q \in R$ function and actor uses the $\theta_u \in R$ function, as in the following Poylack averaging equations [36]:

$$\theta^{Q'} \leftarrow \sigma\theta^Q + (1-\sigma)\theta^{Q'} \tag{12}$$

$$\theta^{u'} \leftarrow \sigma\theta^u + (1-\sigma)\theta^{u'} \tag{13}$$

where $\sigma \in (0,1)$ is a coefficient that determines the speed of change in target network parameters. In other words, $\sigma \in (0,1))$ is a measure of the learning rate of the DDPG system. Therefore, the DDPG algorithm consists of actor-critic, environment, state, and return value as illustrated in Fig. 2.

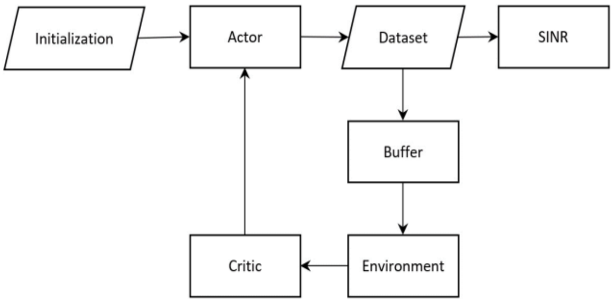


Fig. 2. Proposed DDPG algorithm for massive MIMO.

1) DDPG parameters

At this stage, the DDPG algorithm is structured as shown in Fig. 2, utilizing the parameters outlined in Table I to design the DDPG system.

TABLE I: THE SIMULATION PARAMETERS

Parameter	Value/Description
Operating frequency	1.9 GHz
Bandwidth	20 MHz
Coverage area	1×1 km^2
No. of APs and UEs	M APs, K UEs
Antenna setup	Single antenna per AP and UE
Channel model	Rayleigh fading
DDPG learning rate	10^{-3}
DDPG discount factor	0.99
Buffer size	10^6
Batch size	200
Training epochs	500
Exploration strategy	Ornstein-Uhlenbeck process
Neural network size	256 neurons per layer
Training time	12-24 hours
Hardware setup	NVIDIA RTX 2080
Software setup	TensorFlow and MATLAB

2) Actor

The actor in the proposed algorithm generates a policy from which an action is taken based on current circumstances. This actor serves as an agent that interacts with the massive MIMO system. An agent is in a state that acts based on a behavioral policy towards the environment. An action performed by an agent that modifies the state of the environment irreversibly (without returning to its original state). This actor interacts according to the parameters in Table I.

3) Dataset

The dataset in this paper was generated using the code in [36]. Then, the dataset is a training channel dataset that was generated by computer simulation using MATLAB software. This was done by considering the number of APs, UEs, and transmission power. In this paper, the dataset consists of 500 epochs for each UE.

The datasets are used to train system models to be more sensitive and recognize the input data. This process also generates weight and return values, which the model then stores as a dataset during the training process. Through this data collection, the model becomes more sensitive and able to recognize data that will be used for testing later on. The testing process for the DDPG system model. The DDPG system model at this stage already has weights and values from the previous training process. At this stage, the system model can recognize the data well and produce optimal SE values.

4) Buffer

The DDPG buffer stores all the states and rewards obtained when the DDPG actor interacts with the dataset. The critic and environment use the stored data functions to obtain updated information for each actor interacting at any time.

5) Environment

This function is used by DDPG as a condition or state obtained by the system when actors interact. This function specifically stores state values and is always updated after the DDPG actor interacts with the dataset. With this environment, the DDPG system can recognize a state by using a previously saved state. The goal is to find the best resulting sequence of actions that can provide optimal policy functions in a given situation. The environment in which the agent finds itself provides a value that indicates the quality of the action performed so that actions of the correct and appropriate quality can be selected.

6) Critic

This function is used to evaluate actions or interactions carried out by actors. This criticism is used to update every action the actor takes to get a better score. This function cannot be separated from the actor function because this function greatly influences the actions carried out by the actor on the dataset. The critic's role is to assess the actor's policies and guide the actor toward the optimal path through feedback. This method significantly reduces the burden of manually adjusting hyper-parameters in training and stabilizes its convergence. On the other hand, hyper-parameter tuning and unstable environments are still major challenges for most state-of-the-art DRL models, such as DDPG.

7) SINR

This stage is the output value from the DDPG system, in the form of a SINR value, which is then stored for each user dataset. After obtaining all SINR values for each user, the DDPG interaction process is also completed. The SE value for each user depends on the SINR value. The higher the SINR value, the higher the level of spectral efficiency. The SINR value in a massive MIMO system has a great influence because this massive MIMO system uses many antennas, which requires efficient spectrum usage. An illustration of the simulation stages can be seen in Fig. 3.

Fig. 3 illustrates that the process begins with generating a dataset, which is utilized for both training and testing. Initially, this dataset undergoes training before being fed into the DDPG system for further training. After training, the system is assessed for stability. If it is found to be unstable, the process returns to data training for retraining. Once the system achieves stability, the dataset is then utilized for testing, and the DDPG system also undergoes testing. Finally, the results from the DDPG system testing are analyzed in terms of SINR and SE to evaluate the system's performance.

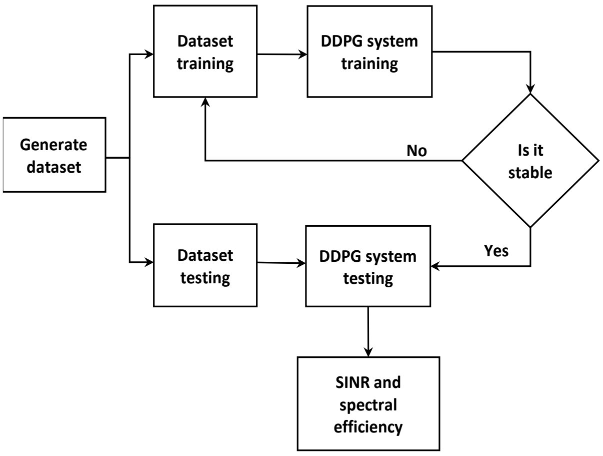


Fig. 3. Flowchart of the simulation stages.

IV. Result and Discussion

A. *Evaluation of Proposed DDPG Model*

The previously generated training channel dataset is then input into the DDPG method. In this process, we optimize the dataset using the DDPG method to achieve better SE performance parameters. In this process, several new parameters are used to design the DDPG algorithm, as shown in Table I.

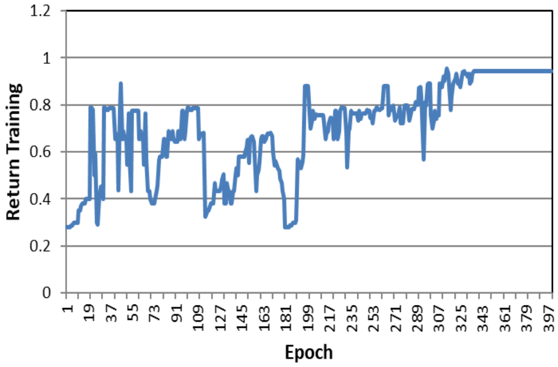


Fig. 4. Return training of the proposed DDPG.

The initial assessment involved testing the return training of the proposed DDPG on a massive MIMO system. The experimental results show that the return value of each epoch user has different values, as shown in Fig. 4. However, the DDPG system has the most optimal return value. For each epoch, ten training iterations are carried out, and the most optimal return value is determined based on the return value of the tenth iteration. The return value from the last iteration is stored in the buffer, and the critical weights are updated so that the DDPG system has the latest reference value for each epoch. The return value of the DDPG system is stable starting at the 336th epoch with a return value of 0.9453 and becomes increasingly stable until the 400th epoch. This stability value shows that the system has reached its optimal value and that the training data carried out is sufficient.

The system that has been tested and validated is then trained. The dataset used to perform these experiments consists of 10% or 50 epochs of data. The DDPG algorithm used for testing is the same as the one used during training and validation. However, the system model already has biases and weights from previous data. Thus, during testing, the model will optimally recognize channel data. The test return value of the proposed DDPG model has a higher value than the return value during testing as shown in Fig. 5. It can be observed that the testing return value for each user epoch is higher than the training return value. The number of epochs used in this testing process is 100 epochs or 20% of the dataset. The system has a stable return value at the 35th epoch with a return weight of 0.9506. The stable return value in this process is higher than the stable value in the training process which is only 0.9453.

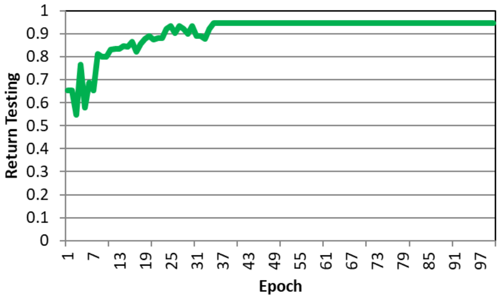


Fig. 5. Return testing of the proposed DDPG.

B. *Spectral Efficiency*

Once training and testing are complete, output data will be obtained as SINR values for each user. Based on Fig. 6, each user has 100 data epochs with sequentially different average SINR values: 19.6892 dBm, 19.4900 dBm, 19.5790 dBm, 19.3727 dBm, and 19.5928 dBm. SINR is the ratio of the strength of the main signal to the interference and noise mixed with the main signal. In this paper, the main transmitted signal strength was 200 mW, which is equivalent to 23.0103 dBm. The total average SINR value for all users is 19.5447 dBm, so compared to the transmitted signal strength, there is only a difference of 15.0606%, or in other words, the accuracy level of the SINR value is 84.9393%. Since the SINR value has a high accuracy value concerning the signal transmission value, it can be concluded that the obtained SE value also has a similar accuracy. Indeed, the SE value depends strongly on the SINR, as in (5).

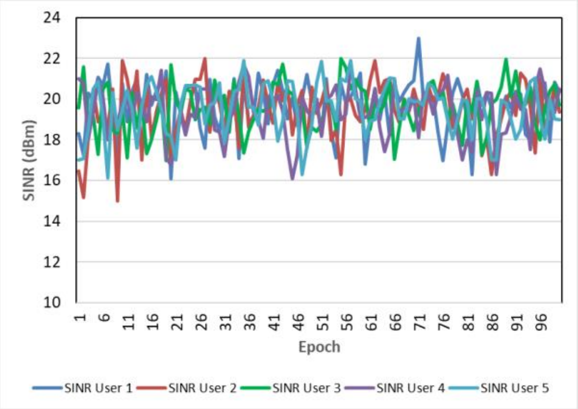


Fig. 6. SINR testing of the proposed DDPG.

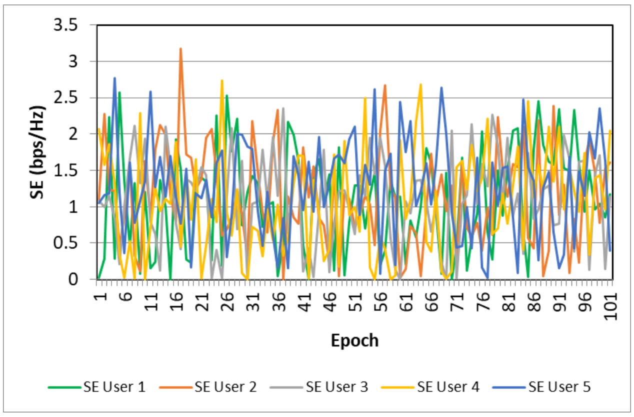


Fig. 7. SE of the conventional system.

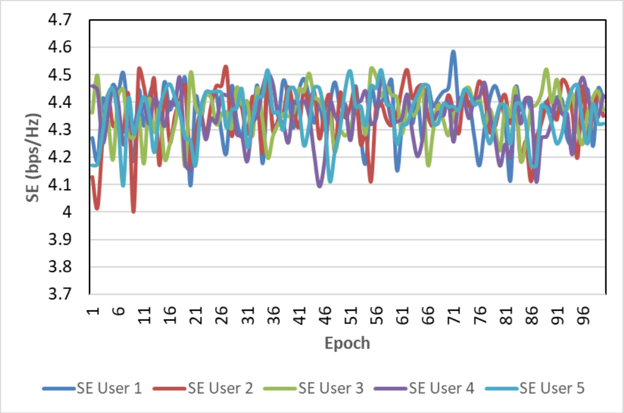


Fig. 8. SE of the proposed DDPG system.

The results for SE are achieved using the DDPG algorithm and the dataset generation algorithm, as shown in Figs. 7 and 8. The spectral efficiency values of these two figures show very clear differences. This is because the spectral efficiency using the DDPG system has a SINR value that is more stable and optimal. The SINR value determines the efficient level of spectral use in a communications network. The stable value in the DDPG system model is because the system model has sufficient initial data used in the training process, so at this testing stage, the system model can already recognize the data used.

The SE values for the conventional and the proposed DDPG models are illustrated in Figs. 7 and 8, respectively, are very different. It can be seen that when using the DDPG method on a cell-free massive MIMO system, each user has an average SE value of 1.1052 bps/Hz, 1.1571 bps/Hz, 1.1010 bps/Hz, 1.0462 bps/Hz, and 1.2563 bps/Hz. Meanwhile, in Fig. 8, the average SE value for user 1 is 4.3678 bps/Hz, user 2 is 4.3235 bps/Hz, user 3 is 4.2607 bps/Hz, user 4 is 4.3459 bps/Hz and user 5 is 4.3569 bps/Hz. From these average values, spectrum utilization seems more efficient and optimal in a massive 5G MIMO communication system. Optimal spectrum usage can significantly enhance the performance of 5G communication systems. The comparison of SE values using the DDPG system yields considerably better and more optimal results.

The proposed DDPG model achieved a significantly higher SE gain compared to the conventional model, with an average difference of 3.7 b/s/Hz, as shown in Fig. 9. The conventional system in Fig. 9 is a massive MIMO system that does not utilize Deep Reinforcement Learning (DRL) techniques like DDPG. Additionally, SE values tend to be more stable than with conventional methods. This shows that a DDPG-based massive MIMO system can effectively maximize spectrum allocation to each user terminal simultaneously.

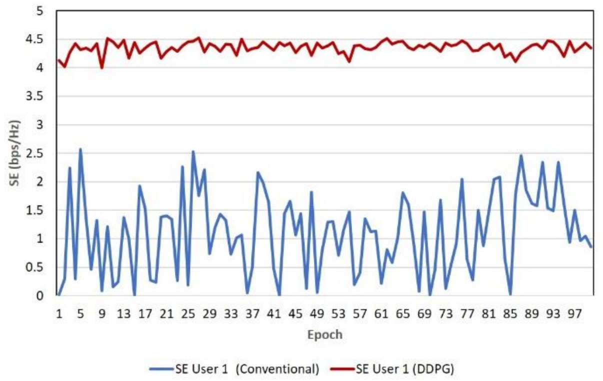
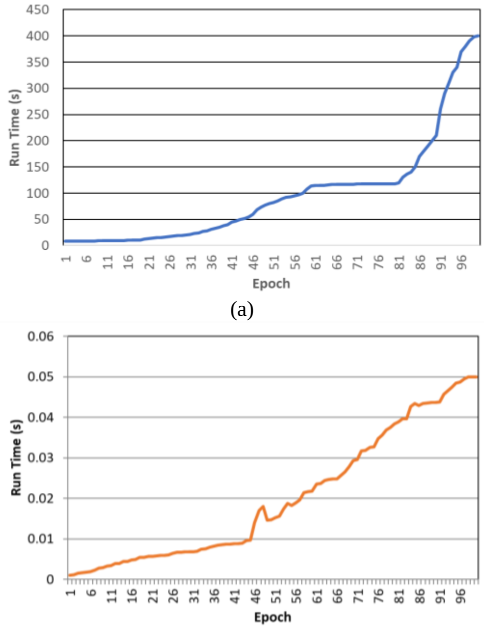


Fig. 9. SE comparison of the proposed DDPG and conventional systems for User 1.

C. Complexity Analysis of the Proposed DDPG

Convex Optimization (CVX) is a method of solving optimization problems, an algorithm to find the solution to a problem. In this paper, we use the CVX algorithm to compare the complexity of the proposed DDPG algorithm with CVX. This algorithm was chosen because it is suitable for finding an optimal solution to a problem. This optimization is the process of maximizing or minimizing the objective function while paying attention to existing value constraints. This optimization plays an important role in the design of a system. Through this optimization, the system can achieve high throughput with fewer inputs and reduce the time required to process data.



(a)

(b)

Fig. 10. Complexity of (a) convex optimization, and (b) proposed DDPG system.

To see how much this DDPG system can solve complex problems, the CVX algorithm has been simulated to compare the optimization process with the proposed DDPG model. The comparison of the complexity of the DDPG algorithm with CVX can be seen in Fig. 10. The more complex the data executed by the system; the response time required by the system is also much longer. As shown in Fig. 10, it can be concluded that the DDPG algorithm can solve the problem 8000 times faster in terms of execution time than the CVX algorithm. This demonstrates the effectiveness of the DDPG algorithm in solving more complex problems, such as massive MIMO systems.

The results demonstrate significant improvements in SE and SINR using the DDPG algorithm compared to conventional methods. However, one limitation we identified is the relatively high training time required for the DDPG model, particularly when scaling to larger numbers of antennas and users in the massive MIMO system. This suggests that further research is needed to explore ways to reduce the computational burden, such as optimizing the neural network architecture or using distributed learning techniques to parallelize the training process. Additionally, while the model performs well under simulated channel conditions, real-world implementation may present challenges, including variability in channel state information and hardware constraints. Future work could focus on testing the model in more dynamic environments or developing hybrid models that integrate DDPG with traditional optimization techniques to improve robustness and scalability. Implementing DDPG on hardware platforms presents significant computational challenges, particularly when processing high-dimensional data in real time. This can be difficult due to limited computational resources and the necessity for efficient algorithms that can function effectively on edge devices or within network constraints.

V. Conclusion

This paper has analyzed Spectral Efficiency (SE) based on the Deep Deterministic Policy Gradient (DDPG) method in the 5G massive MIMO communication system. The use of the DDPG-based DRL method aims to simplify the SE system performance analysis stages, which go through several stages such as training and testing. The testing of the DDPG method resulted in an average SINR value of 19.5 dB, demonstrating high accuracy. This SINR value is used for system testing with the proposed DDPG method to obtain a higher SE value compared to conventional methods. This is because the DDPG system can study the condition of the information delivery channel until it reaches the user. The DDPG system can also overcome complexity problems in 5G massive MIMO communication systems with the basic reference being the short data processing time required. This is because the DDPG method always updates every DDPG actor interacting with the dataset. Therefore, the DDPG system can solve complex problems and provide optimal solutions. The DDPG algorithm significantly enhances SE and SINR in massive MIMO systems, while addressing computational challenges, but further research is needed to optimize training time and test real-world performance. Exploring the application of DDPG in other areas, such as energy-efficient communication or multi-objective optimization in wireless networks, could extend its utility beyond the current context.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors conducted the research; the first author designed the system model and simulation, and all authors analyzed the simulation data. The first author wrote the paper. All authors reviewed the manuscript.

Funding

This work is supported by the professorship research scheme, Universitas Syiah Kuala, under grant no: 41/UN11.2.1/PT.01.03/PNBP/2023.

Acknowledgment

The authors would like to thank the Lembaga Penelitian dan Pengabdian Masyarakat, Universitas Syiah Kuala, Banda Aceh, Indonesia for supporting this work.

References

[1] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.

[2] A. Gupta and R. K. Jha, "A survey of 5G network: Architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.

[3] M. J. Shehab, I. Kassem, A. A. Kutty, M. Kucukvar, N. Onat, and T. Khattab, "5G networks towards smart and sustainable cities: A review of recent developments, applications and future perspectives," *IEEE Access*, vol. 10, pp. 2987–3006, 2022.

[4] P. Salva-Garcia, J. M. Alcaraz-Calero, R. M. Alaez, E. Chirivella-Perez, J. Nightingale, and Q. Wang, "5G-UHD: Design, prototyping and empirical evaluation of adaptive Ultra-High-Definition video streaming based on scalable H.265 in virtualised 5G networks," *Comput. Commun.*, vol. 118, pp. 171–184, Mar. 2018.

[5] I. F. Akyildiz, S. Nie, S.-C. Lin, and M. Chandrasekaran, "5G roadmap: 10 key enabling technologies," *Comput. Netw.*, vol. 106, pp. 17–48, Sep. 2016.

[6] M. A. Mohamed, H. A. Hassan, M. H. Essai, H. Esmaiel, A. S. Mubarak, and O. A. Omer, "Deep learning-based SC-FDMA channel equalization," *International Journal of Electrical and Electronic Engineering & Telecommunications*, vol. 13, no. 1, pp. 67–79, 2024.

[7] S. A. Busari, K. M. S. Huq, S. Mumtaz, L. Dai, and J. Rodriguez, "Millimeter-wave massive MIMO communication for future wireless systems: A survey," *IEEE Commun. Surv. Tutor.*, vol. 20, no. 2, pp. 836–869, 2018.

[8] J. Guo, C.-K. Wen, S. Jin, and G. Y. Li, "Overview of deep learning-based CSI feedback in massive MIMO systems," *IEEE Trans. Commun.*, vol. 70, no. 12, pp. 8017–8045, Dec. 2022.

[9] A. Ly and Y.-D. Yao, "A review of deep learning in 5G research: channel coding, massive MIMO, multiple access, resource allocation, and network security," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 396–408, 2021.

[10] Y. Xin, D. Wang, J. Li, H. Zhu, J. Wang, and X. You, "Area spectral efficiency and area energy efficiency of massive MIMO cellular systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3243–3254, May 2016.

[11] A. He, L. Wang, Y. Chen, K.-K. Wong, and M. Elkashlan, "Spectral and energy efficiency of uplink D2D underlaid massive MIMO cellular networks," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3780–3793, Sep. 2017.

[12] X. Zhang, H. Qi, X. Zhang, and L. Han, "Spectral efficiency improvement and power control optimization of massive MIMO networks," *IEEE Access*, vol. 9, pp. 11523–11532, 2021.

[13] H. T. P. D. Silva, H. S. Silva, M. S. Alencar, W. J. L. D. Queiroz and U. S. Dias, "Energy and spectral efficiencies of cell-free millimeter-wave massive MIMO systems under rain attenuation based on ray tracing simulations," *IEEE Access*, vol. 11, pp. 26979–26995, 2023.

[14] S. Panda, "Spectral efficiency optimization of massive MIMO system under channel varying conditions," *Wirel. Pers. Commun.*, vol. 117, no. 2, pp. 1319–1335, Mar. 2021.

[15] T. A. Sheikh, J. Bora, and M. A. Hussain, "Massive MIMO system lower bound spectral efficiency analysis with precoding and perfect CSI," *Digit. Commun. Netw.*, vol. 7, no. 3, pp. 342–351, Aug. 2021.

[16] B. K. Gül and N. Taşpınar, "Spectral and energy efficiency trade-off in massive MIMO systems using multi-objective bat algorithm," *J. Electr. Eng.*, vol. 73, no. 2, pp. 132–139, Apr. 2022.

[17] L. You, J. Xiong, A. Zappone, W. Wang and X. Gao, "Spectral efficiency and energy efficiency tradeoff in massive MIMO downlink transmission with statistical CSIT," *IEEE Tran. on Signal Processing*, vol. 68, pp. 2645–2659, Apr. 2020.

[18] W. Tan, S. Li, and M. Zhou, "Spectral and energy efficiency for uplink massive MIMO systems with mixed-ADC architecture," *Phys. Commun.*, vol. 50, 101516, Feb. 2022.

[19] B. Jiang, B. Ren, Y. Huang, T. Chen, L. You, and W. Wang, "Energy efficiency and spectral efficiency tradeoff in massive MIMO multicast transmission with statistical CSI," *Entropy*, vol. 22, no. 9, 1045, Sep. 2020.

[20] Q. Cai, C. Cui, Y. Xiong, W. Wang, Z. Xie, and M. Zhang, "A survey on deep reinforcement learning for data processing and analytics," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 5, pp. 4446–4465, May 2023.

[21] M. Naeem, S. T. H. Rizvi, and A. Coronato, "A gentle introduction to reinforcement learning and its application in different fields," *IEEE Access*, vol. 8, pp. 209320–209344, 2020.

[22] A. Iqbal, M.-L. Tham, and Y. C. Chang, "Energy- and spectral-efficient optimization in cloud RAN based on dueling double deep network," in *Proc. 2021 IEEE Int. Conf. on Automatic Control & Intelligent Systems (I2CACIS)*, Jun. 2021, pp. 311–316.

[23] Y. Zhao, I. G. Niemegeers, and S. M. H. De Groot, "Dynamic power allocation for cell-free massive MIMO: deep reinforcement learning methods," *IEEE Access*, vol. 9, pp. 102953–102965, 2021.

[24] T. P., Lillicrap, J. J. Hunt, A. Pritzel *et al*., "Continuous control with deep reinforcement learning," in *Proc. of 4th Int. Conf. on Learning Representations (ICLR)*, May 2016, pp. 1–14.

[25] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press: England, 2005.

[26] G. Zheng, K. K Zhang, and E. Bj örnson, "Improving the scalability of convex optimization-based precoding for massive MIMO systems," *IEEE Trans. on Signal Processing*, vol. 65, no. 15, pp. 4060–4075, 2017.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.

[28] V. Mnih, K. Kavukcuoglu, D. Silver *et al*., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[29] M. Zhang, J. Gao, and C. Zhong, "A deep learning-based framework for low complexity multiuser MIMO precoding design," *IEEE Trans. Wirel. Commun.*, vol. 21, no. 12, pp. 11193–11206, Dec. 2022.

[30] M. Khurana, "Deep learning based low complexity joint antenna selection scheme for MIMO vehicular adhoc networks," *Expert Syst. Appl.*, vol. 219, 119637, Jun. 2023.

[31] S. Kumar, A. Singh, and R. Mahapatra, "DLNet: Deep learning-aided massive MIMO decoder," *AEU - Int. J. Electron. Commun.*, vol. 155, 154350, Oct. 2022.

[32] E. Bjornson and P. Giselsson, "Two applications of deep learning in the physical layer of communication systems [lecture notes]," *IEEE Signal Process. Mag.*, vol. 37, no. 5, pp. 134–140, Sep. 2020.

[33] M. Bashar A. Akbari, K. Cumanan *et al*., "Exploiting deep learning in limited-fronthaul cell-free massive MIMO uplink," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1678–1697, Aug. 2020.

[34] L. V. Nguyen, N. T. Nguyen, N. H. Tran, M. Juntti, A. L. Swindlehurst, and D. H. N. Nguyen, "Leveraging deep neural networks for massive MIMO data detection," *IEEE Wirel. Commun.*, vol. 30, no. 1, pp. 174–180, Feb. 2023.

[35] S. Yu and J. W. Lee, "Deep reinforcement learning based resource allocation for D2D communications underlay cellular networks," *Sensors*, vol. 22, no. 23, 9459, Jan. 2022.

[36] L. Luo, J. Zhang, S. Chen, X. Zhang, B. Ai, and D. W. K. Ng, "Downlink power control for cell-free massive MIMO with deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6772–6777, Jun. 2022.

[37] T. Tiong, I. Saad, K. T. K. Teo, and H. bin Lago, "Deep reinforcement learning with robust deep deterministic policy gradient," in *Proc. 2020 2nd Int. Conf. on Electrical, Control and Instrumentation Engineering (ICECIE)*, Nov. 2020. doi: 10.1109/ICECIE50279.2020.9309539

[38] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 1st ed., Cambridge, USA: MIT Press, 1998.

Nasaruddin Nasaruddin received his B.Eng. degree in electrical engineering from the Sepuluh Nopember Institute of Technology, Surabaya, Indonesia in 1997. Subsequently, he received his M.Eng. and D.Eng. degrees in physical electronics and Informatics, from the Graduate School of Engineering, Osaka City University, Japan, in 2006 and 2009, respectively. He is a full professor at the Electrical Engineering Department, Syiah Kuala University. Currently, he is the director of the Directorate of Education and Learning, Syiah Kuala University. His research interests include digital communications, information theory, and deep learning applications, in addition to computer and communication networks. He has published many papers in cooperative communication networks, communication technologies, and deep learning applications. He is a member of IEEE, IEEE Systems, Man, and Cybernetics Society, and ACM.

Afzal Riski received his B.Eng. degree in electrical engineering from Universitas Syia Kuala, Banda Aceh, Indonesia in 2023. He completed his undergraduate studies by conducting research on deep learning in MIMO communication systems. He is currently a Health, Safety and Environment (HSE) officer at PT Cijantung Anugrah Sukses Mandiri which is an experienced business entity working on national projects located in East Jakarta.

Yunida Yunida received her B.Eng. degree in electrical engineering from Universitas Syiah Kuala, Banda Aceh, Indonesia in 2013. Then she received her Ph.D. degree in electrical and computer engineering from Universitas Syiah Kuala in 2020 through the "Magister Program of Education Leading to Doctoral for Excellent Graduates (PMDSU)" scholarship from the Ministry of Research, Technology and Higher Education of the Republic of Indonesia. Since 2016, she has published about 7 articles, of which are 4 articles in Scopus Indexed Journals and 3 articles in the Proceedings of the IEEE. She is currently a lecturer in the Electrical and Computer Engineering Department at Universitas Syiah Kuala. Her research interests include digital communications, wireless communications, and information theory.

Ramzi Adriman was born in Banda Aceh, on January 30, 1979. He graduated with a bachelor's degree in electrical engineering Department of Universitas Syiah Kuala in 2003. In 2009 he completed his master's program in computer science and information engineering, at the University of Asia-Taiwan. Then in 2015 completed a doctoral degree in computer science and information engineering, at the University of Asia-Taiwan, since 2005 he served as a lecturer at the Faculty of Engineering, Department of Electrical and Computer Engineering Universitas Syiah Kuala. He has research focused on optimization, algorithms, multi-agent systems, network security, the Internet of Things, and intelligent transportation systems.