Deep Reinforcement Learning Based Voltage Controls for Power Systems under Disturbances

by

Yuling Wang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved January 2024 by the
Graduate Supervisory Committee:

Vijay Vittal, Chair
Raja Ayyanar
Anamitra Pal
Mojdeh Hedman

ARIZONA STATE UNIVERSITY

May 2024

ABSTRACT

In recent years, there has been an increasing need for effective voltage controls in power systems due to the growing complexity and dynamic nature of practical power grid operations. Deep reinforcement learning (DRL) techniques now have been widely explored and applied to various electric power operation analyses under different control structures. With massive data available from phasor measurement units (PMU), it is possible to explore the application of DRL to ensure that electricity is delivered reliably. For steady-state power system voltage regulation and control, this study proposed a novel deep reinforcement learning (DRL) based method to provide voltage control that can quickly remedy voltage violations under different operating conditions. Multiple types of devices, adjustable voltage ratio (AVR) and switched shunts, are considered as controlled devices. A modified deep deterministic policy gradient (DDPG) algorithm is applied to accommodate both the continuous and discrete control action spaces of different devices. A case study conducted on the WECC 240-Bus system validates the effectiveness of the proposed method. System dynamic stability and performance after serious disturbances using DRL are further discussed in this study. A real-time voltage control method is proposed based on DRL, which continuously regulates the excitation system in response to system disturbances. Dynamic performance is considered by incorporating historical voltage data, voltage rate of change, voltage deviation, and regulation amount. A versatile transmission-level power system dynamic training and simulation platform is developed by integrating the simulation software PSS/E and a user-written DRL agent code developed in Python. The platform developed facilitates the training and testing of various power system algorithms and power grids in dynamic simulations with all the modeling capabilities available within PSS/E. The efficacy of the proposed method is evaluated based on the developed platform. To enhance the controller's resilience in addressing communication failures, a dynamic voltage control method employing the Multi-agent DDPG algorithm is proposed. The algorithm follows the principle of centralized training and decentralized execution. Each agent has independent actor neural networks and critic neural networks. Simulation outcomes underscore the method's efficacy, showcasing its capability in providing voltage support and handling communication failures among agents.

ACKNOWLEDGMENTS

I want to express my heartfelt appreciation to my advisor, Dr. Vijay Vittal, for his indispensable support and guidance throughout my research journey. He is an esteemed professor, consistently providing invaluable help throughout my entire PhD study and research. His expertise, patience, and encouragement have significantly contributed to the success of my doctoral work.

I also want to acknowledge and thank my committee members—Dr. Raja Ayyanar, Dr. Anamitra Pal, and Dr. Mojdeh Hedman—for their insightful feedback and valuable contributions to my research. Their expertise has played a crucial role in enhancing the overall quality of my work.

Finally, I extend profound thanks to my parents, family members, and friends for their unwavering love, support, and understanding. Their consistent encouragement and belief in me have served as a continuous wellspring of motivation and inspiration throughout my doctoral journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Chapter 1

INTRODUCTION

1.1 Background

Power system resilience [2; 3] and reliability are vital to the economic viability of society. The US-Canada power system outage on August 14, 2003 cost 10 billion US dollars [4]. More and more essential services, such as electrical transportation, rely on electricity, so it is of great importance to guarantee power system stability and dynamic performance [5; 6].

With the increasing integration of utility-scale renewable energy and distributed energy resources [7; 8; 9; 10; 11], such as wind and solar, the power system variability has further increased due to the nonlinearity and unpredictable consumer patterns of these new types of resources and loads, which reduces the system inertia and leads to faster dynamics [12; 13; 14]. More loads interfaced with the system through electronics converters and the growing capacity of the High-Voltage Direct Current (HVDC) system also contribute to the complexity of the power grid [15; 16; 17; 18].

The system operation requires real-time monitoring and control to respond to unexpected dynamic changes at both the demand and supply sides [19; 20]. When the system undergoes a large disturbance, such as an abrupt change in load or generation, grid operators are often faced with the challenge of maintaining the system bus voltage magnitudes within secure ranges and can be overwhelmed with the task of dispatching generation to maintain the power balance and relieve transmission congestion, leaving them insufficient bandwidth to attend to voltage violations. In addition, the increased energy exchange requirement has impacted power system security [21], resulting in the system operating closer to its security limits in some instances. These factors enhance the chances of power system instability and pose severe challenges to real-time voltage control [4]. Advanced control techniques are needed to ensure that electricity is transmitted and delivered reliably and avoid negative economic and societal results.

Phasor measurement units (PMUs) [22], which work as communication and measurement devices, make it possible to transfer synchronized dynamic data across power systems. The advanced communication infrastructure in power systems, computation structure, and power system devices provide the possibility for the implementation of the advanced control methods. Hence, online stability prediction [23] and corrective control can be achieved [24].

Artificial intelligence (AI) [25; 26; 27] techniques have matured and are being applied to various power system applications, representing a significant advancement in how we manage and optimize power grids. AI plays an important role in addressing the system operation and control challenges by enhancing the power system performance, resilience, and robustness of system stability. This data-driven technology opens possibilities to design power system control by learning and updating the control action and policies.

1.2 The Development of Artificial Intelligence Implementation in Power Systems

The concept of AI was first proposed in the 1950s to 1960s when researchers were trying to mimic human intelligence [25]. Early machine learning (ML) [28] arose and mainly focused on theories like perceptron and decision trees to implement into rule-based systems. In the 1970s to 1980s, challenges arose in AI research, leading to a heightened focus on expert systems, which were used in the fields of medicine and finance. Later in the 1990s to 2000s, AI technology underwent a resurgence driven by advancements in computational capabilities and the increased availability of data, which contributed to the development of machine learning. The Supervisory Control and Data Acquisition (SCADA) [29] systems began to be implemented by power utilities for power grid remote monitoring and control. These systems established the groundwork for the digitization of data within power systems.

After the 2000s, various algorithms and neural networks became more advanced and widely used, ushering in a new era of data-driven across different domains, from medicine and finance to manufacturing and transportation. Deep learning [30], a subset within the realm of machine learning, started to gain recognition as neural networks developed and the availability of Graphics Processing Units (GPUs) enabled faster computational processing. Meanwhile, ML was applied to power systems in predicting equipment failures and recommending maintenance actions [31].

With more implementation of ML to power systems, utilities began building data management systems to store and preprocess the increasing data. Then, ML algorithms were used to identify power system unusual operation patterns or faults in real-time data, including dynamic state estimation and event detection. As PMUs that offers synchronized measurements of electrical quantities were integrated into power grids, high-resolution data was available, and wide-area monitoring and situational awareness were further analyzed based on ML models.

With the growth and expansion of the power grid, the penetration of renewable energy sources is increasing[32]. ML models were utilized to address the challenges in resource integration, providing innovative schemes to predict the pattern of renewable energy generation, optimize energy storage, and enhance grid stability. ML models play a significant role in demand response, which provides information to utilities and customers in response to supply energy and pricing fluctuations. The distribution of electricity within a smart grid can be optimized by ML algorithms, and energy efficiency is enhanced.

Reinforcement Learning (RL) [33] has emerged to be a powerful tool for power grid operation and control. The RL agents are able to make decisions based on real-time data and power system feedback and can autonomously adjust various grid parameters to ensure grid normal operation, minimize energy losses, and enhance system efficiency. Deep Reinforcement Learning (DRL) [34] represents a state-of-the-art method that combines RL with deep neural networks, enabling agents to manage more complex grids and challenging scenarios. The integration of DRL is shifting the approaches to manage the electrical grids.

1.3 Previous Research on Power System Voltage Control

In recent years, the demand for robust and reliable voltage control methods in power systems has surged considerably due to the growing complexity and dynamic nature of practical power grid operations.

Early approaches to regulating the voltages have mainly relied on utility-owned devices in power systems, such as transformers equipped with tap changers [35], shunt reactors and shunt capacitors [36], automatic voltage regulators (AVRs) [37], static var compensators (SVCs) [38],and flexible alternating current transmission system (FACTS) devices [39]. On-load tap changers (OLTCs) are typically built-in high-voltage power transformers to adjust the transformer turns ratio. The OLTCsc can be adjusted manually or automatically to regulate the output voltage to maintain it within the desired range. Adjustable-voltage-ratio(AVR) transformers have been reported in [40], where the voltage ratios can be adjusted quickly and continuously using magnetic flux valve(MFV) based characteristics. Achieving the optimal configuration for the device typically involves tackling mixed-integer programs, generally known to be NP-hard. In [41] and [42], a semidefinite relaxation heuristic was employed to optimize the tap positions. Control rules founded on heuristics were formulated in [43]] and [44]. Nonetheless, it is worth noting that these methods may entail substantial computational requirements and may not ensure optimal performance.

Smart power inverters, like PV and wind turbines, are equipped with integrated computing and communication modules. These modules can be instructed to modify their reactive power output. Determining the optimal settings for controlling the inverters' reactive power output is a non-convex optimal power flow problem, as discussed in [45]. To address challenges related to renewable energy variability and communication obstacles, such as delays and packet loss, there have been growing studies on stochastic, online, decentralized, and localized reactive control approaches [46; 45; 47; 48; 49; 50; 51].

Excitation system control is of significant importance in maintaining generators' voltages and can impact power system dynamic stability directly[52]. Excitation control is considered to be one of the most economical and effective methods for maintaining voltage and dynamic performance enhancement[53]. Numerous excitation control methods have been conducted in terms of voltage regulation considering system dynamic stability after disturbances. A decentralized nonlinear voltage controller is proposed in [54] to achieve both voltage regulation and system stability improvement. Global control(GC) where a stable controller is used for the fault period and a voltage controller is activated for voltage level regulation in [55]. Different controllers need to be switched at different operating stages to guarantee a satisfactory voltage level and system dynamic performance. Lyapunov-function-based methods can achieve voltage regulation and dynamic stability control simultaneously by designing the excitation control without switching[56]. A Lyapunov-based decentralized control (LBC) is proposed in [57] to enhance power system dynamic performance by simultaneously controlling the excitation and governor systems. The time-derivative of the

Lyapunov function is designed by the feedback control of synchronous generators, and voltage deviation is considered as the feedback variable to realize voltage regulation as well as dynamic performance improvement. The majority of these model-based methods have been claimed to achieve promising performance. However, they rely heavily on accurate information of power system topology and parameters. Furthermore, power systems are experiencing uncertainties of load changes and contingencies and it is quite challenging to apply the above model-based methods. Therefore, a voltage regulation method that is flexible and scalable to the application and operational uncertainties needs to be developed.

1.4 Previous Research on DRL-based Power System Steady State Control

Artificial intelligence (AI) techniques are now being applied to various power system applications in order to solve control or data-related problems[58]. These early study efforts include [59] on reactive power and voltage control, [60] on power system stability control, [61] on load-frequency control, and [62; 63; 64] on the electricity price prediction. These data-driven, model-free methods are particularly well-suited for highly non-linear and high-dimensional power systems, especially with the availability of phasor measurement units (PMUs) that enable the synchronized transfer of dynamic data across the grid.

Advanced control schemes for enhancing power system stability based on AI methods have been developed, and the recent success of reinforcement learning (RL) has shown promise in addressing various power system challenges. An RL agent can be trained to respond instantaneously to a range of system operating conditions based on knowledge obtained by interacting with the power system environment during the training process. Therefore, a real-time application based on RL is possible.

Q-learning, a conventional RL method, has been utilized in [65] and [66] to learn a reactive power optimal control scheme and keep the voltage within the normal range. Reference [67] proposed a fully automated energy management system (EMS) algorithm based on RL, which learns how to make optimal decisions for consumers. A novel EMS formulation based on a request inventory model using Q-learning is proposed in [68]. It balances energy cost and the delay in energy usage in the same way that the consumer would, but without the consumer having to make the decision. Q-learning was also adopted in [69] for optimal tap setting of on-load tap changers of step-down transformers (connecting electric distribution systems with the rest of the system) to control the distribution system voltages under uncertain load dynamics. Reference[70] proposed a control scheme of active power generations to prevent system cascading failure based on Q-learning. The controller operates in the system's normal state and takes actions in the form of preventive control to make adjustments in case of cascading failure when the system suffers large disturbances.

However, conventional RL methods only work in environments with discrete and finite state and action spaces and thus are not suitable for large, complex problems, such as real-time control problems for large-scale power systems. To overcome this disadvantage, deep reinforcement learning (DRL) has been developed by researchers, which utilizes powerful deep neural networks as function approximators that enable high-dimensional feature extraction. Reference [71] proposed a two-time-scale voltage control scheme, including fast inverter control and switching of shunt capacitors at a slower time control based on the Deep Q-Network (DQN) algorithm. Reference [58] applied DQN and Deep Deterministic Policy Gradient (DDPG) for subsystem voltage control and found that DDPG performed better with sufficient training scenarios. Reference [59] adopted multi-agent deep deterministic policy gradient (MADDPG), which is a multi-agent continuous actor-critic-based algorithm, to realize voltage regulation among different regional zones based on power flow data. These works focused on the steady-state performance of the system,

1.5 Previous Research on DRL-based Power System Dynamic Control

There have been many explorations and attempts in the area of steady-state voltage control based on reinforcement learning, however, ignoring the influence of the dynamic behaviors in the transient process when subjected to a disturbance. Reference [72] proposed the scheme of two coordinated wide-area damping controllers (CWADCs) for damping low-frequency oscillations (LFOs) based on DRL. While it learns by a pre-prepared data set and does not realize on-line training and implementation. References [73] and [74] addressed transient stability issues to keep the system in synchronism by controlling power system components, such as wind turbines and generators. Approximate Dynamic Programming (ADP) is used in [73] to optimize the closed-loop performance of a wind-integrated power grid by providing supplementary damping control. Another study[74] proposed a wide-area control architecture that includes a local supervised PSS control and an RL-based global wide-area control, which ensures coherent damping of local and inter-area oscillations using a priority scheme. In [75], the authors used DRL methods to implement dynamic braking and under-voltage load shedding for power system emergency control. While these methods have been tested on the IEEE 39-bus system or the 68-bus system, practical regional power grids are larger and more complex, which need significant information exchange between DRL agents and the power grid environment, especially considering the dynamic performance and real-time control application.

1.6 Research objectives and Aims

To address the challenges discussed above, a data-driven control framework with multiple types of control devices is proposed to support system voltage regulation following disturbances. Two types of control devices are considered - transformers capable of continuously adjusting voltage ratios and shunt capacitors taking discrete switching actions. Although transformer taps are generally adjusted in a slow and discrete manner, the quick and continuous change of voltage ratios can be achieved using the magnetic flux valve (MFV)[40]. A modified DDPG algorithm is applied to accommodate both the continuous and discrete controls of different devices while maintaining the ability to provide control in a large action and state space. During each training period, a reward function is defined to evaluate the effectiveness of the control actions—the ratio of the controlled transformers and the group size of the switched shunts. Compared with past studies, this work has developed a DRL framework where both the continuous and discrete controls collaborate to conduct power system voltage regulation. In this framework, a well-trained agent can control multiple equipment instantaneously under different operating conditions and provide quick and effective operational assistance when voltage violations occur.

For power system dynamic control, this dissertation aims to propose a real-time voltage control framework that continuously regulates the excitation system based on DRL. The dynamic performance attributes are considered to include dynamic stability factors that may influence power system operation in practical power grids. The voltage control function is achieved by adjusting the generators' excitation system under system disturbances. The DDPG algorithm, which deals with continuous action spaces, is used in this report to continuously control the voltage reference of the generator excitation system. To focus on the dynamic process, a transmission-level dynamic power system training and simulation platform is built based on the commercial power system software package PSS/E and user-written code in Python. By using DRL, this study proposes a controller that allows generators to change their reactive power output within specified limits in real-time, enabling the system to satisfy operational requirements and provide voltage support in response to disturbances or load changes.

1.7 Main Contribution of this work

This work aims to address some of the key issues identified in the current literature. The main contribution of this study is:

1. A data-driven control framework with multiple types of control devices is proposed to support system voltage regulation following a disturbance. Two types of control devices are considered - transformers capable of continuously adjusting voltage ratios and shunt capacitors taking discrete switching actions.

2. A modified DDPG algorithm is applied to accommodate both the continuous and discrete controls of different devices while maintaining the ability to provide control in a large action and state space.

3. A DRL framework where both the continuous and discrete controls collaborate to conduct power system voltage regulation is developed.

4. A novel real-time voltage control method based on DRL is proposed, which not only regulates and controls the voltage but also considers the dynamic performance of the power system after the control implementation. By leveraging DRL algorithms, the proposed method achieves improved dynamic performance, addressing the challenges of practical power grids characterized by large size, complexity, and real-time control requirements.

5. A transmission level power system dynamic training and testing platform is built in this study using a combination of a commercial power system software package PSS/E and a user-written DRL agent code developed in Python. This

platform provides a versatile environment that enables the training and testing of various power system algorithms in different power grid environments. The platform supports different scenarios that enable the simulation of various system conditions.

6. A large-scale power system is tested and verified based on the dynamic training and testing platform to investigate the control performance for large power grids. The platform's ability to handle large and complex dynamic power system environments further ensures the practicality and effectiveness of the tested methods in real-world scenarios.

7. A dynamic voltage control method employing the Multi-agent DDPG algorithm is proposed to enhance the controller's resilience in addressing communication failures. Centralized training and decentralized execution features of Multi-agent DDPG enable independent actor and critic neural networks for the controller. After being well trained, each agent possesses the capability to autonomously generate control commands utilizing only local information, which significantly improves the robustness of the control method.

1.8 Report Organization

The rest of the report is organized as follows:

- Chapter 2 gives a brief review of the concepts and mathematical formulations relevant to DRL. The first section includes the basic concepts of deep neural networks, different activation functions, and the relevant theory of neural network training. The theoretical background of DRL is introduced in section 2. Section 3 further introduces the theory and formulation of DDPG.

- Chapter 3 presents the DRL-based steady-state voltage control using multiple control devices. The formulations for building the multi-device voltage control problem into a Markov Decision Process are introduced first. Following this, the definitions of state space, action space, and reward functions are presented. The discretization for the action space of switched shunts to be implemented in the DDPG algorithm is discussed. The training platform and the data interaction between the DRL agent and the power system environment are presented. Finally, simulations are conducted for the result analysis.

- Chapter 4 discusses the DRL-based excitation control considering the system's dynamic performance. The power system dynamic operation control is first discussed. The DRL formulations, which include the state space, action space, and the design of the reward functions for the system dynamic voltage control, are further presented. The dynamic simulation platform is introduced in detail for the DRL agent training. The simulation results are analyzed to demonstrate the effectiveness of the proposed method.

Chapter 5 discusses a dynamic voltage control method employing the Multi-agent DDPG algorithm. The Markov Game theory is first introduced. The detailed algorithm of the Multi-agent DDPG algorithm is then discussed in detail. The design of action, state, and reward function are discussed and finally simulations based on different test systems are analyzed.

- Chapter 6 concludes the report and provides the potential for the future work of this study.

Chapter 2

DEEP REINFORCEMENT LEARNING THEORY BACKGROUND

2.1 Deep Neural Networks

2.1.1 Artificial Neural Networks

Artificial neurons are the basic function component or building nodes in a neural network, the mathematical model of which is inspired by the biological neurons found in the human brain. A neural network is a group of algorithms representing the underlying relationship among data similar to the brain. It can learn to perform tasks from examples of data. When the input changes, the neural networks are able to give the best result without redesigning the output procedure when neural networks are well-trained [76].

In a neural network, multiple inputs will be given, and a weighted sum of these inputs will be connected with an activation function to produce an output, as shown in Figure 2.1. Each input is associated with a weight, determining its importance in the computation. The weighted sum of inputs, often denoted as z, is computed as [76]:

$$z = \sum_{i=1}^{n} (w_i \cdot x_i) + b,$$

where w_i are weights, x_i are inputs, n is the number of inputs, and b is a bias term.

The structure of the neural network includes 3 types of layers [76]:

- **Input layer** — This layer refers to the first layer of nodes in the neural network and will receive the initial raw data that is input to the system. It passes the data directly to the hidden layer, where the data is multiplied by the first hidden

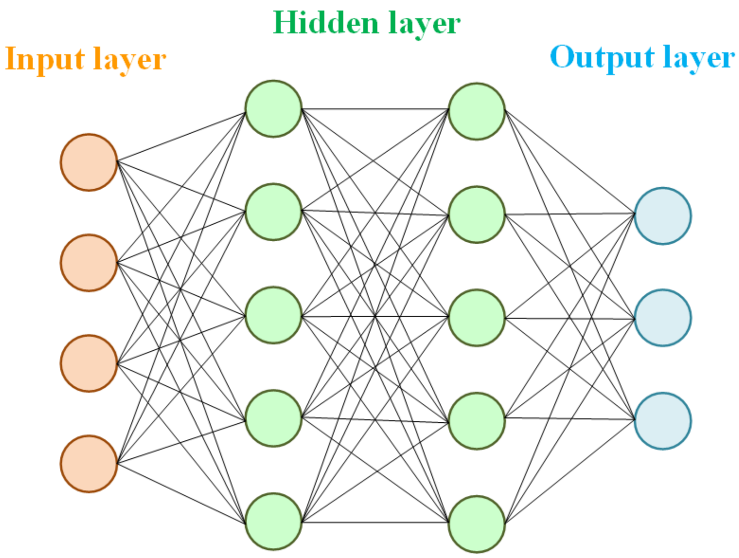

Figure 2.1: Structure of neural networks.

layer's weights.

- **Hidden layers** — Hidden layers are intermediate layers between the input and output layers, where all data processing is done. They are key components in the neural network to extract information and learn complex tasks.
- **Output layer** — The output layer inputs the processed data and produces the final result for neural networks.

2.1.2 Activation Functions

Under the above structure, the network represents a linear relationship between the input and the output even after applying a hidden layer. The activation function does the non-linear transformation to the input, making it capable of learning more information. The activation function will introduce non-linearity into the output of a neuron.

Various non-linear activations are in use, such as Sigmoid, ReLU(Rectified Linear

Unit), Tanh (Hyperbolic Tangent), Leaky ReLU, and Softmax [77].

The sigmoid function can output only positive values between 0 and 1 which is often used in the output layer for binary classification problems. The formulation can be described as [77]

$$f(z) = \frac{1}{1 + e^{-z}} \tag{2.1}$$

The function is plotted as an ’S’-shaped Curve, as shown in Figure 2.2. The small changes in z would bring about large changes in the value of $f(z)$ when z is around 0, so the predicted result would easily be 1 if the value is greater than 0.5 and 0 otherwise.

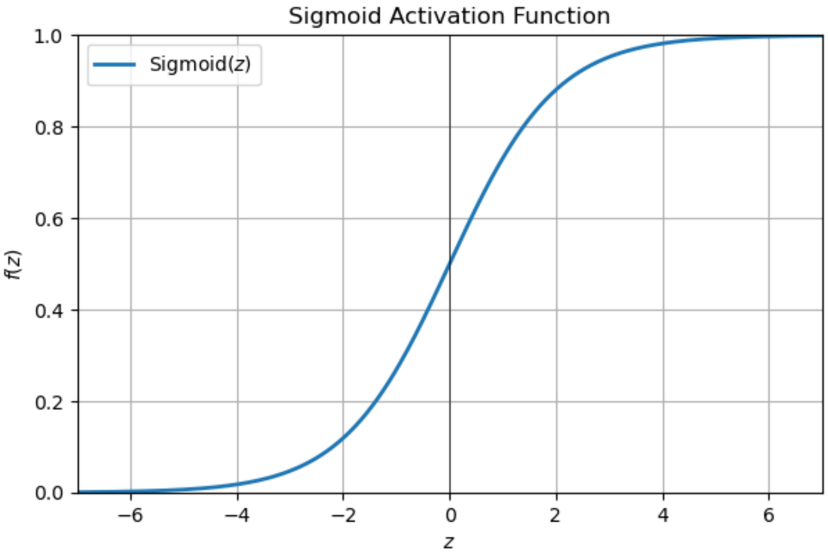

Figure 2.2: The Sigmoid activation function.

The ReLU function returns the value of the positive inputs and 0 for negative inputs, as shown in Figure 2.3. It is defined as [77]

$$f(z) = \max(0, z) \tag{2.2}$$

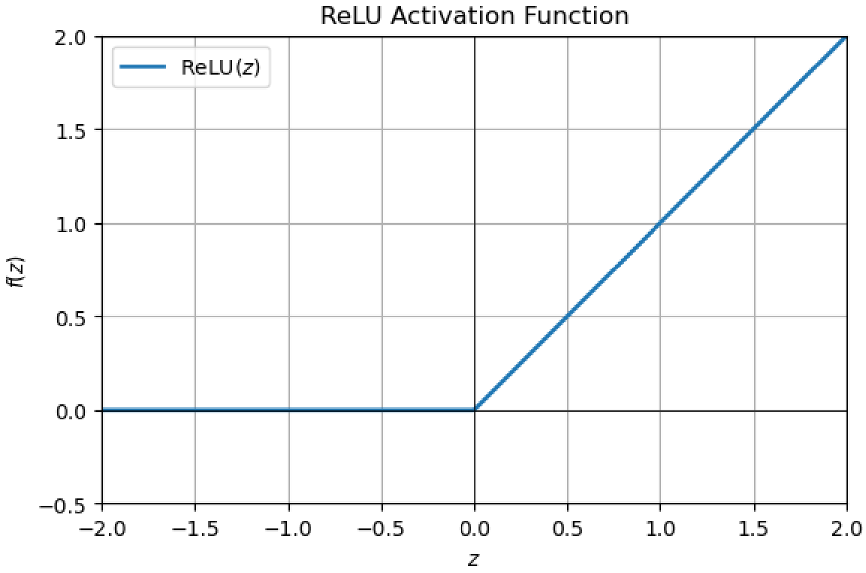

Figure 2.3: The ReLU activation function.

The ReLU function is the most widely used activation function and is usually used in the hidden layer of the neural network. Since it is simpler in mathematical structures, Relu learns faster due to its simplicity and effectiveness and does not saturate.

The Tanh function is described as [77]

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{2.3}$$

As shown in Figure 2.4, the Tanh function squashes the output between -1 and 1 and is usually used in hidden layers of a neural network. The mean of its output is 0 or very close to it, hence helping in centering the data by bringing the mean close to 0. It works better than Sigmoid and is mathematically shifted from the Sigmoid function.

The Leaky ReLU solves the 'dying ReLU' problem since the output is zero for all

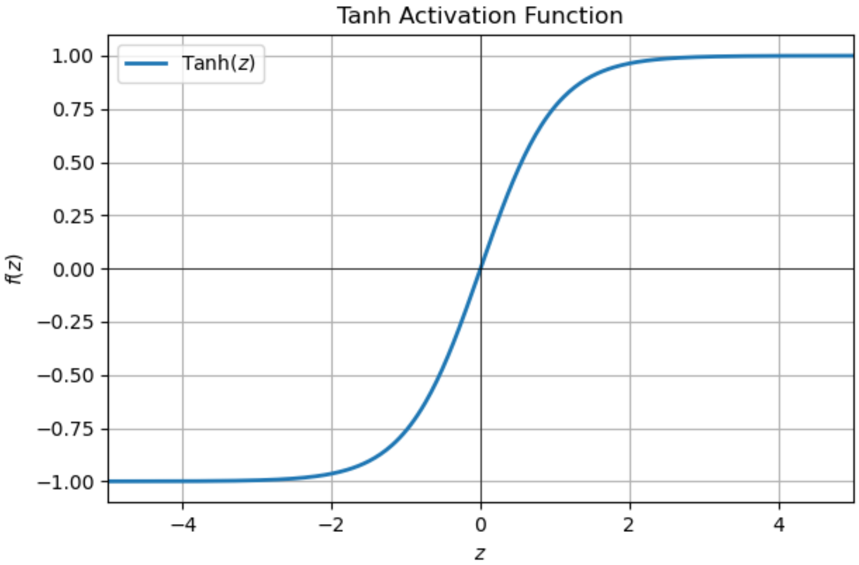

Figure 2.4: The Tanh activation function.

negative inputs in ReLU activation function. It is formulated as follows [77]

$$f(z) = \begin{cases} z & \text{if } z > 0 \\ \alpha z & \text{if } z \leq 0 (2.4) \end{cases}$$

where α is a small positive hyperparameter, which allows a small gradient for negative inputs for the Leaky ReLU activation function, the curve of the Leaky ReLU can be seen in Figure 2.5 where α is set as 0.01.

The Softmax function is a type of Sigmoid function that is used in the output layer for multi-class classification problems where probability distribution to define the class of each input is obtained [77].

Activation functions form the core components responsible for information processing and feature extraction within the network. They introduce non-linearity into neural networks, which allows them to learn intricate patterns among massive data and solve complex problems.

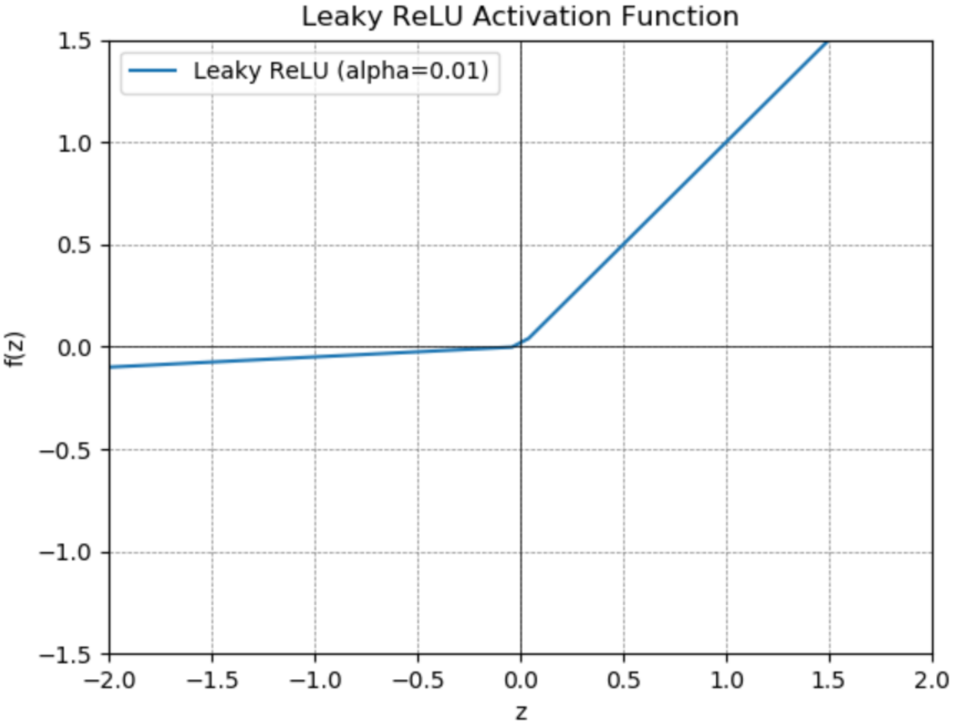

Figure 2.5: The Leaky ReLU activation function.

2.1.3 Feedforward and Backpropagation

Given a neural network architecture, the training process is to teach the network to make accurate predictions or classifications by updating its internal parameters (weights) based on the available training dataset.

The neural network training process starts from the initialization of the network's parameters (weights) in each layer with small random values. After initialization, the input data is propagated through the network's layers to generate predictions, which is called forward pass [76]. The forward pass computes the predictions by multiplying the weight vector with the given input vector through all layers and passing the sum of the product in every layer through the activation function.

The result of the forward pass is compared to the ground truth through loss calculation using loss functions. The loss function is utilized to quantify how different the network's predictions are from the ground truth values. The loss function can be a single function, such as mean square error, mean absolute percentage error, and cross-entropy, as well as a combination of several functions tuned by hyperparameters.

The foundational goal in neural network training is to minimize the loss function, achieved by backpropagation [76]. Backpropagation includes gradient computation, backward propagation, and parameter (weights) updates. The gradient of the loss with respect to the output of the last layer is computed using calculus and the chain rule. The computed gradient is then propagated backward through the network from the output layer to the input layer. In each layer, the gradient is adjusted based on the layer's parameters (weights) and the activation function's derivative. The adjusted gradients in each layer are utilized to update the corresponding parameters (weights) in that layer. The updates are based on the selection of optimization algorithms, such as gradient descent.

In each training iteration (epoch), forward pass, loss calculation, and backpropagation are performed sequentially to minimize the difference between prediction and the ground truth. The network's performance is periodically evaluated on a separate validation dataset to monitor progress and prevent overfitting. If the network performs well on the validation data, training can stop.

2.1.4 Gradient Descent

Gradient descent is an iterative optimization algorithm in training neural networks and minimizing the loss or cost functions. The model parameters are interactively adjusted in the direction of the steepest decrease in the cost function. Figure 2.6 shows the gradient descent iteration.

The iteration process is described as follows [76]:

- Calculate the gradient value of the cost function to the model parameters. The gradient is the direction and magnitude that the cost value decreases most.

- Subtracting a learning rate of the gradient from the value of the model parameters to update the parameters into the latest value.
- Repeat the above steps until the criterion is met. The criterion can be a predefined number of iterations or the cost function converging to a minimum.

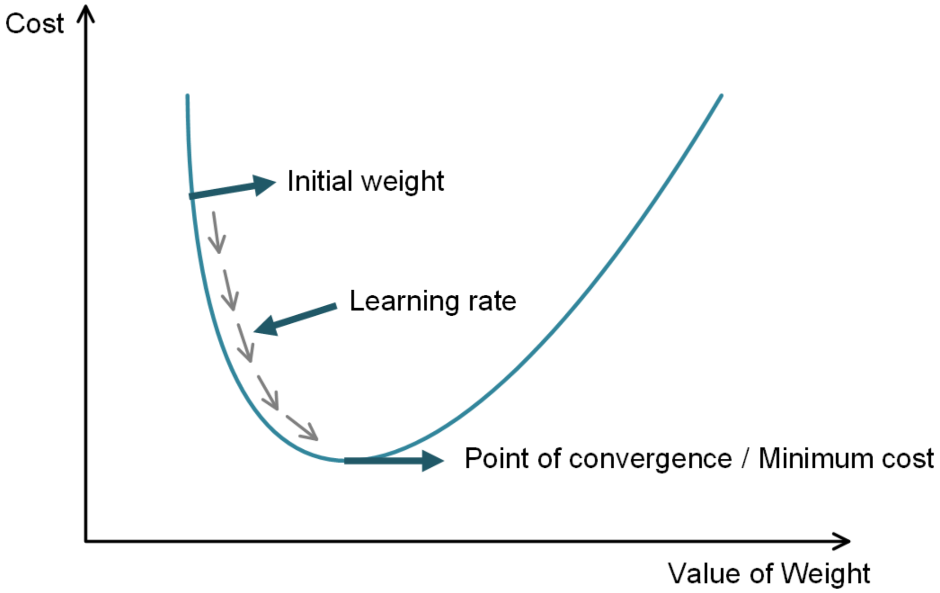

Figure 2.6: Diagram of gradient descent.

2.2 Deep Reinforcement Learning

Reinforcement learning is a subfield of artificial intelligence (AI) that deals specifically with training agents to make a sequence of decisions in dynamic environments to maximize a cumulative reward.

RL agent learns by interacting with the environment and making sequential decisions through a trial-and-error process. During the training process, the learned policy is continuously evaluated to guide the agent toward adjusting its control pol-

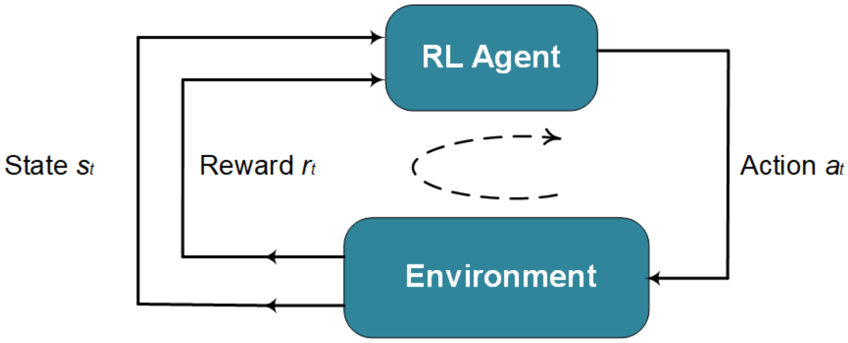

Figure 2.7: Interaction between RL agent and environment.

icy in the right direction. The RL agent aims to maximize the value of a reward function that is carefully designed to capture the objectives of the task. The agent explores different actions and extracts information about the state representations of the environment in real-time or through simulation to achieve this goal. If an action results in an increase in the reward value, the agent reinforces the trend of the action; otherwise, the action is attenuated. By adding various event scenarios to the data set, the RL agent can be fully trained to learn a behavior that yields maximum rewards.

The environment follows the Markov Decision Process (MDP). The formulation is defined as a finite MDP[78], M:

$$M \in (S, A, P, R, \gamma) \tag{2.5}$$

which includes a continuous or discrete state space S and action space A. The environment transition probability P maps a state-action pair at time t to a probability distribution over possible next states. A reward R is given for each state-action pair and a discount factor $\gamma \in [0, 1]$ is used to balance immediate and future rewards.

Figure 2.7 illustrates the interaction between the RL agent and the environment. At each step t, the agent observes the current state s_t from the environment and selects an action a_t based on its current policy. The agent obtains a reward r_t based on its action and state, and the environment transitions to a new state s_{t+1}. This process is repeated iteratively with the agent continuously updating its policy based on the observed states, actions, and rewards until a preset number of episodes is reached to end the training.

The agent aims to choose the optimal action given the current state to achieve the maximum accumulated discounted reward R_t over time:

$$R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i \tag{2.6}$$

where T is the time step. The key concept in searching for the optimal policy is evaluating the state-value function V and the action-value function Q, which is also known as the Q-function. The state-value function evaluates the goodness of a state for an agent under policy π, as shown in (2.7)

$$V^{\pi}(s) = E[R_t | s_t = s] \tag{2.7}$$

The Q-function $Q(s, a)$ represents the expected cumulative future discounted reward for an agent under policy π and estimates the value of performing a certain action a_t in a given state s_t:

$$Q^{\pi}(s_t, a_t) = E[R_t | s_t = s, a_t = a] \tag{2.8}$$

The Q-function is updated by the recursive relationship in the Bellman equation[79]:

$$Q_{t+1}(s, a) = E[R + \gamma max_{a'} Q_t(s^{'}, a^{'}) | s, a] \tag{2.9}$$

The Bellman equation will eventually converge to the optimal solution $Q^*(s, a)$ as the iterations proceed if the states follow the Markov property.

2.3 Deep Deterministic Policy Gradient Algorithm

DDPG is a reinforcement learning algorithm that is well-suited for continuous action spaces. It uses an actor-critic structure that concurrently learns a Q-function (modeled by the critic neural network) and a policy (modeled by the actor neural network). To improve the stability of the approach, DDPG utilizes a copied actor neural network and critic neural network to calculate the target values, which are periodically updated with the weights from the main neural networks to ensure consistency. In total, DDPG includes four networks to estimate the policy and value function: actor, target-actor, critic, and target-critic. Equation (2.10) is used to update the critic $Q(s, a)$ value.

$$Q_{j+1}^{(s,a)} = Q_j^{(s,a)} + \alpha[R_j + \gamma max Q_j^{(s^{'},a^{'})} - Q_j^{(s,a)}] \tag{2.10}$$

where α is the learning rate, γ is the discount rate, and $Q_j^{(s^{'},a^{'})}$ represents the target critic neural network.

The control action is obtained from the actor neural network, which enables DDPG to handle a continuous action space in a practical large-scale system. The actor neural network uses a parameterized actor function to determine a deterministic action based on the system states. During training, the policy π is updated in the direction suggested by the critic neural network to maximize the expected reward by taking steps in the direction of $\nabla_{\theta\mu} J$ with respect to the actor parameters. It is formulated as:

$$\nabla_{\theta\mu} J = \frac{1}{N} \sum \nabla_a Q(s, a)|_{s=s_j, a=\mu_{(s_j)}} \nabla_{\theta\mu} \mu(s|\theta^{\mu})|_{s=s_j} \tag{2.11}$$

where J is the starting distribution, $\mu(s|\theta^{\mu})$ is the parameterized actor function, and θ^{μ} is the policy neural network parameter.

The weights of the target neural networks are periodically updated using a soft update method: $\theta^{'} \leftarrow \rho\theta + (1 - \rho)\theta^{'}$, where ρ is a fraction weight that lies between 0 and 1.

During the action exploration, a decaying noise is added to the policy to improve the agent's ability to explore the range of actions available to solve the environment:

$$\mu^{'}(s_j) = \mu(s_j|\theta_j^{\mu}) + \xi_j \tag{2.12}$$

where $\xi_{j+1} = r_d * \xi_j$ and r_d is the decay rate.

Both the critic and actor are approximated with parameterized neural networks.

Chapter 3

DEEP REINFORCEMENT LEARNING BASED VOLTAGE CONTROL USING MULTIPLE CONTROL DEVICES

3.1 Problem Formulation for Multi-device Voltage Control

For voltage control, AVR transformers and switched shunts are considered controlled devices that provide reactive power support. The control objective is to find a policy that simultaneously determines the ratios of the transformers and the group size of the switched shunts that are in service to minimize the voltage deviation from the normal range. The states, actions, and rewards are defined below under the DDPG-based control framework.

3.1.1 Definition of States

Different measurements obtained by meters are usually used as the system states to represent the system's operating condition. Voltage magnitudes have been widely used for reactive power and voltage control problems, they are the direct indicators of the system conditions, and other electrical statuses can be somehow reflected in the voltage change[59], [80], [81]. This report also considers the bus voltage magnitudes as states.

3.1.2 Definition of Action Space

For voltage control using AVR transformers and switched shunts, the control actions are defined as a vector of the transformer ratios and the group size of the switched shunts. The DDPG algorithm considers a continuous action space, the continuous transformer ratio $a_{tf} = [a_{tf1}, a_{tf2}, ..., a_{tfn_{tf}}]^T$ can be directly controlled as part of the action. However, the switched shunts are controlled by the group which is discrete. In order to control multiple types of devices, the discretization for the actions of the switched shunt should be done for further implementation of the DDPG algorithm. Table 3.1 shows how the continuous actions of the switched shunts are discretized and implemented. A_l and A_u represent the lower and upper bounds of the continuous action space, L is the entire group of the switched shunts, g is the group value of the switched shunts in service. When actions $a_s = [a_{s1}, a_{s2}, ..., a_{sn_s}]^T$ of the switched shunts are generated by the DDPG agent, the values in the different ranges defined in Table 3.1 correspond to different groups that should be connected to the system. As a result, the total action space is formed by $a = [a_{tf1}, a_{tf2}, ..., a_{tfn_{tf}}, a_{s1}, a_{s2}, ..., a_{sn_s}]^T$, where n_{tf} and n_s are each the number of transformers and switched shunts under the control of the DDPG agent, respectively.

Table 3.1: Switched Shunt Action Discretization

a$_s$	Group in Service
$(A_l, A_l + \frac{A_u - A_l}{L})$	$g = 1$
(A$_l$ $+ \frac{A_u - A_l}{L}, A_l + 2 * \frac{A_u - A_l}{L})$	$g = 2$
...	...
$(A_l + (n-1) * \frac{A_u - A_l}{L}, A_{lo} + n * \frac{A_u - A_l}{L})$	$g = n$

3.1.3 Reward Function

The reward function r_t is designed to evaluate the effectiveness of the control actions when they are implemented. To restore the voltage level under control, the reward is designed to motivate the controller to reduce the deviation of the bus voltage magnitude from the bus reference value V_{ref}. As shown in (3.1), if the system power flow diverges after applying the control action, a significant negative reward will be imposed. Otherwise, with less bus voltage deviation, the reward will become larger according to the first term of (3.1) in the case of system convergence. The reward function guides the controller to regulate its actions to reach better states. Additionally, we hope to reach the goal with less regulation so that the second term in (9) reflects the amount of regulation required in the control process. In (9), a_{ref} is usually the initial setting of the controlled parameter; c_1 and c_2 are weights selected based on the expert knowledge of the system as well as outcomes of the trials and errors[**?**].

$$r_t = \begin{cases} Huge\ penalty, \quad power\ flow\ diverges \\ -c * \sum_i \Delta v_i(t) - c_2 * \sum_j \Delta a_j(t),\ otherwise \end{cases} \tag{3.1}$$

The definition of Δv and Δa are given in (3.2)-(3.3).

$$\Delta v_i(t) = |v_i(t) - V_{ref}| \tag{3.2}$$

$$\Delta a_j(t) = |a_j(t) - a_{ref}| \tag{3.3}$$

where i is the number of the state observed, and j is the action dimension, which corresponds to the number of controlled transformers and switched shunts.

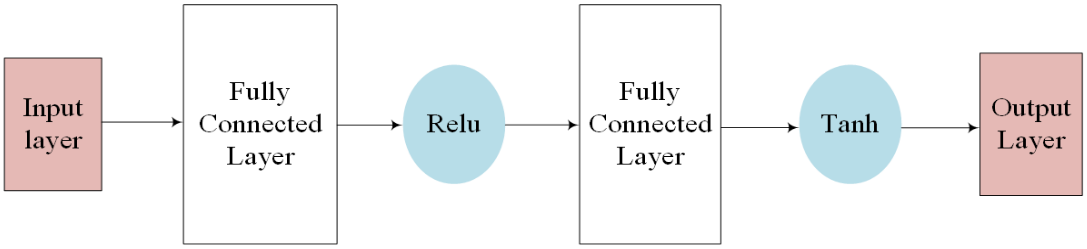

Figure 3.1: Actor neural network structure

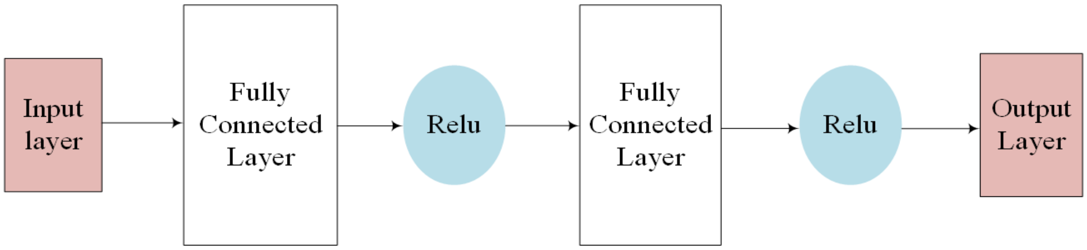

Figure 3.2: Critic neural network structure

3.1.4 Neural Network Architecture

The neural network structures adopted for the DDPG algorithm in this study are shown in Figure 3.1 and Figure 3.2. Both actor and critic neural networks have two hidden layers, which are connected with activation functions. The actor neural networks adopt Relu and Tanh activation functions and critic networks adopt Relu as the activation function.

3.2 Implementation of the DDPG-based Voltage Control Method

The overall implementation of the DDPG-based multiple devices voltage control is described in Algorithm 1 [79]. The power flow results generated by PSS/E for different scenarios of system load demand are used as the training data. Indices of M and T are each the episode number of training and the step number that indicates the maximum iteration count of each episode, respectively. At the start of the training process, four neural networks with different sets of random weights and

Algorithm 1 DDPG-based algorithm for multiple devices voltage control

```
input  : system voltage states
output: AVR transformer ratio and switched shunt group size
Initialize the critic network $Q$, $Q'$ and actor network $\mu$, $\mu'$ with random weights $\theta$,
  $\theta' \leftarrow \theta$ and $\phi$, $\phi' \leftarrow \phi$.
Initialize the experience replay buffer $D$.
for episode 1 to $M$, do
    Initialize the environment and obtain initial state $S_0$
    Initialize a random process $N$ for action exploration
    for step 1 to $T$, do
        Select action $a_t = \mu(s_t|\theta + N_t)$ according to the current policy and exploration
          noise
        Execute action $a_t$, observe $r_t$ and next state $s_{t+1}$
        Store transition ( $s_t$, $a_t$, $r_t$, $s_{t+1}$) in $D$
        Sample a random minibatch of $B$ transition ( $s_j$, $a_j$, $r_j$, $s_{j+1}$) from $D$
        Compute the critic target:
            $\mathrm{y}_j = R_j + \gamma Q'(s_{j+1}, \mu'(s_{j+1}|\theta^{\mu'})|\theta')$
        Update the critic Q-function by gradient descent using:
            L=1/N $\sum_j (y_j - Q(s_j, (a_j|\theta^Q))^2$
        Update the target networks as:
        $\nabla_{\theta^\mu} J = \frac{1}{N} \sum \nabla_a Q(s,a)|_{s=s_j, a=\mu_{(s_j)}} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s=s_j}$
        Update the network parameters:
            $\theta' \leftarrow \rho\theta + (1-\rho)\theta'$,    $\phi' \leftarrow \rho\phi + (1-\rho)\phi'$
```

the replay buffer size are initialized. For each episode, the power flow is solved to obtain the initial states (bus voltage magnitudes). In other words, this initializes the

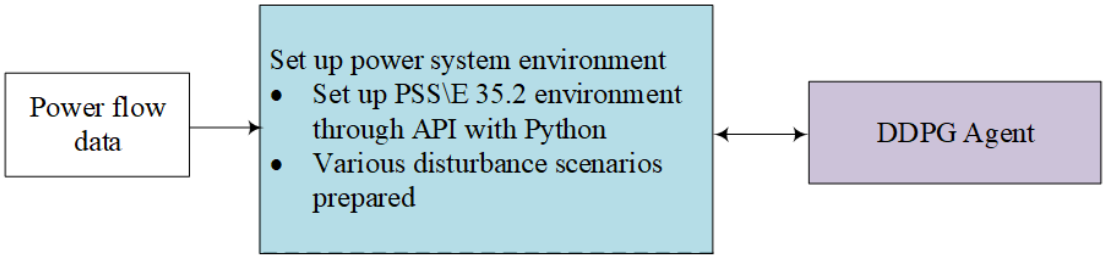
Power flow data
Set up power system environment
Set up PSS\E 35.2 environment through API with Python
Various disturbance scenarios prepared
DDPG Agent

Figure 3.3: Simulation platform for training DRL algorithm in power system environment

environment. A loop for a defined number of steps per episode begins with the action generated by the actor-network. The action is implemented in the power system environment by adjusting the transformer ratios and changing the dispatched group size of the switched shunts, as realized through the Python API with PSS/E. The training of these episodes terminates when no more voltage violations are detected, or the power flow diverges. Then, another episode is initiated. The structure of the simulation and training platform is shown in Figure 3.3. The agent learns from this repetitive process and keeps updating the parameters of the critic and actor neural networks by maximizing the accumulated reward that was designed to adjust the policy that generates actions until the maximum limit on episode M is reached.

3.3 Case Study

The proposed approach is verified on a WECC 240-bus system. The detailed configuration of this system can be found in [82]. The effectiveness of the proposed DDPG-based voltage control method is tested under different scenarios of load levels. The training data are generated from feasible power flow solutions considering reasonable constraints. The process for generating the power flow files is as follows. Starting with a base case, each load is randomly perturbed to vary in the range of 80 % to 120

% of the base case load using PSS/E. Then, to balance the power generation and load, generators are re-dispatched and their active power is adjusted based on a specified reserve requirement. All generators are adjusted simultaneously to compensate for the power imbalance where the output change of each generator is in proportion to its reserve capacity. Then the power flow is solved and the convergence is checked. Feasible power flow cases will be saved and added to the training data set. The desired voltage normal range is conservatively set to 0.98-1.02pu in this study, which can also be adjusted according to the system requirement. The maximum number of training episodes is set to 9000 with randomly selected power flow files generated based on the above description. Another set of 1000 cases is randomly chosen for testing. The maximum step number, which indicates the maximum iteration count of each episode during training, is set to 50.

3.3.1 Simulation parameters

The training parameters are crucial to the convergence of the algorithm. Generally, as shown in Table 3.2, the hyperparameters include the learning rate for both actor and critic networks, the discount rate γ, batch size, memory capacity, and training step. Exploration noise, which is used to enrich the training exploration, is also an important parameter that influences learning performance.

The learning rate determines the learning speed of the agent. The larger the learning rate is, the faster the agent will learn, but this could also lead to oscillations and result in a loss of the optimal solution. A smaller learning rate can make the learning process more precise but the drawback is the learning speed is slower. The process could be easily trapped into an overfitting situation. Generally, the learning rate is set to the range of 0.01-0.001. We used a learning rate of 0.001 due to the good performance during training.

Table 3.2: Training Parameters

Hyper parameters	Parameter values
Layer	2, 2 (actor, critic)
Activation Function	([ReLU, Tanh], [ReLU, ReLU])
Units of MLP per Layer	32
Learning Rate Actor	0.001
Learning Rate Critic	0.001
Discount rate γ	0.9
Batch Size	128
Memory Capacity	10000
Max Step	50
Exploration Noise	1.6

The discount rate essentially determines how much the reinforcement learning agent cares about rewards in the distant future relative to those in the immediate future. If it is set as 0, the agent will only learn about actions that produce an immediate reward. If it is set as 1, the agent will evaluate every action based on the total sum of all the future rewards. Most actions do not have long-lasting repercussions and need to be traded off to avoid irrelevant information. The discount rate is set as 0.9 in the training conducted and it provides satisfactory results.

The batch size indicates the number of training examples utilized in one iteration. Since the number of states and actions is not large in our case, the batch size is set as 128. The memory capacity is the capacity of the datasets that the agent will randomly sample from to train the network. It is used to break the correlation between different data to avoid inefficient learning. A value of 10,000 memory capacity is suitable for

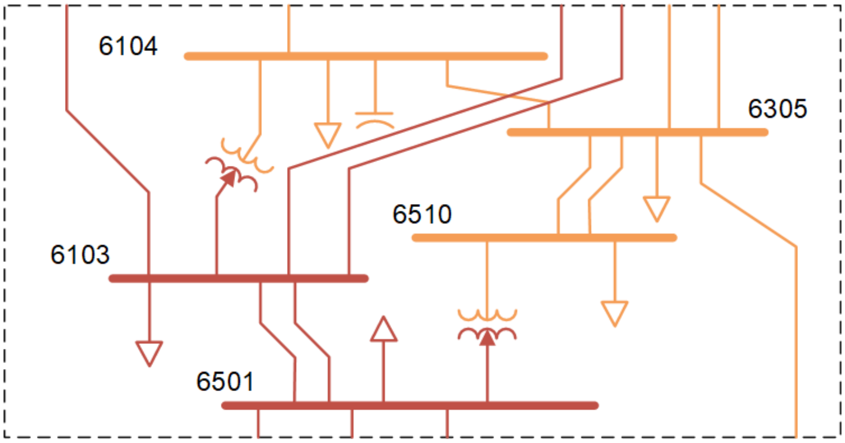

Figure 3.4: The controlled area of the WECC 240-bus system

a 128-batch size.

The step interval is set as 1 second, so the maximum step setting as 50 means every round of dynamic simulation will last 50 seconds after the disturbance is introduced. A 5-second initialization time is set so the total simulation will be 55 seconds for each episode.

Exploration noise is used to introduce more explorations during the training process, which will enrich the data set during the information exchange with the power system environment. The exploration noise is set by trial and error and a final value of 1.6 is obtained.

3.3.2 Case I: Voltage Control with Transformers Only

Since voltage control is a local problem, two transformers of the WECC 240-bus system Figure 3.4 are controlled to perform actions, and voltages of bus 6510 and the 6104 are observed. The action bound is set to 0.5-1.5, which corresponds to the lower and upper limits of the transformer ratio. Figure 3.5 and Figure 3.6 show the training and testing results. The blue line is the actual agent reward; the black line is the smoothed average reward that shows the reward trend; the red scatter in

Figure 3.6 represents the steps each episode takes to correct the voltage violations. At the start of the training, the reward is negative and small, the agent is not capable of outputting the appropriate control actions, and power flows diverge easily. The divergence will terminate the episode so that the agent only takes small incremental steps in the first 2300 episodes. With the training is proceeding, the agent gradually finds the right action policy, resulting in the reward having an upward tendency and finally reaching a high level, the episodes of these cases usually end due to reaching the maximum number of steps or with voltage violations being resolved. Figure 3.7 shows the bus voltages after each episode. After the reward increases significantly, the voltages are eventually restored within the range of 0.98-1.02pu. However, this process takes some long steps, as shown in the red scatter data in Figure 3.5. During testing (last 1000 episodes), the average of the agent's steps is 24.282.

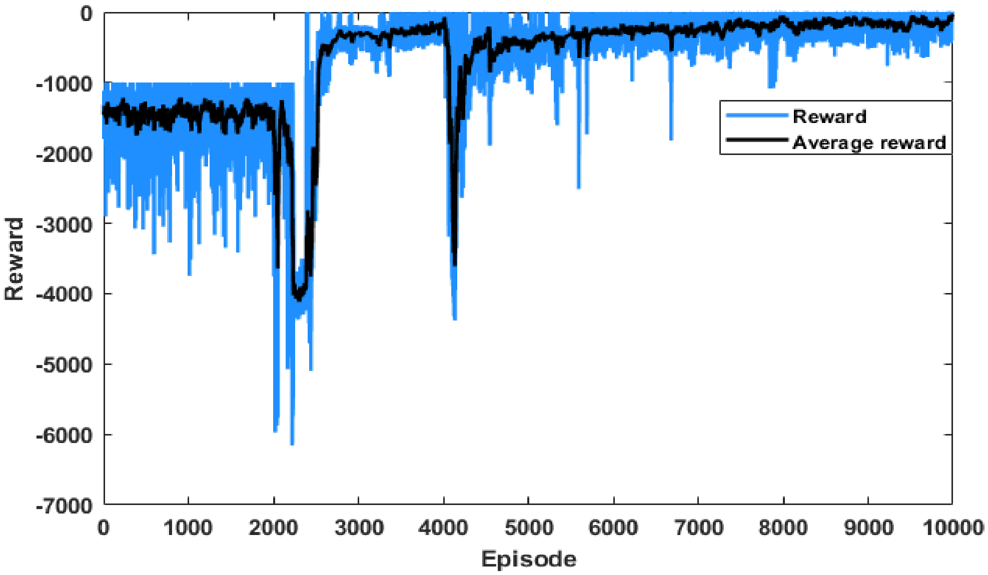

Figure 3.5: Rewards with transformers controlled only

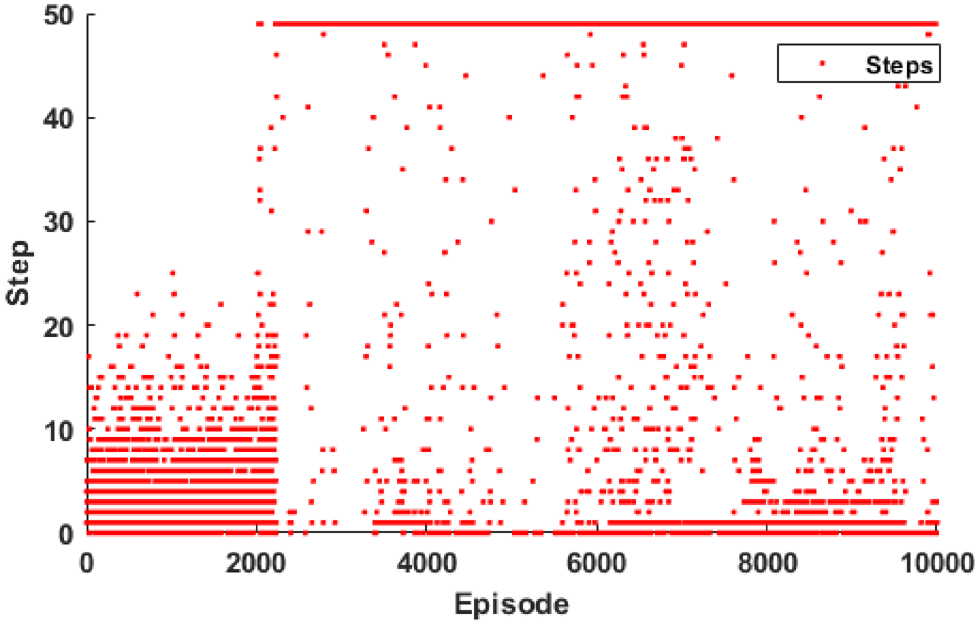

Figure 3.6: Number of steps with transformers controlled only

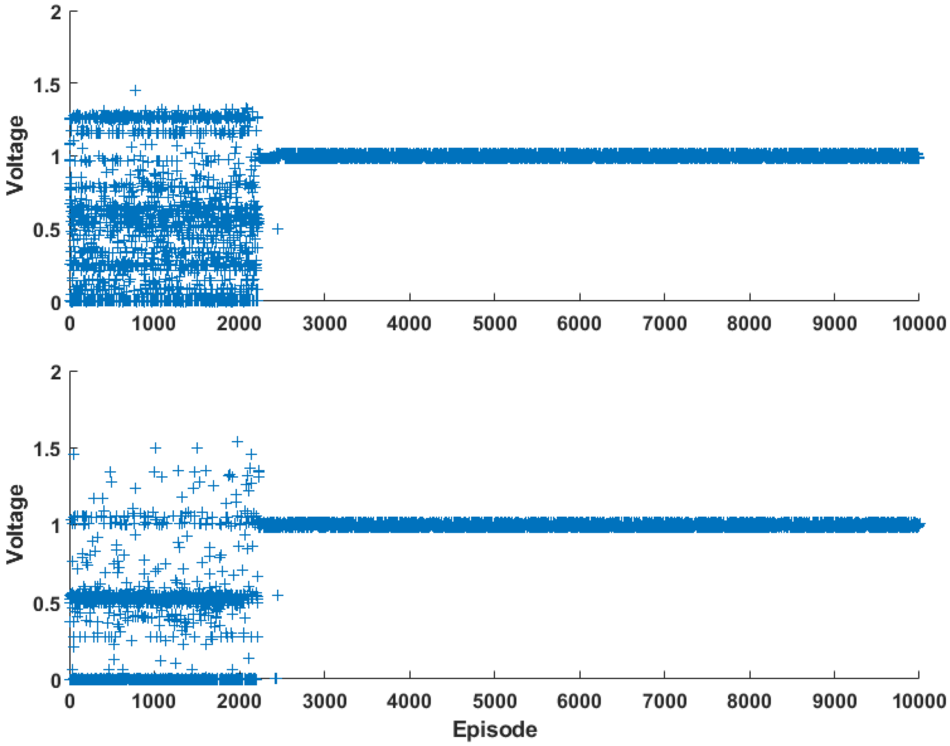

Figure 3.7: Bus voltages with transformers controlled only

3.3.3 Case II: Voltage Control with Transformers and switched shunt combined

Case II controls multiple types of devices with the DDPG-based agent, including two transformers and a switched shunt with five 100-MVar blocks. The switched shunt is connected to bus 6104. The action bound is set as 0.5-1.5. Figure 3.8 to Figure 3.10 show the results of Case II. The reward has an upward tendency and the value is closer to zero after being well trained. Compared with Case I, the performance of Case II shows a more stable convergence. The voltages can be regulated within the defined range of 0.98-1.02pu. Compared with Figure 3.5, Figure 3.8 demonstrates that the agent, which controls both the transformers and switched shunt, takes fewer steps to resolve the voltage violation.

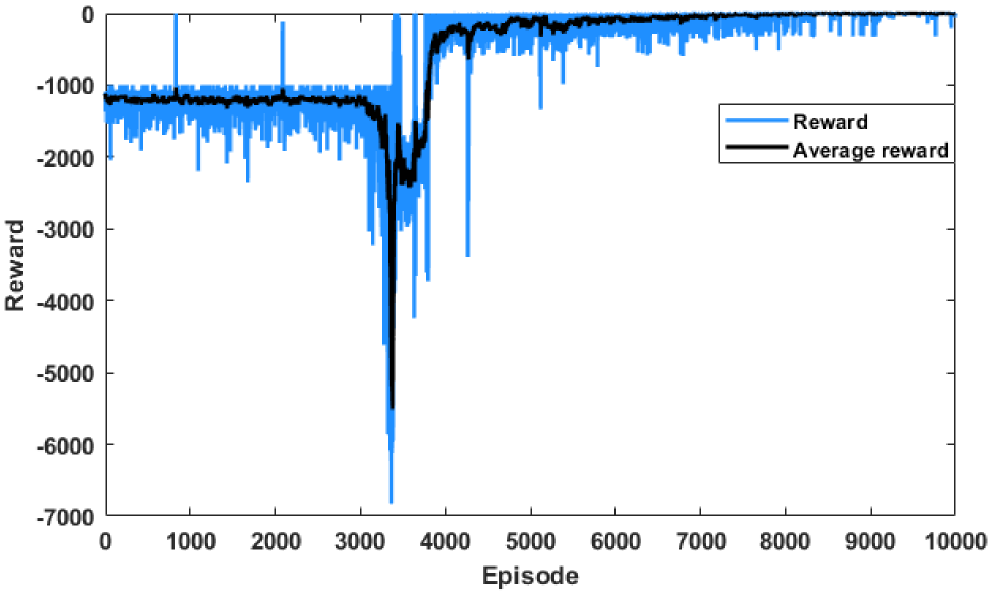

Figure 3.8: Rewards with transformers and switched shunt

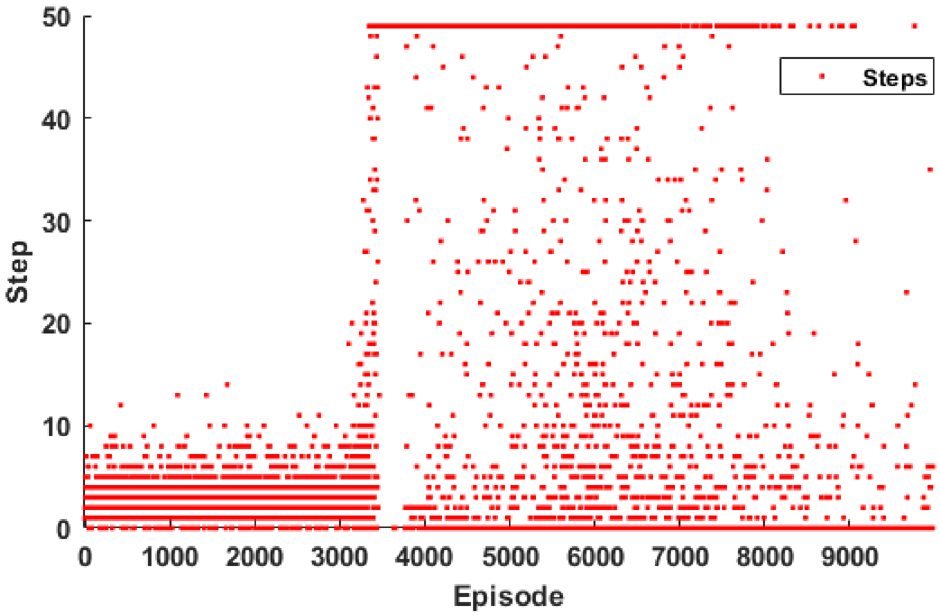

Figure 3.9: Number of steps with transformers and switched shunt

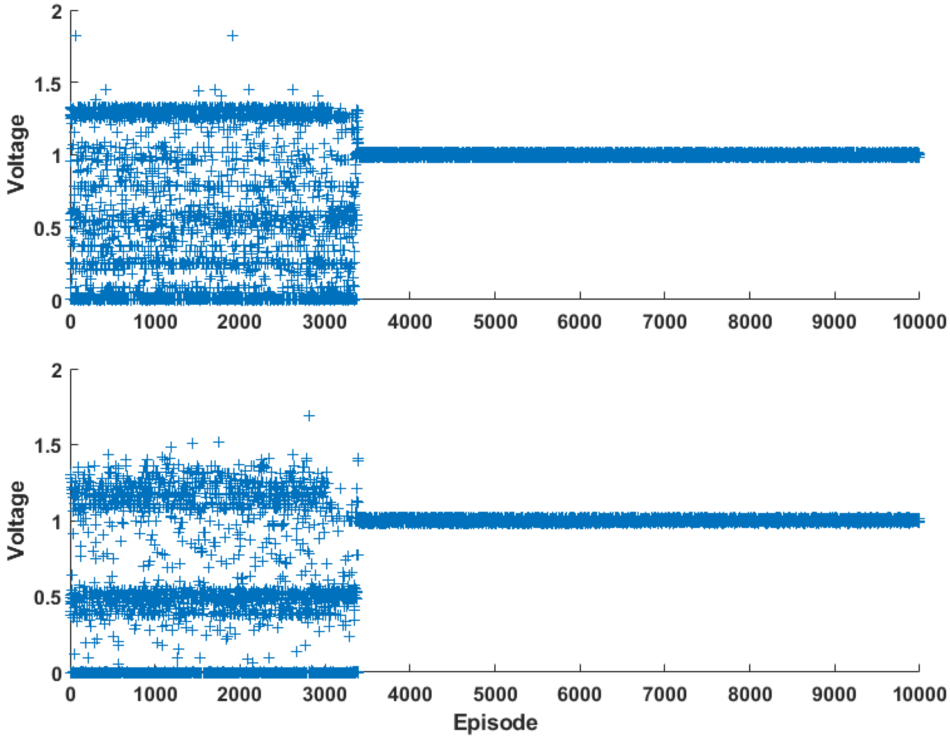

Figure 3.10: Bus voltages with transformers and switched shunt

During the testing, the agent takes an average of 1.536 steps to remove the violation. In Case II, the voltages of 94.7% of the testing cases can be regulated to the defined range within one step under the combined control of transformers and switched shunts. Meanwhile, in Case I, only 28.1% of scenarios can achieve voltage recovery within one-step control. The results demonstrate that the control of the DDPG-based agent using multiple devices can significantly improve the voltage control performance compared with the single device control.

3.4 Conclusions

This work proposes a DDPG-based voltage control scheme when the power system undergoes abrupt changes in the generation or load. Multiple voltage control devices are considered by the DDPG agent. The continuous and discrete actions are combined to showcase the proposed control method's capability to incorporate both continuous and discrete device types. The well-trained DDPG-based agent achieves robust performance in eliminating voltage violations with quick actions for different operating conditions. The proposed approach can make full use of the reactive power resources with different response characteristics to provide more reliable voltage support. Simulations on the WECC 240-Bus system verify the effectiveness of the proposed method.

Chapter 4

REAL-TIME EXCITATION CONTROL CONSIDERING SYSTEM DYNAMIC PERFORMANCE

4.1 Power System Dynamic Operation With Excitation System

A dynamical system is a complex system where the behavior evolves over time, and the power system is an example of such a system. It involves interactions between subsystems with an enormous number of variables that are constantly changing during operation. Thus, the dynamic process of power grid operation possesses a highly non-linear characteristic, which is essentially a process of sequential decision-making. In the event of a disturbance, it becomes essential to take appropriate control measures to ensure optimal control while considering power system stability, control cost, and variation of the dynamic variables of the power grid. This decision-making process can be described as a Markov decision process (MDP)[75] and solved by DRL algorithms, which will be discussed in more detail in Section III.

As for the action for the control of the excitation system, numerous parameter-setting methods have been discussed by researchers. However, the parameters are usually set as a constant before the generators are put into operation, which results in inflexibility and underutilization of reactive power[83]. To address this issue, DRL can be implemented to continuously optimize the excitation system parameters in real time during system operation. This allows the DRL algorithm to interact with the power system environment, exchange information, and learn the control policy of highly non-linear power systems without requiring detailed power grid model information.

4.2 Power System Model During Dynamic Operation

The dynamic process of power grid operation possesses a highly non-linear characteristic, which is essentially a process of sequential decision-making under uncertainty. The power system model can be formulated as follows[75]:

$$\mathbf{P} : min \int_{T_0}^{T_c} C(\mathbf{x}_t, \mathbf{y}_t, \mathbf{s}_t) dt \tag{4.1}$$

subject to

$$\dot{\mathbf{x}} : f(\mathbf{x}_t, \mathbf{y}_t, d_t, \mathbf{s}_t) \tag{4.2}$$

$$0 = f(\mathbf{x}_t, \mathbf{y}_t, d_t, \mathbf{s}_t) \tag{4.3}$$

$$\mathbf{x}_t^{min} \leq \mathbf{x}_t \leq \mathbf{x}_t^{max}, \forall t \in [T_0, T_c] \tag{4.4}$$

$$\mathbf{y}_t^{min} \leq \mathbf{y}_t \leq \mathbf{y}_t^{max}, \forall t \in [T_0, T_c] \tag{4.5}$$

$$\mathbf{a}_t^{min} \leq \mathbf{a}_t \leq \mathbf{a}_t^{max}, \forall t \in [T_0, T_c] \tag{4.6}$$

where $\mathbf{x_t}$ denotes dynamic state variables in the power system; $\mathbf{y_t}$ represents the algebraic states in the power system, such as the voltage of the buses of the power grid; $\mathbf{a_t}$ is the control action of the power system, such as generator regulation; d_t represents the system disturbance or fault that occurs during system operation; T_0 and T_c represent the time horizon of this dynamic process.

Equation (4.1) represents minimizing the total cost of the corrective control, including the cost of control actions and the control effectiveness in terms of system states (the control effectiveness can be reflected by system states). Equation (4.2) describes the dynamic system model, such as the behavior of generators and the relevant control systems. Equation (4.3) represents the power system constraints that describe the power balance between generators, loads and transmission branches. Equation (4.4)-(4.6) are the operational constraints of the system's dynamic states, algebraic states, and control actions. Equations (4.1) - (4.6) together describe the optimal decision-making model during power system operation[75; 84].

4.3 Definition of Action, State and Observation

Voltage magnitudes are commonly used to represent the operating condition of a power system in reactive power and voltage control problems, since other electrical statuses in system operation can be appropriately reflected in the voltage change[59; 80; 81]. Partial states in DRL algorithms can still work well for streaming valuable information, allowing for flexibility in data measurement and communication[59]. Thus, this study adopts bus voltage magnitudes as the observation states in the Markov decision process.

The control actions are defined as a vector of excitation system voltage reference values of the controlled generators. Each element of this vector is updated continuously.

The power system environment state transition is realized by a set of differential algebraic equations from (4.2) and (4.3). The limits on the value of the voltage references of the excitation system defined in (4.6) are considered in the definition of the action space by a predefined range of minimum and maximum values in considering the reactive power regulation capacity of each generator.

4.4 Definition of Reward

4.4.1 Consider voltage magnitude deviation and regulation cost

The reward function r_t is designed to evaluate the effectiveness of the control actions at each training step. To restore the voltage level under the control of the DRL agent, the reward is designed to motivate the agent to reduce the deviation of the observed bus voltage magnitude from the reference value V_{ref}. As shown in (4.7), if the system diverges after applying the control action, a significant negative reward will be imposed. Otherwise, with less bus voltage deviation, a smaller negative value will be added to the reward at each training step according to the first term of (4.7) in the case of system convergence. This results in a larger accumulated reward after each training episode composed by a predefined amount of steps. The reward function will gradually guide the agent to regulate its actions to reach better states. Besides the voltage magnitude level, we hope to reach the goal with less regulation cost, so the second term considers the amount of regulation during the control process, a_{ref} is the initial setting of the controlled parameter. c_1 and c_2 are the weights of these two parts, and they are chosen based on the expert knowledge of the system as well as trial and error selection[75]. The definition of Δv and Δa can be seen in (4.8)-(4.9).

$$r_t = \begin{cases} Huge\ penalty, \quad power\ system\ diverges \\ -c_1 * \sum_i \Delta v_i(t) - c_2 * \sum_j \Delta a_j(t),\ otherwise \end{cases} \tag{4.7}$$

$$\Delta v_i(t) = |v_i(t) - v_{ref}| \tag{4.8}$$

$$\Delta a_j(t) = |a_j(t) - a_{ref}| \tag{4.9}$$

4.4.2 Consider voltage magnitude deviation, regulation cost and historical voltage data

Power systems possess significant inertia. The dynamic process of the system during system operation is sequential, which means the current state of the system is affected by both the control actions as well as the previous system states. Significant information lies in the massive historical state data for an operating power grid or a given simulation. For voltage control problems, historical information can be provided by observing the history of bus voltage magnitudes. Therefore, the historical voltage magnitude data is added to the input to help the DRL agent learn a more accurate policy to cope with system disturbances. The reward function considering the historical data is formulated as (4.10):

$$r_t = \begin{cases} Huge\ penalty, \quad power\ system\ diverges \\ -c_1 * \sum_i \Delta v_i(t) - c_2 * \sum_j \Delta a_j(t) - \\ \quad c_3 * \sum_{t-c_t}^{t} \sum_i \Delta v_{h-i}(t),\ otherwise \end{cases} \tag{4.10}$$

$$\Delta history_k(t) = |v_{history_k}(t) - V_{ref}| \tag{4.11}$$

where $\Delta v_{history-i}$ is the historical voltage magnitude difference of bus i with bus reference value V_{ref}, c_t is the historical time range considered for a certain past time during system operation, and c_3 is the weight related to the historical data in the reward function.

4.4.3 Consider voltage magnitude deviation, regulation cost, historical voltage data, and voltage rate of change

During the system's dynamic evolution and control implementation after a disturbance or load change, the dynamic performance is also of significant importance. In order to avoid system oscillations and voltage fluctuations so as to facilitate the system voltage recovery in a more stable fashion, both the rates of voltage changes and their historical values are considered in the reward function (4.12) to guide the agent to generate a control policy that is able to aid in the recovery of the system voltage with more desirable dynamic performance. The reward function considering both voltage historical data and voltage rate of change is shown as (4.12):

$$r_t = \begin{cases} Huge\ penalty, \quad power\ flow\ diverges \\ -c_1 * \sum_i \Delta v_i(t) - c_2 * \sum_j \Delta a_j(t) - \\ \qquad c_3 * \sum_{t-c_t}^{t} \sum_i \Delta v_{h-i}(t) - \\ c_4 * \sum_{t-c_t}^{t-\Delta t} \sum_i \frac{v_{h-i}(t) - v_{hi-i}(t-\Delta t)}{\Delta t}, \ otherwise \end{cases} \tag{4.12}$$

where c_4 is the weight related to the rate of voltage change in the reward function, Δt is the time interval of every learning step in the training process. When applied to a practical power system, Δt could be the data sampling time step of the measurement device.

4.5 Simulation Platform Development and Implementation

The overall implementation of the DDPG-based real-time excitation system control is described in Algorithm 2. A transmission-level power system dynamic simulation and training platform is developed for the training and implementation of the algorithm in the power system dynamic simulation environment. The time-domain simulation software Siemens-PTI PSS/E is used as the power system simulator to conduct power system dynamic simulations and emulate the power grid environment. PSS/E provides application programming interfaces (APIs) with Python, which can communicate the power system simulation environment to the DRL agent in real time to exchange information, as shown in Figure 4.1.

Algorithm 2 Deep Deterministic Policy Gradient algorithm for Real-time Dynamic Voltage Control

```
input  :  power system environment states
output:  control action applied to the power system environment

Initialize the critic network $Q$, $Q'$ and actor network $\mu$, $\mu'$ with random weights $\theta$,
  $\theta' \leftarrow \theta$ and $\phi$, $\phi' \leftarrow \phi$.
Initialize the experience replay buffer $D$.
for episode 1 to $M$, do
    Initialize the environment and obtain initial state $S_0$
    Initialize a random process $N$ for action exploration
    for step 1 to $T$, do
        Select action $a_t = \mu(s_t|\theta + N_t)$ according to the current policy and exploration
          noise
        Execute action $a_t$, observe $r_t$ and next state $s_{t+1}$
        Store transition ( $s_t$, $a_t$, $r_t$, $s_{t+1}$) in $D$
        Sample a random minibatch of $B$ transition ( $s_j$, $a_j$, $r_j$, $s_{j+1}$) from $D$
        Compute the critic target:
                $\mathrm{y}_j = R_j + \gamma Q'(s_{j+1}, \mu'(s_{j+1}|\theta^{\mu'})|\theta')$
        Update the critic Q-function by gradient descent using:
                  $\mathrm{L{=}1/N} \sum_j (y_j - Q(s_j, (a_j|\theta^Q))^2$
        Update the target networks as:
        $\nabla_{\theta^\mu} J = \frac{1}{N} \sum \nabla_a Q(s,a)|_{s=s_j, a=\mu_{(s_j)}} \nabla_{\theta^\mu} \mu(s|\theta^\mu)|_{s=s_j}$
        Update the network parameters:
                    $\theta' \leftarrow \rho\theta + (1-\rho)\theta'$,
                    $\phi' \leftarrow \rho\phi + (1-\rho)\phi'$
```

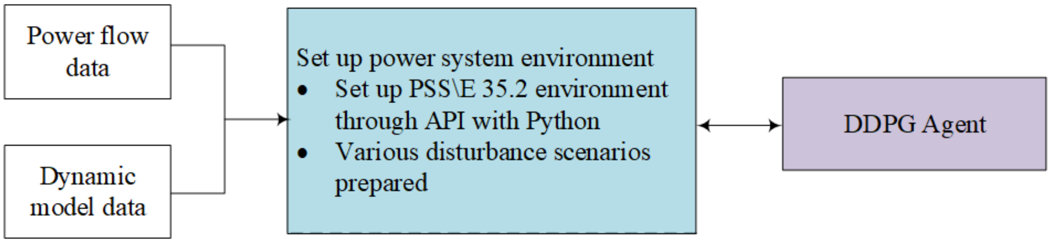
Power flow data
Dynamic model data
Set up power system environment
• Set up PSS\E 35.2 environment through API with Python
• Various disturbance scenarios prepared
DDPG Agent

Figure 4.1: Simulation platform for training DRL algorithm in the power system environment.

Figure 4.2 shows the training procedure and the data interaction between the power system simulator and DRL agent in the training platform. The blue and purple blocks represent the actions conducted in PSS/E and Python, respectively. The two software elements constantly exchange information using the application programming interface (API) in training. The green arrows show the interaction data flow between them. Power flow and dynamic model files are prepared to perform power system dynamic simulation. At the start of the training process, four neural networks with different sets of random weights and the replay buffer size are initialized. For each episode, the power flow is solved, and dynamic simulation is initialized based on the selected study case. The disturbance is randomly introduced, and the initial states are obtained for each training episode, in which one round of dynamic simulation begins. A loop for a predefined number of steps per episode starts with the action generated by the DRL agent. The action then will be sent to the power system simulator and implemented in PSS/E by adjusting the voltage reference input of the excitation system. Then, the dynamic simulation will be run for one training step interval to update the states of the power system environment, and the most updated states are sent back to the DRL agent. The reward will be calculated based on the system observation to evaluate the performance of the learned policy. The data will be collected and stored after each round of interaction between the power system

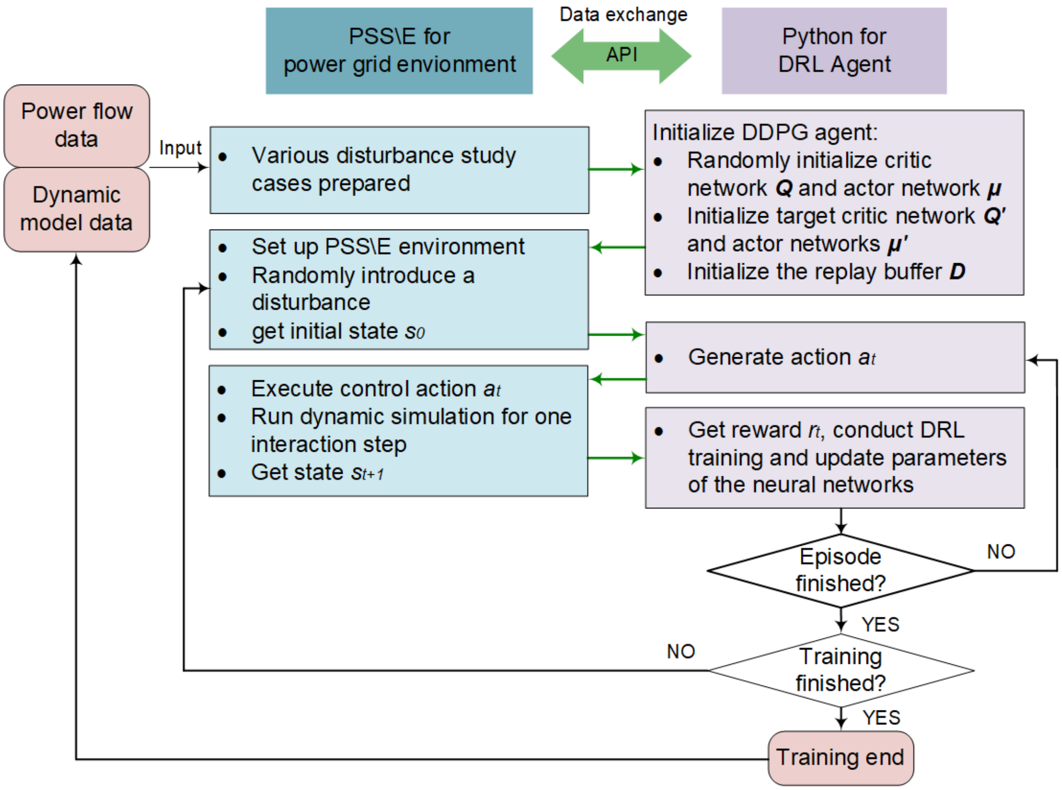
PSS\E for
power grid envionment
Data exchange
API
Python for
DRL Agent
Power flow data
Dynamic model data
Input
Various disturbance study cases prepared
Initialize DDPG agent:
Randomly initialize critic network Q and actor network μ
Initialize target critic network Q′ and actor networks μ′
Initialize the replay buffer D
Set up PSS\E environment
Randomly introduce a disturbance
get initial state s0
Generate action at
Execute control action at
Run dynamic simulation for one interaction step
Get state st+1
Get reward rt, conduct DRL training and update parameters of the neural networks
Episode finished?
NO
YES
Training finished?
NO
YES
Training end

Figure 4.2: Data flow of the simulation platform for training DRL algorithm in power system environment.

simulator with the objective of further training. The DRL agent will then learn and update the parameters of the neural networks based on the observation data. Another round of learning begins until reaching the predefined number of steps, and then another episode is initiated. The agent learns from this repetitive process and keeps updating the parameters of the critic and actor neural networks by maximizing the accumulated reward that was designed to adjust the policy of the action generation until the maximum limit on the episodes is reached.

For each training time step interval in the platform, the dynamic simulation will run for one time step to update the system states, and there will be one round of interaction between the power system simulator (PSS/E) and DRL agent (Python), during which data exchange happens.

This training platform is based on power system dynamic simulation (both power flow data and dynamic data are required) and is used for the emulation of the real-time power system operation environment. Dynamic characteristics of systems can be observed by continued interaction and data exchange during detailed time-domain simulations. Different power system control problems can be addressed by applying and testing various state-of-the-art DRL algorithms based on this platform across a range of power grid simulations varying in scale.

4.6 Simulation and Results

The IEEE 9-bus system[85] and the 2000-bus Texas synthetic grid systems[86; 87; 88] are used as the test systems, based on which time-domain simulations are conducted and interfaced with the DDPG controller. All the case studies, including training and testing, were performed in the simulation environment based on the platform described in Section IV.

4.6.1 Simulation Parameters

With careful tuning by trial and error, the set value of the training parameters are shown in Table 4.1. The learning rates for both the actor and critic are set as 0.001 with a 0.9 discount rate. The batch size, which indicates the number of sampled training data utilized from the reply buffer in one iteration, is set as 128 in considering the number of states and actions space in this study. A value of 10,000 memory capacity is adopted to adapt for a 128-batch size. Exploration noise is set as 3 to introduce explorations that can enrich the data set.

Both actor and critic neural networks have two hidden layers, which are connected with activation functions. The actor neural networks adopt Relu and Tanh activation

Table 4.1: Training Parameters

Hyper parameters	Parameter values
Layer	2, 2 (actor, critic)
Activation Function	([ReLU, Tanh], [ReLU, ReLU])
Units of MLP per Layer	32
Learning Rate Actor	0.001
Learning Rate Critic	0.001
Discount rate γ	0.9
Batch Size	128
Memory Capacity	10000
Max Step	50
Exploration Noise	3

functions, and critic networks adopt Relu as the activation function. Each layer includes 32 units to store and update the data.

The training step interval is set as 1 second, which means the power grid environment will exchange information with the DDPG agent, send current states, and get action commands every 1 second. The maximum step number, which indicates the maximum iteration count of each episode during training, is set to 50. The dynamic simulation will first run for 5 seconds to provide the initial states to start the training in each episode. Then, the disturbance is added at 5s. Therefore, each round of dynamic simulation will run for 55 seconds in total in every episode.

4.6.2 IEEE 9-Bus System

The IEEE 9-bus system includes three generators and nine buses, as shown in Figure 4.3. The system parameters are shown in Table 4.2. Generator 1, a hydraulic unit with the salient-pole generator model GENSAL, is connected to slack bus 1. Generators 2 and 3 are steam turbines with the round-rotor synchronous generator model GENROU. They are controlled by the DDPG agent to participate in voltage control. All three generators are equipped with an IEEE type 1 excitation system model (IEEET1) and an IEEE standard governor model (IEESGO). The maximum action output is set as 1.3. Different load models can be applied to the system. In the simulation of this report, the system loads include an active power component of constant current load and a reactive power component of constant impedance load. All loads are located on buses 5, 6, and 8. The reactive power load is randomly perturbed as the disturbance, which results in around 3% - 5% voltage fluctuations. The desired voltage normal range is conservatively considered as 0.98-1.02pu in this study, so the voltage reference in (4.8) and (4.11) is set as 1.00 pu to guide the DRL agent to control the voltage within the set range.

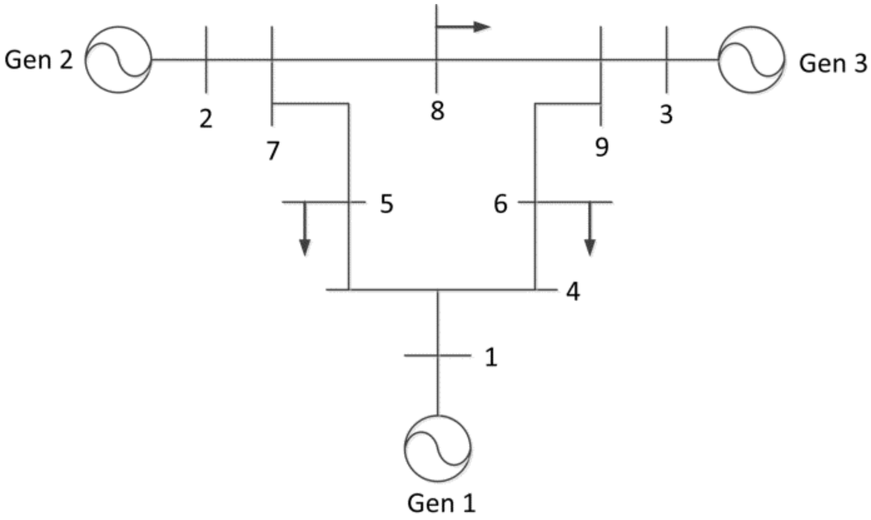

Figure 4.3: IEEE 9-bus test system.

Table 4.2: Parameters of the IEEE 9-bus system

Bus Number	Voltage(kV)	Generator Output(MVA)	Load (MVA)
1	16.5	247.5	/
2	18	192	/
3	13.8	128	/
4	230	/	/
5	230	/	$125 + j50$
6	230	/	$90 + j30$
7	230	/	/
8	230	/	$100 + j35$
9	230	/	/

Case 1: Considering voltage magnitude deviation and regulation cost

The agent is trained with the reward function of (4.7) that considers bus voltage magnitude deviation and generator regulation cost. Figure 4.4 shows the moving average reward finally reaches a satisfactory level after 2000 episodes of training. The DDPG agent is applied to the system after being well-trained for testing by adding load disturbance at 5s to induce voltage changes. The test results, depicting the response to a 90 MVar reactive power load increase, are illustrated in Figure 4.5 and Figure 4.6.

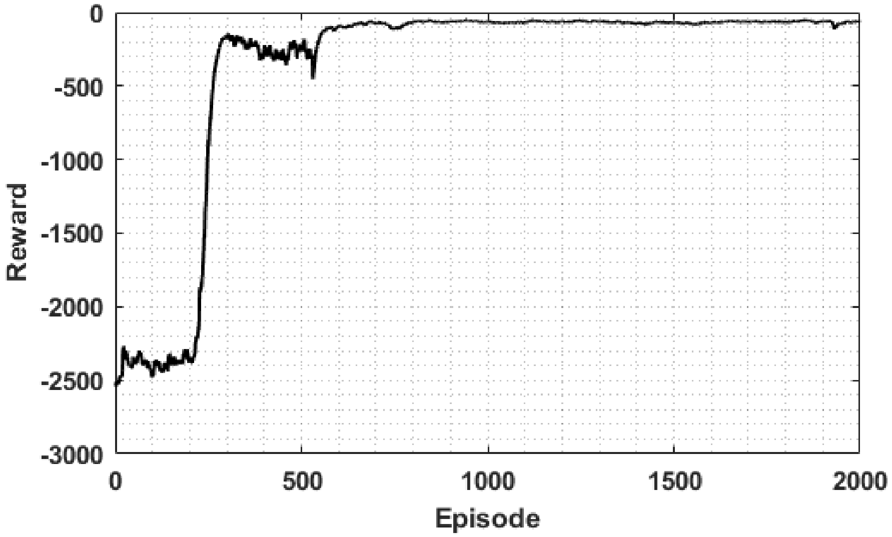

Figure 4.4: Case 1 of IEEE 9-bus system: Average reward.

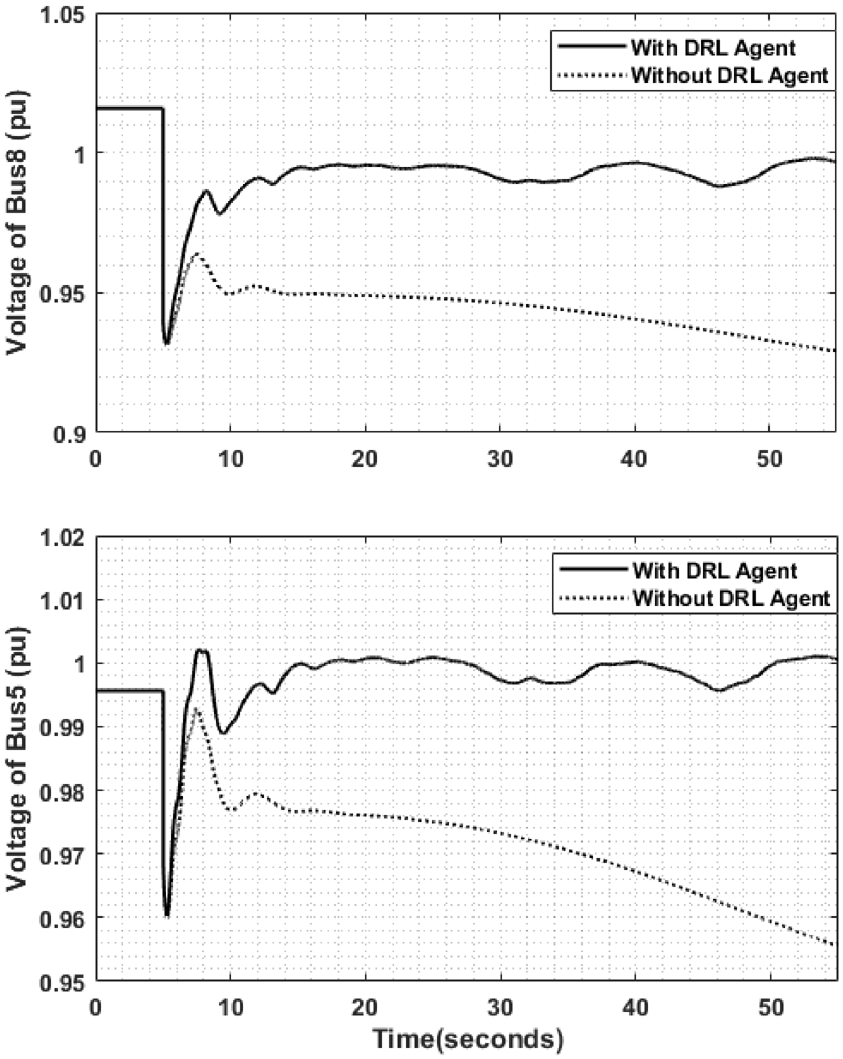

Figure 4.5: Case 1 of IEEE 9-bus system: Voltage of bus 8 and bus 5 with and without DRL agent.

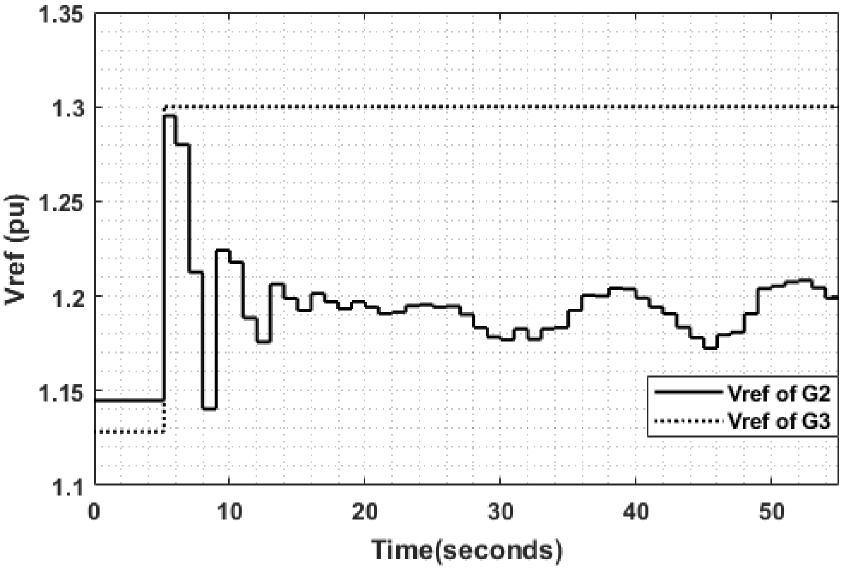

Figure 4.6: Case 1 of IEEE 9-bus system: Generator voltage reference commands from DRL agent.

Under generator control with constant exciter parameters, the system bus voltage magnitudes are significantly impacted and keep decreasing after the disturbance, which puts the system at high risk of losing stability. With the DDPG agent participating in the voltage control, bus voltages can be regulated to normal levels. The change of the excitation system voltage reference value of the two controlled generators can be seen in Figure 4.6. Generator 3 provides full voltage support after detecting the disturbance, and generator 2 is responsible for the voltage regulation in real time according to the system operating. The two generators cooperate under the control of the DDPG agent to help the system restore voltage.

Case 2: Consider voltage deviation, regulation cost and historical voltage data

To further analyze the impact of historical data on agent control performance, we trained the DDPG agent with the reward function (12) that considers historical voltage data, bus voltage magnitude, and generator regulation cost. c_t in (4.10) is set as 5, meaning the last 5 seconds of data are considered. After 2000 episodes of training, the moving average reward shown in Figure 4.7 reached and maintained a high level. After the training converges, the DDPG controller is implemented in the dynamic simulation of the system. This test simulation involves introducing the same 90 MVar reactive load change, enabling a comparison with case 1. The results of Figure 4.8 show that the DDPG agent's control policy considering historical voltage data can provide support to the system, helping it recover to a normal voltage level. It's worth noting that in case 2, the bus voltage recovered faster with fewer oscillations, which demonstrated better dynamic performance compared to case 1. This

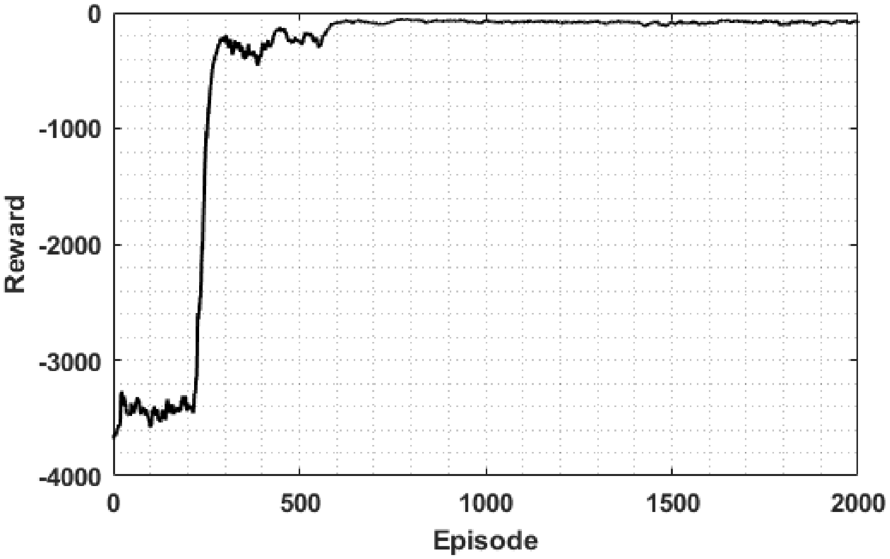

Figure 4.7: Case 2 of IEEE 9-bus system: Average reward.

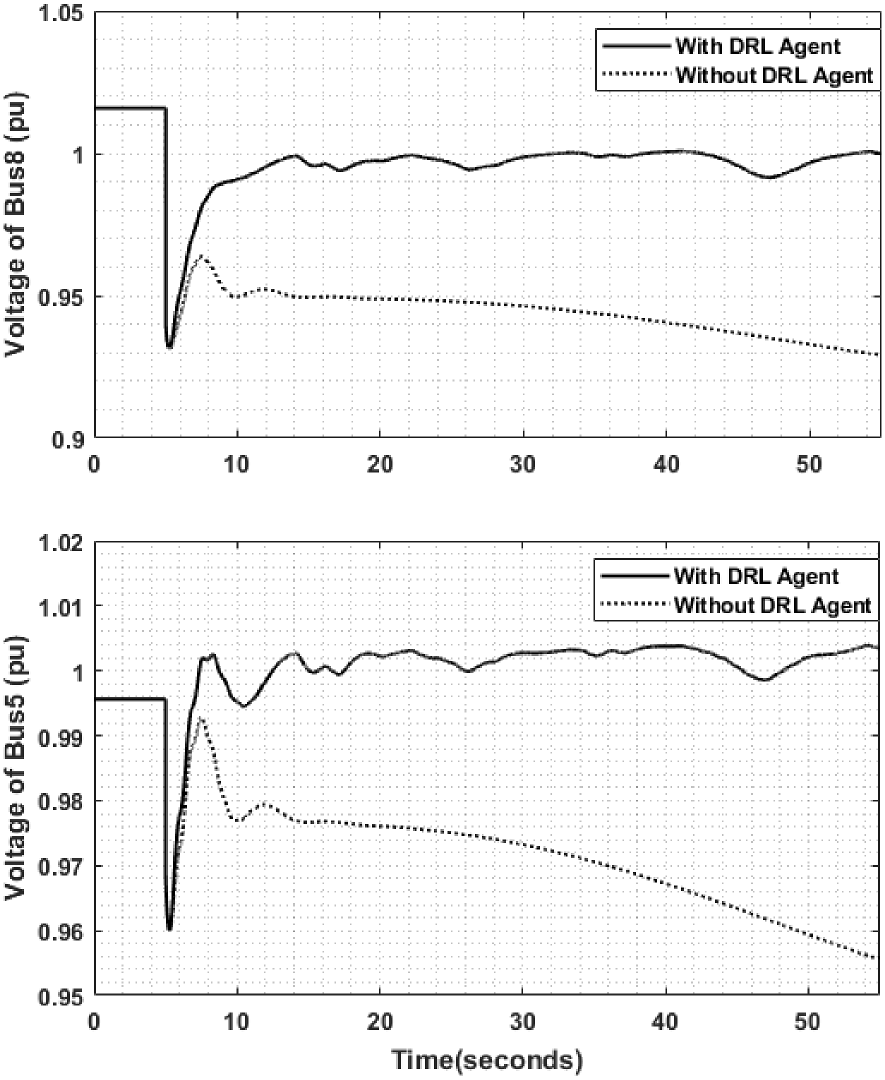

Figure 4.8: Case 2 of IEEE 9-bus system: Voltage of bus 8 and bus 5 with and without DRL agent.

provides evidence that historical data can provide valuable information to the DDPG agent, improving its policy accuracy in managing voltage oscillations and fluctuations during system operation.

Case 3: Consider voltage deviation, regulation cost, historical voltage data and rate of change of voltage

To explore the impact of the rate of voltage change data on the DDPG agent's performance, further training using the reward function in (4.12) is conducted. This function considers the rate of voltage change in historical data, which is calculated using the previous and present voltage values. For the preceding 5 seconds of historical data, there are four rates of voltage change data for each controlled bus. Following the completion of training, which is shown in Figure 4.9, the DDPG controller is tested with a reactive power load increase of 90 MVar as well. As shown in Figure 4.10, the results demonstrate that the agent can effectively support the system voltage recovery

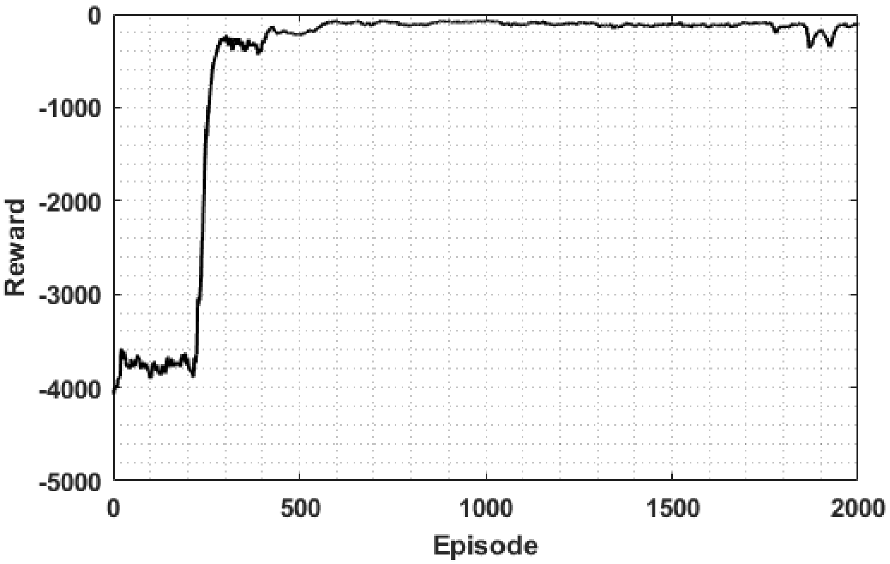

Figure 4.9: Case 3 of IEEE 9-bus system: Average reward.

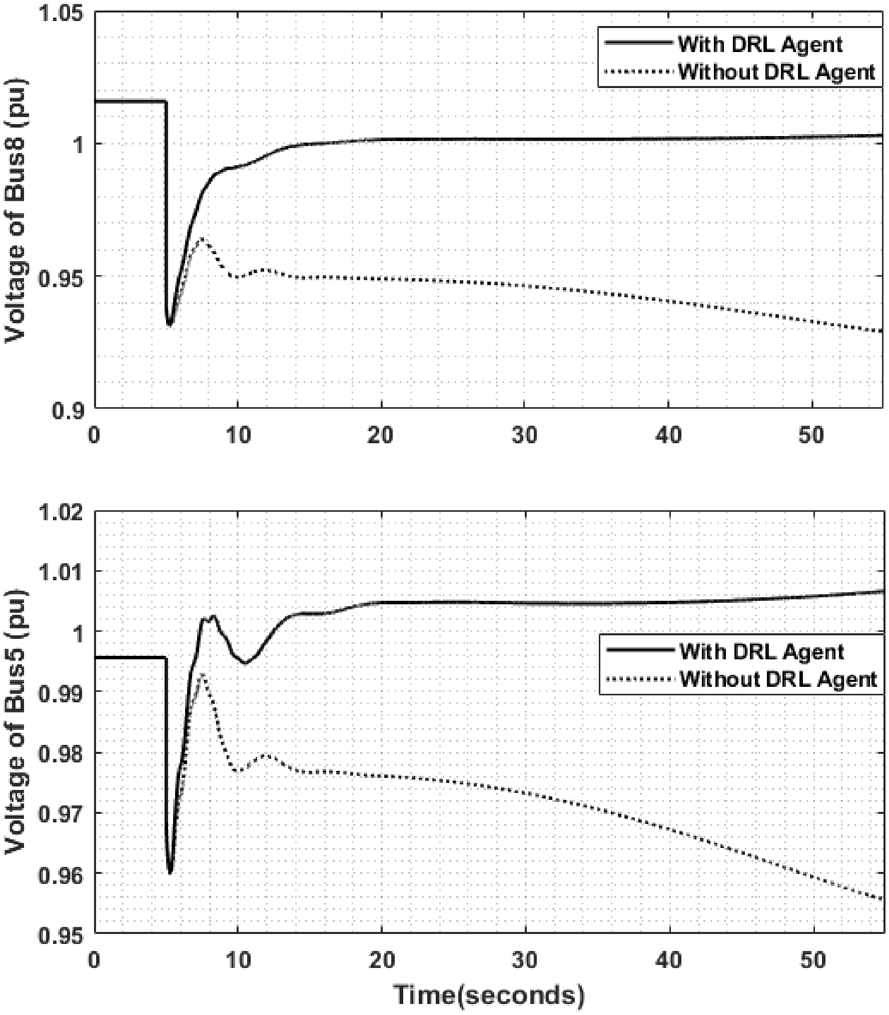

Figure 4.10: Case 3 of IEEE 9-bus system: Voltage of bus 8 and bus 5 with and without DRL agent.

to the desired range in a more stable manner.

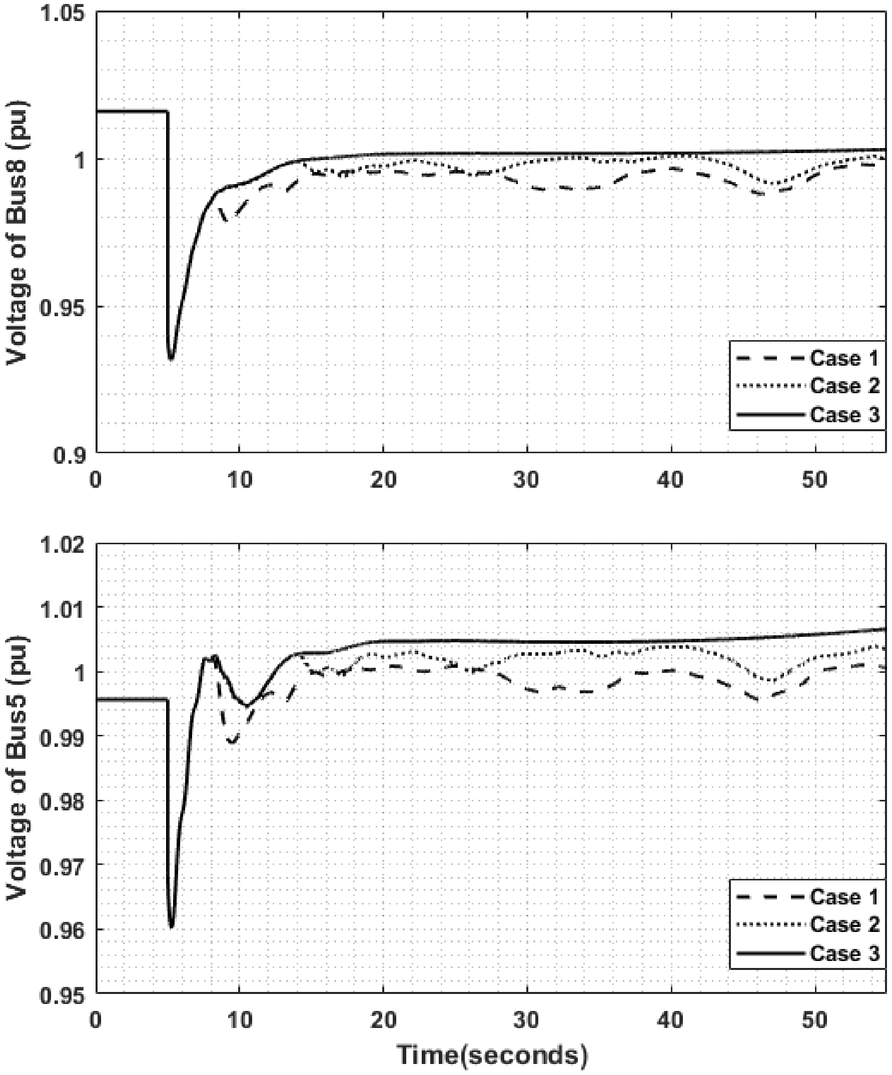

Figure 4.11: Case 1 to Case 3 comparison of IEEE 9-bus system: Voltage of bus 8 and bus 5.

In the analysis of the DDPG agent's dynamic control performance, different types of information in the reward functions are analyzed in Case 1 through Case 3. Figure 4.11 shows a comparison of the voltage control performance when the DDPG agent is tested with the same disturbance. The solid curve of Case 3, which considers both historical voltage data and voltage rate of changes, exhibits the smoothest

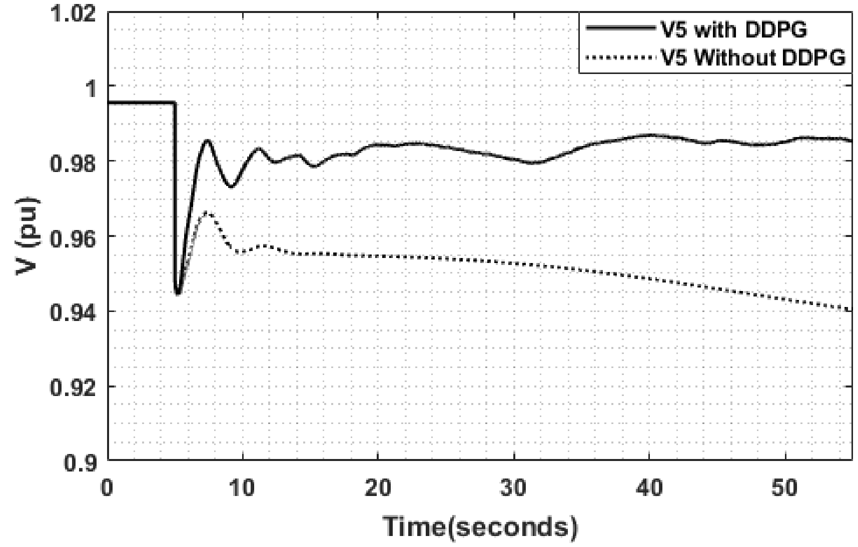

Figure 4.12: Case 4-Voltage of bus 5 when a disturbance occurs at bus 5.

voltage curve with the least fluctuation under the control of the DDPG controller. Additionally, Case 3 is capable of regulating and recovering the voltage faster due to the controller's ability to more accurately predict voltage changes based on dynamic features learned during training. The agent provided with extra information on the rate of voltage change can generate more effective actions to not only control the voltage level but also achieve better dynamic control performance.

Case 4: Simulation with the randomly selected location of the disturbance

The disturbance of Case 1 to Case 3 is located at bus 8. This section analyzes the scenario with randomly selected disturbance locations. The disturbance is randomly added to bus 5, bus 6, and bus 8 with random load change amount. Only voltage magnitude and regulation cost are considered in this scenario, which uses (4.7) as the reward function. After the agent is well trained, the test results are shown in Figure 4.12 and Figure 4.13. We can see from the results that the voltage can still be recovered with the random location of the disturbance. This scenario is more likely to

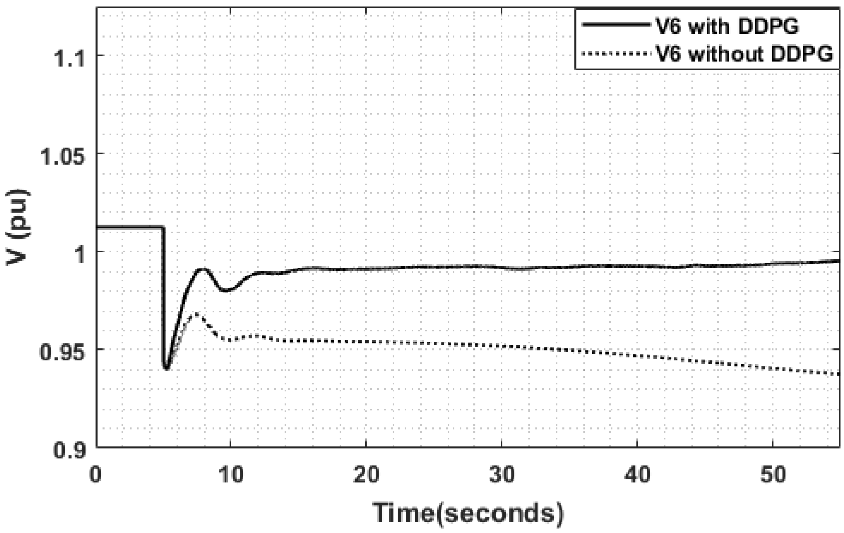

Figure 4.13: Case 4-Voltage of bus 6 when a disturbance occurs at bus 5

happen in the real power grid since faults always appear with significant uncertainty.

4.6.3 Texas 2000-Bus Synthetic Test System

To evaluate the effectiveness of the proposed DRL-based dynamic voltage control method on a more realistic system, simulations are conducted on the Texas 2000-bus synthetic power system, which is a large-scale representation of an actual power grid. This serves as a crucial step to test the proposed control method and the training platform.

The whole 2000-bus power test system is synthetic and has four voltage levels of 500/230/161/115 kV. The total generation capacity in this system is 98GW with a load of 67GW and 19GVAr. The heavily loaded area is in southeast Texas around the Houston area and the Northern part of the Texas grid.

In Figure 4.14, the structure of the Texas 2000-bus synthetic power system is depicted, where disturbances are introduced in the heavily loaded Houston area (highlighted in red) to simulate scenarios with voltage issues. Among the generators, 7098

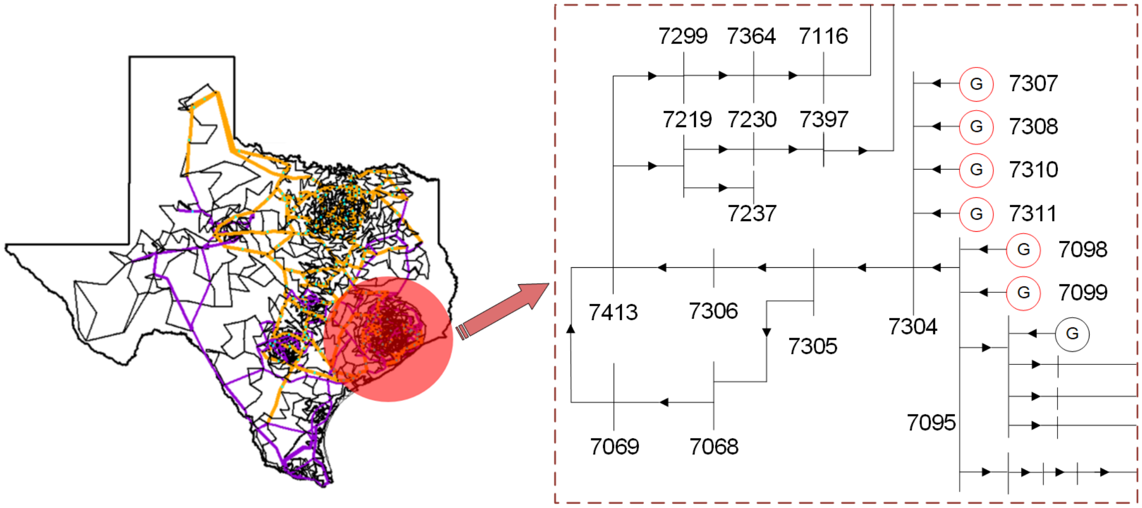

Figure 4.14: Diagram of disturbance area of 2000-bus system.

and 7099 are well-suited as controlled generators due to their large capacity and ample reactive power capability. As generator 7098 is connected to the swing bus of the system, generator 7099 is selected as the controlled generator, along with generator 7310, which is located at a short electrical distance from the Houston area. These two generators are chosen as the controlled generators for this case study. Generators 7099 and 7310 are both represented with the GENROU generator model. Generator 7099 employs an IEEET1 exciter model and IEEE type 1 speed-governing model (IEEEG1). While generator 7310 utilizes the ESST4B exciter model and a general turbine-governor model(GGOV1). The system includes the same load model as the 9-bus system. To train and evaluate the controller's response to voltage changes, system disturbances are induced by altering the reactive power loads.

Parameter adjustments for the Texas 2000-bus system

Dynamic simulations are conducted based on this data and an oscillation is found in the system when no disturbance is added, as seen in Figure 4.15. This indicates that a suitable initial condition was not determined for the time domain simulation. As a result, some system parameters or control settings can be erroneous.

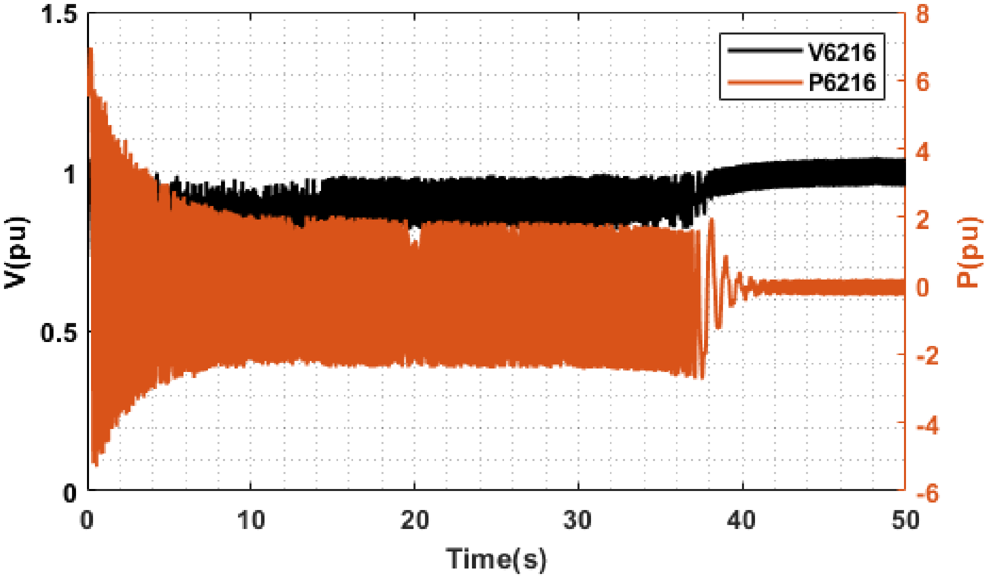

Figure 4.15: Diagram of disturbance area of 2000-bus system.

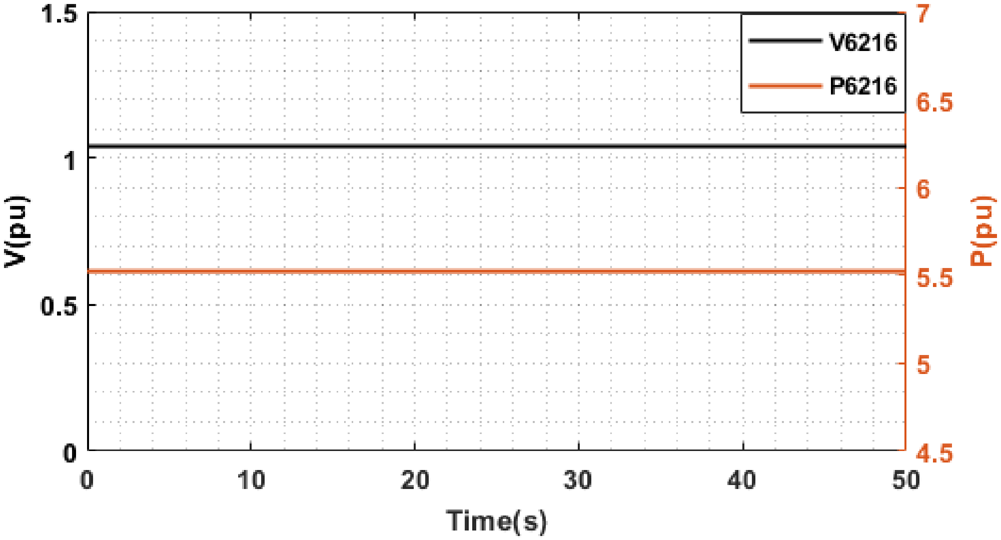

Figure 4.16: Diagram of disturbance area of 2000-bus system.

To remedy this problem and make preparations for the agent training, the appropriate parameter(s) of the test system is adjusted. The test case was simulated in post PSS/E and PSLF, which have excellent initial condition analysis capability for dynamic initialization. According to the warning information that PSS/E and PSLF provided, we adjusted the maximum limit of the governor and the excitation system, and this reduced the oscillation. Generator 6216 greatly influences system stability. The lead and lag time constant and "Vrmax" of generator 6216's excitation system were further adjusted to appropriate values. As a result, the initialization is successful and a flat run indicating this is obtained, as shown in Figure 4.16. The following training and testing simulations are based on this corrected data.

Case 1: Considering voltage magnitude deviation and regulation cost

The simulation begins with the base case that utilizes equation (4.7) as the reward function, which considers voltage magnitude and regulation cost. After training the DDPG agent, as shown in Figure 4.17, where the reward reaches a high level, the agent is tested and the results are shown in Figure 4.18 to Figure 4.19. The results indicate that the agent can improve the voltage to a satisfactory level compared to the

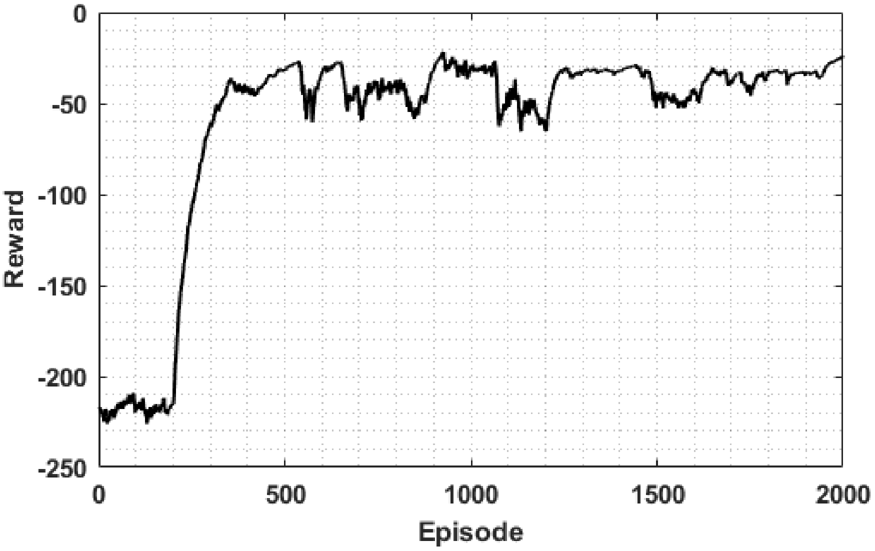

Figure 4.17: Case 1 of the 2000-bus system: Average reward.

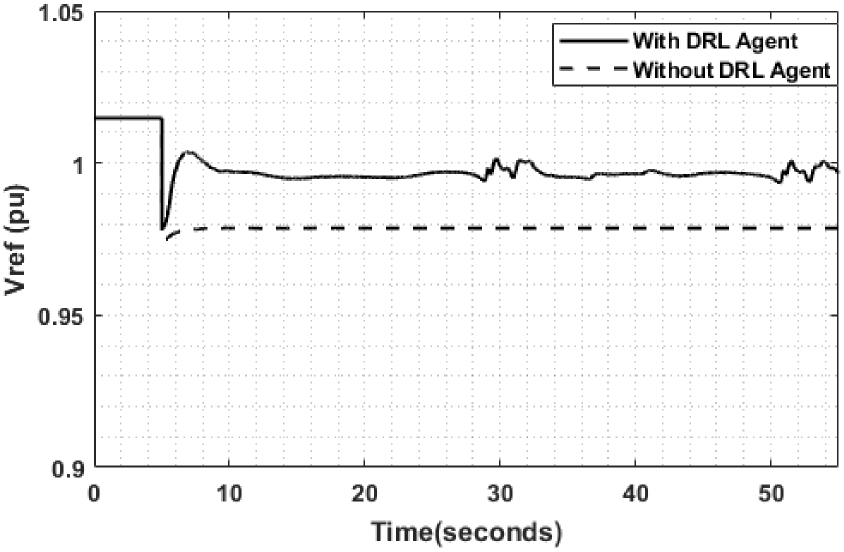

Figure 4.18: Case 1 of the 2000-bus system: Voltage of bus 7068 with and without DRL agent.

conventional control mode. When a 230 MVar reactive load increases at 5s, the agent can detect the voltage change and generate commands to improve the generators' output immediately. The voltage of bus 7068 is shown in Figure 4.18 and Figure 4.19 as the voltage level representative for analysis. The voltage is restored to a normal level in about 2 seconds after the disturbance, and the generators can continuously regulate the excitation systems to achieve real-time voltage control in the recovery process. The two generators can respond quickly to voltage fluctuations under the control of the DDPG agent, which performs well in both situations of quick voltage control during sudden disturbances and minor voltage regulation in the process of system recovery.

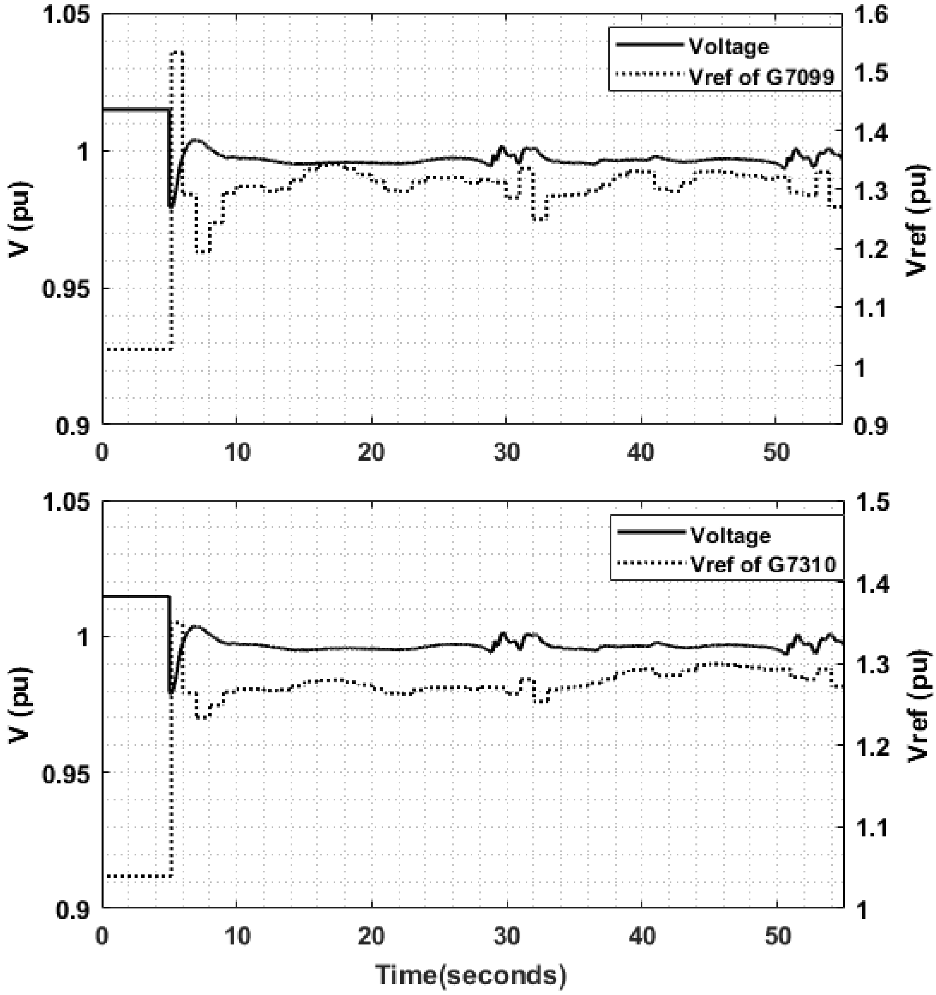

Figure 4.19: Case 1 of the 2000-bus system: Generator voltage reference commands from DRL agent.

Case 2: Consider voltage deviation, regulation cost, historical voltage data and rate of change of voltage

Various reward functions are employed for the DDPG controller in the Texas 2000-bus system, including considering historical voltage deviation and adding voltage rate of change in addition to the base case. The simulation results with the same load disturbance as case 1 are presented in Figure 4.20. The addition of voltage rate of change

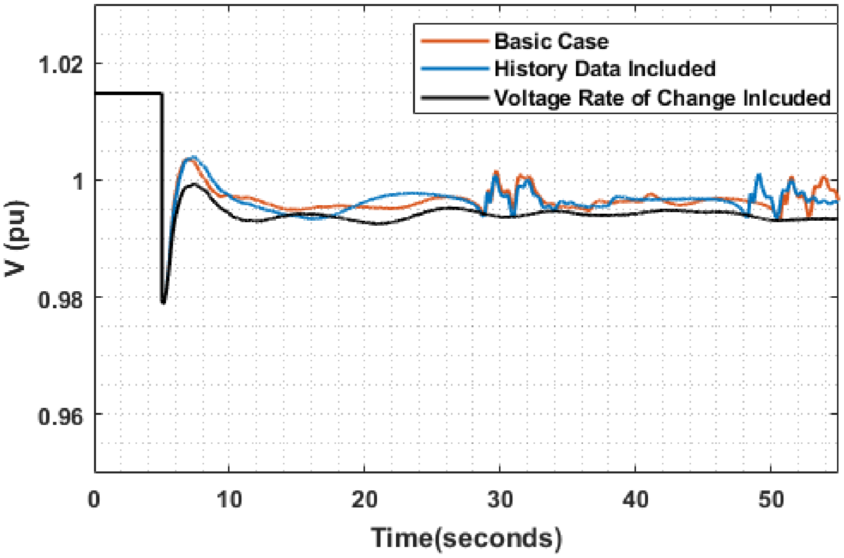

Figure 4.20: Comparison of 2000-bus system test results

in the reward function leads to voltage recovery with a smoother curve, compared to the basic case and the case that includes historical voltage deviation. These two cases exhibit minor voltage oscillations and deviations, which do not exhibit satisfactory dynamic performance, though the voltage level has recovered to the normal level. The reward function, which includes the voltage rate of change can guide the agent to achieve a maximum reward value and mitigate the oscillations, which improves the system's dynamic performance during control.

Case 3: Simulation with randomly selected location of disturbance

A random disturbance at buses 7219, 7306 and 7069 is introduced, respectively. Only voltage magnitude and regulation cost are considered. After the agent is well trained, the test results are shown in Figure 4.21 and Figure 4.22. We can see from the results that the voltage can still be supported with a random disturbance. The voltage is not completely restored to around 1 pu, but it still can help improve the system voltage, which is also very important to the safe operation of the system.

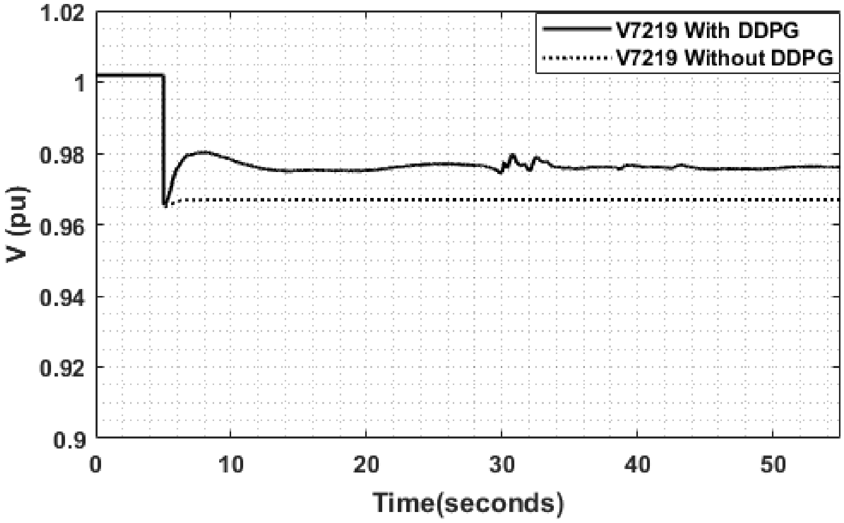

Figure 4.21: Case 3: Voltage of bus 7219 when a disturbance occurs at bus 7219

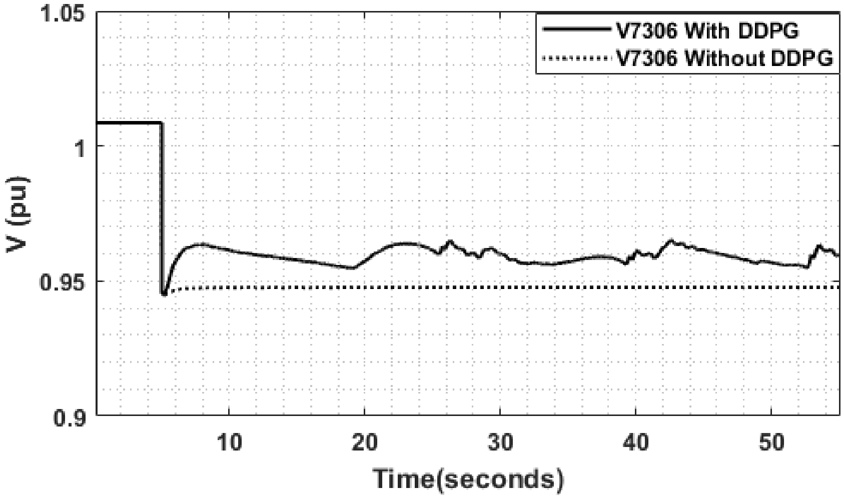

Figure 4.22: Case 3: Voltage of bus 7306 when a disturbance occurs at bus 7306

Case 4: Simulation with more severe disturbance

A larger disturbance with 280MVar load change is introduced into the system to fully test the effectiveness of the proposed controller. The voltage curve compared

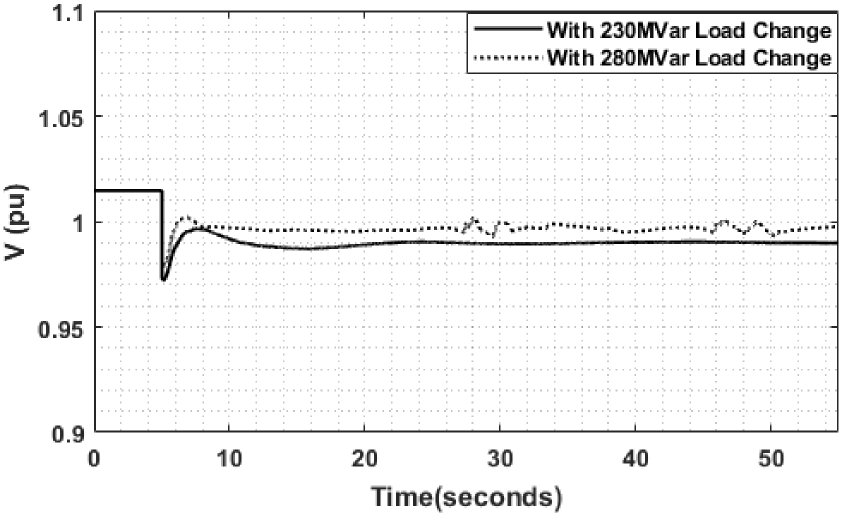

Figure 4.23: Case 4: Voltage compare between Case 1 and Case 4

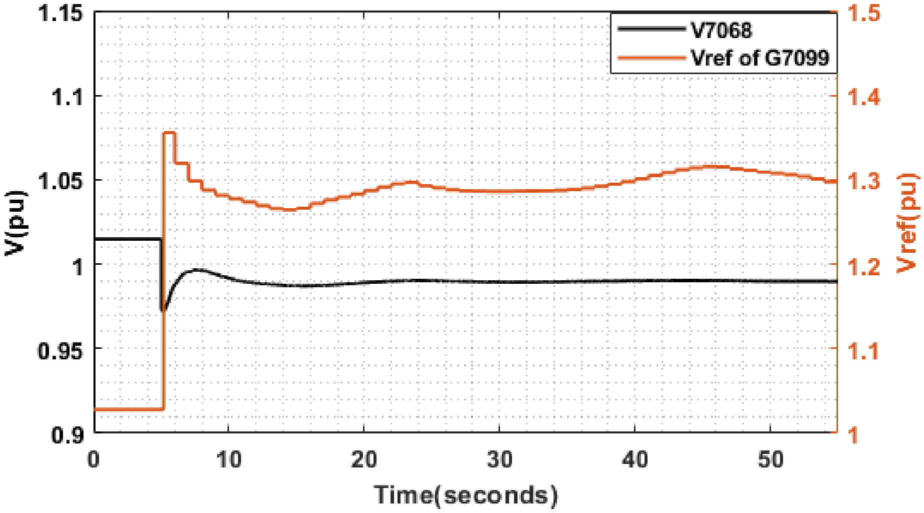

Figure 4.24: Case 4: Voltage reference command of the excitation system for generator 7099

with Case 1 of 230MVar load change can be seen in Figure 4.23, and the red curve is the voltage with 280MVar load change, which decreased more compared to the green curve of Case 1 after the disturbance is added into the system at 5 seconds. The excitation system voltage reference values of the two controlled generators can

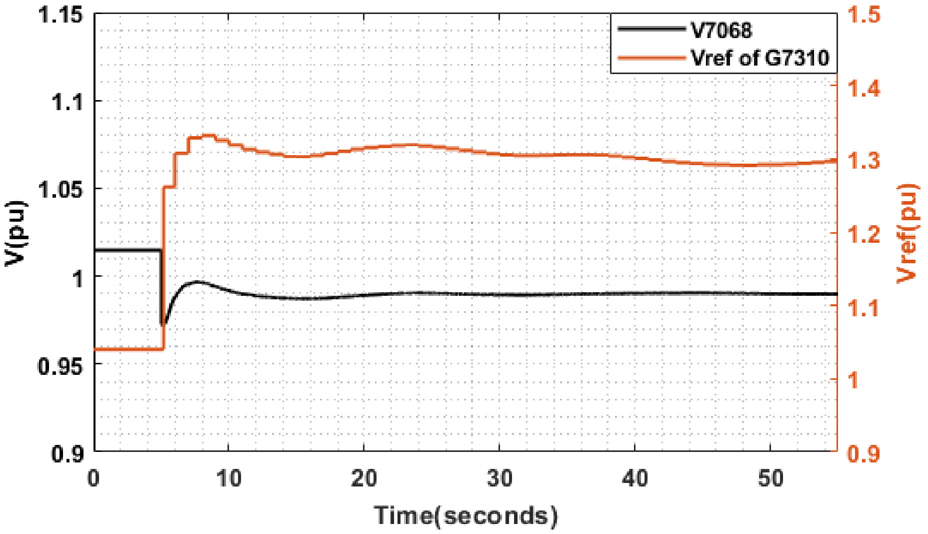

Figure 4.25: Case 4: Voltage reference command of the excitation system for generator 7310

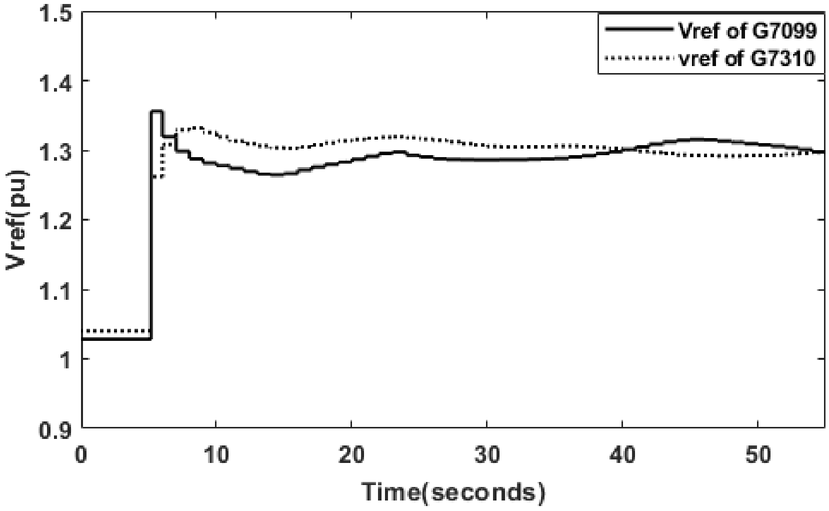

Figure 4.26: Case 4: Trend of the two generators voltage reference command

be seen in Figure 4.24 and Figure 4.25. The voltage curve (green one) is also retained in the figure to make the results more intuitive. It can be observed that after the voltage drop, the controller will output high value commands to make the two generators provide better voltage support to the system after a serious disturbance. The voltage can be improved and can finally reach nearly 1 pu under the control of the controller. The controller will keep monitoring and regulating the system voltage during the whole dynamic simulation. The output commands from the controller are also presented in Figure 4.26, which shows the trends of the whole control time range. From all the results presented, it is seen the controller can effectively provide support and precisely help the system restore voltage.

4.7 Conclusions

This study proposes a DRL-based data-driven excitation control scheme to realize real-time voltage regulations. The voltage control problem is formulated as a Markov Decision Process that considers historical voltage data and the voltage rate of change information besides the voltage deviation and regulation cost, which leads to better dynamic performance during voltage recovery after disturbances. The development of a dynamic simulation training and test platform provides a reliable environment for the training and testing of different scales of systems regarding various control problems based on DRL algorithms. The results show that the proposed DRL-based dynamic voltage control method outperforms conventional voltage control methods in terms of faster and more accurate voltage control without relying on complex system models. The method demonstrates promising dynamic performance and can be readily generalized to large-scale power systems, which has the potential to be applied in practical power systems for real-time voltage control.

Chapter 5

MULTI-AGENT DECENTRALIZED EXECUTION REAL-TIME EXCITATION CONTROL

5.1 Markov Games

The reinforcement learning agent acquires knowledge by engaging with the environment and making sequential decisions through a process of trial and error. Throughout the training, the acquired policy undergoes constant evaluation, guiding the agent to refine its control strategy in the optimal direction. The multi-agent environment could be extended from the Markov Decision Process(MDP) to a Markov game. A Markov game for N agents is a tuple $< \mathbb{N}, \mathbb{S}, \mathbb{A}, \mathbb{R}, \mathbb{P} >$ where $\mathbb{N}$ is a set of agents indexed $1, ..., N$. $\mathbb{S}$ represents state space of N-agents. $\mathbb{A} = [A_1, ..., A_N]$ represents the action space of N-agents. $\mathbb{P} : \mathbb{S} \times A_1 \times ... \times A_N \to \mathbb{S}$ is a stochastic transition function. $\mathbb{R}$ contains stage reward r for each agent and agent N obtains stage reward r_N as a function of the state and action $r_N : S \times A_N \to \Re$ and each agents N wants to maximize its own total expected return by

$$R_N = \mathbb{E}\left[\sum_{t=0}^{+\infty} \gamma^t r_N^t\right] \tag{5.1}$$

5.2 Decentralized Actor Centralized Critic Multi-Agent DDPG

The multi-agent learning algorithm needs to consider the algorithm's robustness. As in Fig. 5.1, the algorithmic framework of the centralized critic with the decentralized actor is shown, where the information from all the critic neural networks can be used to ease the training for each actor agent. Different from centralized DDPG, which shares the same critic and actor neural network, multi-agent DDPG utilizes all

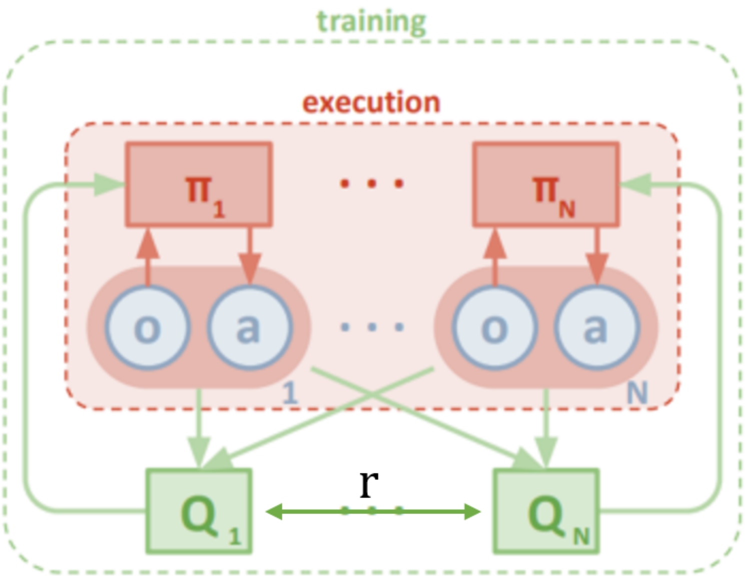

Figure 5.1: The schematic of MADDPG[1]

the observations and actions of all agents that have independent neural networks to update the critic network during the training by:

$$\begin{aligned} \mathcal{L}\left(\theta_N\right) &= \mathbb{E}_{\mathbf{s},a,r,\mathbf{s}'}\left[\left(Q_N^{\pi_N}\left(\mathrm{s}, a_1, \ldots, a_N\right) - y\right)^2\right] \\ y &= r_N + \gamma Q_N^{\pi'_N}\left(\mathrm{s}', a'_1, \ldots, a'_N\right)\Big|_{a'_N=\pi'_N(s_N)} \end{aligned} \tag{5.2}$$

Then, based on the centralized critic network, the decentralized action network is updated by

$$\nabla_{\theta_N} J\left(\theta_N\right) = \mathbb{E}_{\{\mathrm{s},a\}\sim D}[\nabla_{\theta_N} \pi_N\left(a_N \mid s_N\right) Q_N^{\pi_N}\left(s, a_1, \ldots, a_N\right)] \tag{5.3}$$

The training process is detailed in Algorithm 3.

Algorithm 3 Multi-Agent Deep Deterministic Policy Gradient algorithm for Real-time Dynamic Voltage Control

```
input  : power system environment states
output: control action applied to the power system environment

Initialize the critic network $Q$, $Q'$ and actor network $\mu$, $\mu'$ with random weights $\theta$, $\theta' \leftarrow \theta$ and $\phi$, $\phi' \leftarrow \phi$.
Initialize the experience replay buffer $D$.
for episode 1 to $M$, do
    Initialize the environment and obtain initial state $S_0$
    Initialize a random process $N$ for action exploration
    for step 1 to $T$, do
        Select action $a_t = (a_{1t}, ..., a_{Nt})$ according to the current policy and exploration noise
        Execute action $a_t$, observe $r_t$ and next state $s'$  Store transition ( $s_t$, $a_t$, $r_t$, $s'$) in $D$
        for agent 1 to $N$, do
            Sample a random minibatch of $B$ transition ( $s_j$, $a_j$, $r_j$, $s'$) from $D$
            Compute the critic target:
                $\mathrm{y}{=}\mathrm{r}_N + \gamma Q_N^{\pi_N'}\left(\mathrm{s}', a_1', ..., a_N'\right)\Big|_{a_N'=\pi_N'(o_N)}$
            Update the critic Q-function by gradient descent using:
                $\mathrm{L}(\theta_N) = \mathbb{E}_{\mathbf{s},a,r,\mathbf{s}'}\left[\left(Q_N^{\pi_N}\left(\mathrm{s}, a_1, ..., a_N\right) - y\right)^2\right]$
            Update the target networks as:
                $\nabla_{\theta_N} J\left(\theta_N\right) = \mathbb{E}_{\{\mathrm{s},a\}\sim D}[\nabla_{\theta_N} \pi_N\left(a_N \mid o_N\right) Q_N^{\pi_N}\left(s, a_1, ..., a_N\right)]$
        Update the network parameters:
                $\theta' \leftarrow \rho\theta + (1-\rho)\theta'$,
                $\phi' \leftarrow \rho\phi + (1-\rho)\phi'$
```

5.3 Definition of Action, State and Reward

5.3.1 Definition of Action and State

Similar to Chapter 4, this study also adopts bus voltage magnitudes as the observation states in the Markov game since voltage stability is the problem we are discussing.

The control actions are defined as the excitation system voltage reference values of the controlled generators. Different from the centralized DDPG, in which the actions are defined as one vector of multiple generators but share the same neural network, each control action of the multi-agent DDPG has an independent actor neural network. Each generator is controlled by an actor to output the action, which means every action is an independent vector within its own actor neural network.

5.3.2 Definition of Reward

The reward function r_t has the same structure as in Chapter 4, as shown in (5.4). The voltage variation, the action regulation amount, the history voltage data, and the rates of voltage changes are considered in the reward function to accommodate both the final voltage regulation level and the dynamic performance of the controller.

$$r_t = \begin{cases} Huge\ penalty, \quad power\ flow\ diverges \\ -c_1 * \sum\limits_i \Delta v_i(t) - c_2 * \sum\limits_j \Delta a_j(t) - \\ \qquad c_3 * \sum\limits_{t-c_t}^{t} \sum\limits_i \Delta v_{h-i}(t) - \\ c_4 * \sum\limits_{t-c_t}^{t-\Delta t} \sum\limits_i \dfrac{v_{h-i}(t) - v_{hi-i}(t-\Delta t)}{\Delta t}, \ otherwise \end{cases} \tag{5.4}$$

$$\Delta v_i(t) = |v_i(t) - V_{ref}| \tag{5.5}$$

$$\Delta a_j(t) = |a_j(t) - a_{ref}| \tag{5.6}$$

where c_1, c_2 and c_3 are the weights of each part. The definition of Δv and Δa is given in (5.5)-(5.6). Δt is the time interval of every learning step in the training process.

5.4 Simulation and Results

The test systems are the IEEE 9-bus system and the 2000-bus Texas synthetic grid systems as well, which is the same as Chapter 4. All the case studies, including training and testing, were performed in the simulation environment based on the platform described in Section IV.

5.4.1 Simulation Results of IEEE 9-Bus System

Case 1: Voltage Control Performance

The agent is trained and Figure 5.2 illustrates that the moving average reward ultimately reaches a satisfactory level after 2000 training episodes. This indicates that the performance of the agents, evaluated using the designed reward function, is commendable. Generators 2 and 3 are under the control of different agents, with each agent possessing independent critic and actor neural networks to execute control actions. A disturbance is introduced at 5 seconds to induce voltage changes.

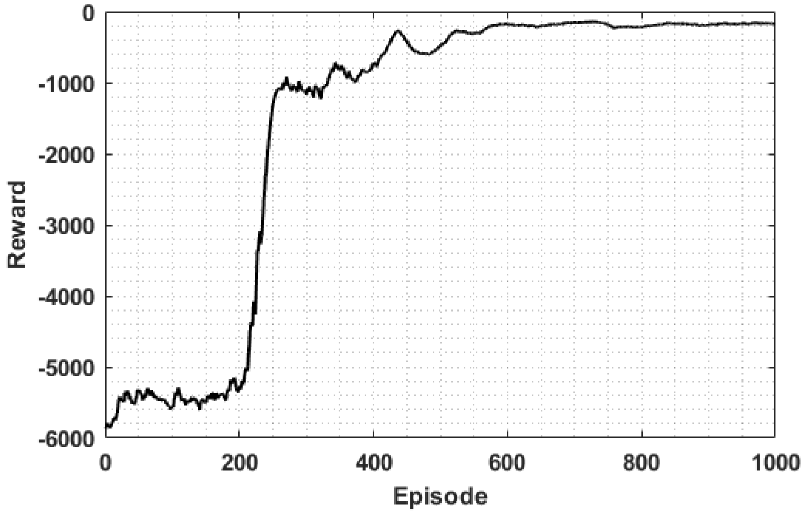

Figure 5.2: Training average reward of IEEE 9-bus system.

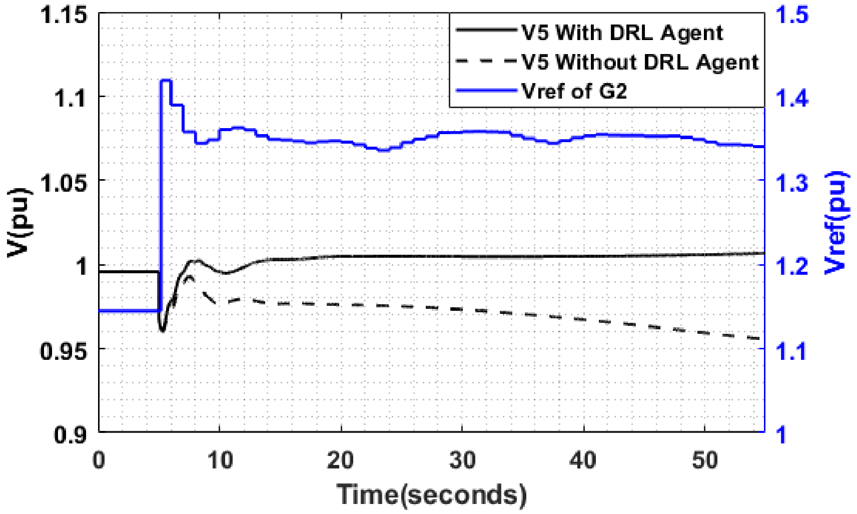

Figure 5.3: Case 1 of IEEE 9-bus system: Voltage of bus 5 and Vref command for Generator 2.

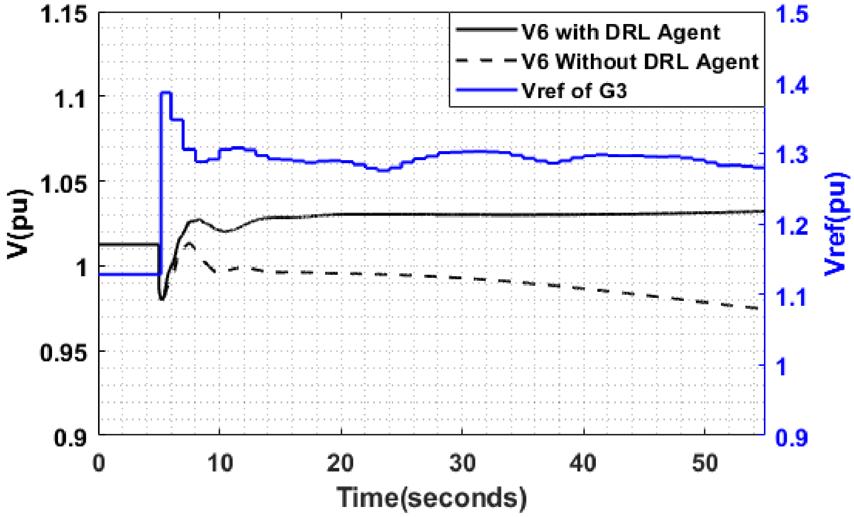

Figure 5.4: Case 1 of IEEE 9-bus system: Voltage of bus 6 and Vref command for Generator 3.

The test results depict the system's response to a 90 MVar reactive power load increase in the IEEE 9-bus system, as illustrated in Figure 5.3 and Figure 5.4. A comparison of system performance with and without the MADDPG agents is presented.

In the scenario where only a constant voltage reference is applied in the excitation system, represented by the dotted line in Figure 5.3 and Figure 5.4, the system experiences a decrease in voltages following oscillations. Moreover, it fails to recover to the normal level after the disturbance.

However, when MADDPG agents are introduced, a significant increase in voltage reference commands is observed as the system voltage decreases during the disturbance. The controlled generators respond dynamically, providing crucial support that leads to an improvement in voltages. Subsequently, the controlled generators continue to regulate the output in real time, effectively maintaining the voltage level around one pu.

The results demonstrate that the MADDPG controller effectively enhances and sustains system voltage during disturbances, showcasing its capability to respond and adapt to challenging voltage disturbance conditions.

Case 2: With Time-Varying Load Changes

To conduct a thorough evaluation of control performance in the face of dynamic load fluctuations, we introduce varying load changes following the initial disturbance in the dynamic simulation. In particular, after the initial 50 MVar load change, an additional 30 MVar load change at 20 seconds and a subsequent 20 MVar load change at 35 seconds are introduced. This sequence of load variations provides a comprehensive scenario to assess the system's resilience and the effectiveness of the control mechanisms in adapting to evolving and dynamic conditions.

The test results illustrating bus voltage and MADDPG control commands can be observed in Figure 5.5 and Figure 5.6. With time-varying disturbances, the system voltage remains supported consistently under the control of MADDPG agents, effectively responding to each disturbance event. The agents exhibit prompt and precise detection of system voltages, enabling timely adjustments to maintain a stable voltage level.

In contrast, the scenario with constant voltage reference commands fails to support system voltage effectively; the voltage continuously decreases, posing a threat to the stability of the system. The comparison underscores the performance of the MADDPG-controlled system, emphasizing its capacity to respond accurately to dynamic conditions, which decreases the risk of voltage instability under varying disturbances.

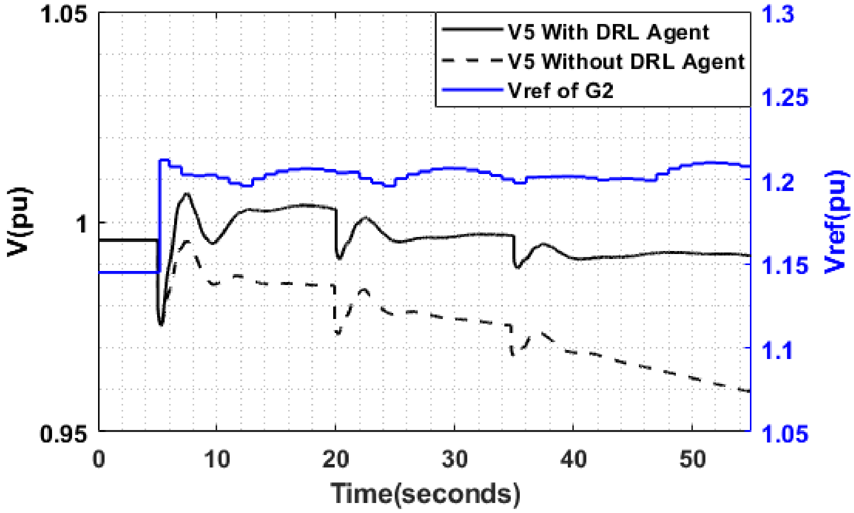

Figure 5.5: Case 2 of IEEE 9-bus system: Voltage of bus 5 and Vref command for Generator 2.

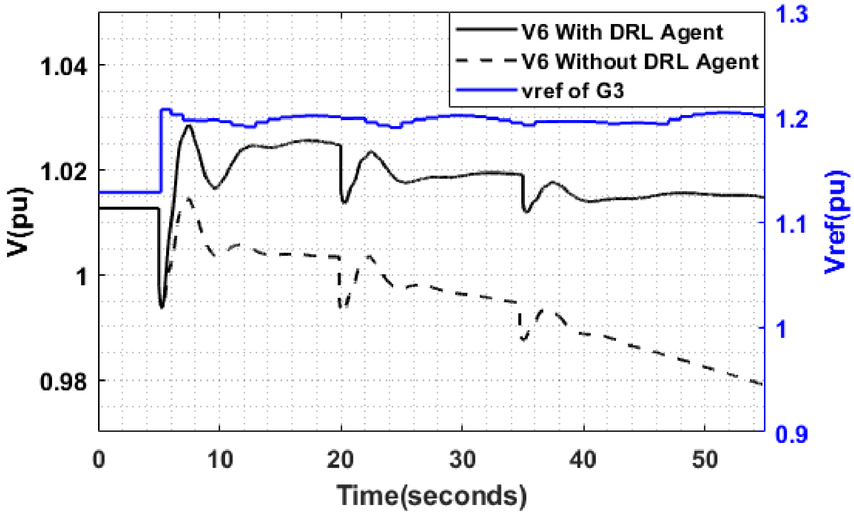

Figure 5.6: Case 2 of IEEE 9-bus system: Voltage of bus 6 and Vref command for Generator 3.

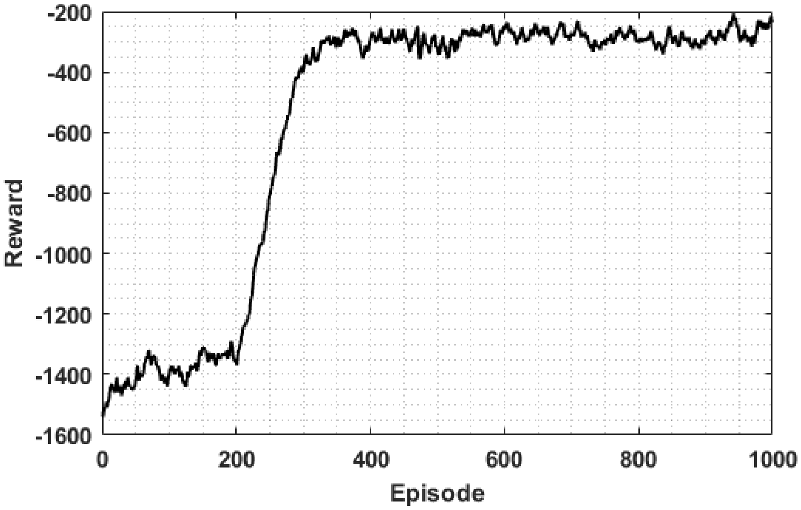

Figure 5.7: Training average reward of Texas 2000-bus system.

5.4.2 Simulation Results of Texas 2000-Bus Synthetic Test System

Simulations are executed on the Texas 2000-bus synthetic power system to assess the efficacy of the multi-agent DDPG controller in handling larger systems. Three distinct scenarios, encompassing instances of communication failure, are thoroughly examined to demonstrate the robustness and effectiveness of the proposed control policy.

Case 1: Voltage Control Performance

The voltage control performance is evaluated in this case. Firstly, the multi-agent DDPG agents are trained, and the result is shown in Figure 5.7, where the reward can finally stabilize at a high level. Then, the controller is implemented into the Texas 2000-bus system with load disturbance. The results can be seen in Figure 5.8 to Figure 5.9.

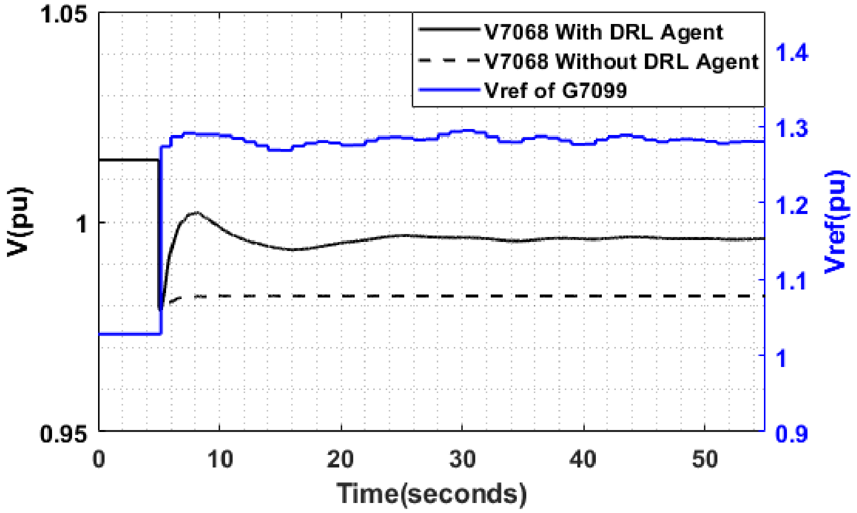

Figure 5.8: Case 1 of 2000-bus system: Voltage of bus 7068 and Vref command for Generator 7099.

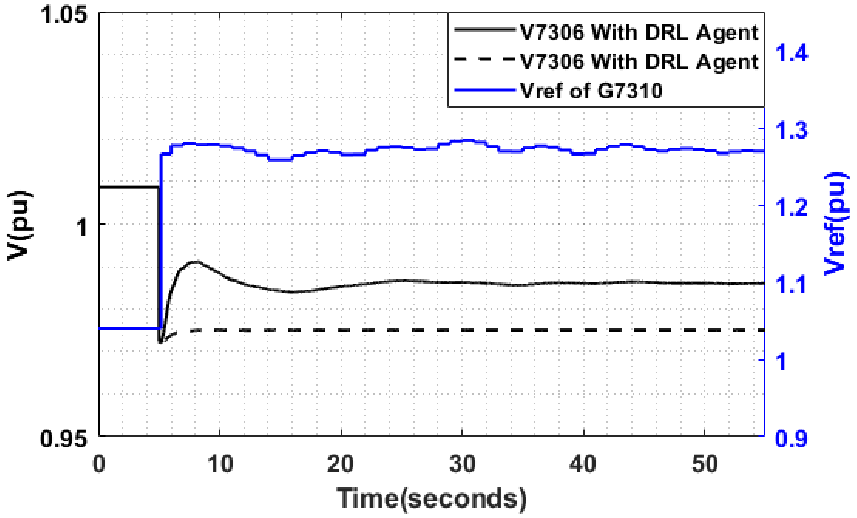

Figure 5.9: Case 1 of 2000-bus system: Voltage of bus 7306 and Vref command for Generator 7310.

The results demonstrate that both generators, each under the control of distinct agents, can effectively respond to the voltage drop during disturbances to provide

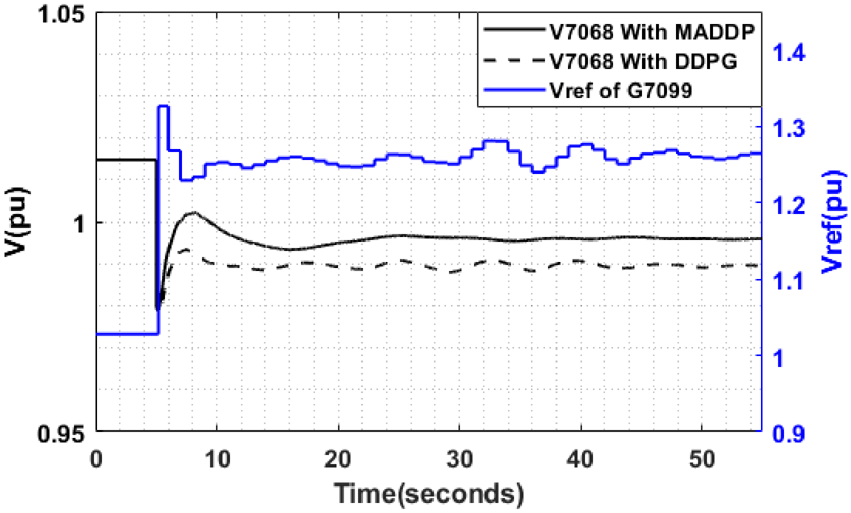

Figure 5.10: Case 2 of 2000-bus system: Voltage of bus 7068 and Vref command for Generator 7099.

timely support. Within just 2 seconds after the fault, the voltage is restored to approximately 1 per unit, and this level is consistently maintained during the system operation. The two controlled generators keeps regulating the system voltage in real-time, precisely adjusting the voltage reference values generated by the multi-agent DDPG agents during system operation. The voltage recovery is achieved without unnecessary fluctuations. In contrast to conventional excitation system control with a constant voltage reference, the multi-agent DDPG controller significantly enhances the voltage level. The proposed control strategy not only ensures a rapid recovery from disturbances but also maintains a stable voltage level, highlighting its superiority in improving system performance.

Case 2: With Communication Failure in One Generator

The multi-agent DDPG control adopts centralized training and decentralized execution in its control structure. This signifies that each distributed agent generates

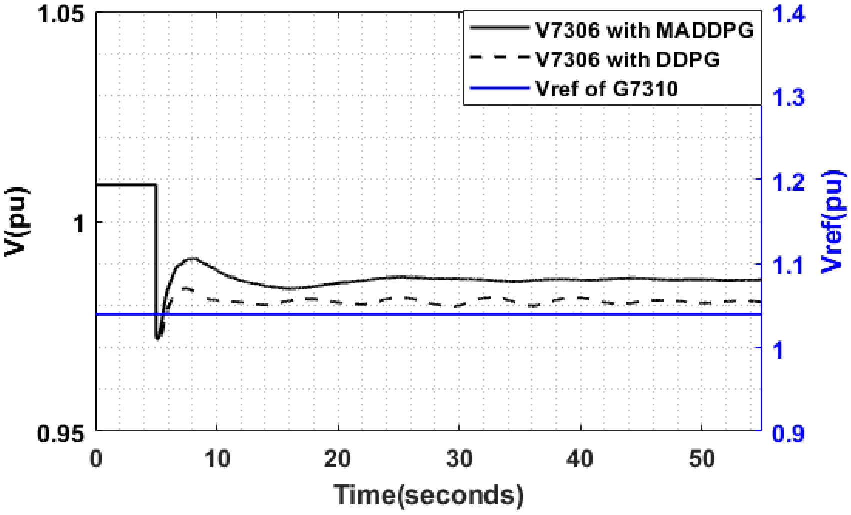

Figure 5.11: Case 2 of 2000-bus system: Voltage of bus 7306 and Vref command for Generator 7310.

control commands based only on local information after the agent is well-trained. Simulations conducted under both multi-agent DDPG and centralized DDPG are visually represented in Figure 5.10 through Figure 5.11, providing a comparative analysis of their performance under distinct control settings.

In the multi-agent DDPG configuration, independent critic and actor neural networks are employed for each agent. Conversely, the centralized DDPG utilizes shared critic and actor neural networks across all agents. A noteworthy scenario is introduced where a communication failure occurs in the actor neural network responsible for controlling the generator G7310, leading to the inability to transmit critic neural network information to the actor during the control process.

In Figure 5.10 and Figure 5.11, the solid black curves represent the system voltage under the control of the multi-agent DDPG, while the dotted curves depict the system voltage when controlled by the centralized DDPG. Additionally, the blue curve represents the reference voltage command generated by the centralized DDPG controller. When a communication failure occurs between the actor and the critic in the centralized DDPG configuration, the affected actor lacks real-time system status information, rendering it unable to output commands normally. Consequently, the voltage reference for the G7310 excitation system remains unchanged and fails to adapt to system voltage fluctuations, as shown in Figure 5.11.

In contrast, the multi-agent DDPG continues to operate normally despite the communication failure. Relying only on local information, it produces control commands unaffected by the disrupted communication link. Consequently, the system's voltage is effectively recovered and maintained. With the centralized DDPG controller now only controlling G7099, the voltage support is constrained, leading to a lower voltage level, as illustrated in Figure 5.10 and Figure 5.11.

This comparison underscores the resilience of the multi-agent DDPG approach, showcasing its ability to enhance control system robustness during communication failures. The decentralized execution with centralized training proves advantageous in maintaining effective control even in communication failure scenarios, thereby highlighting the robustness of the proposed multi-agent DDPG architecture.

Case 3: With Communication Failure in One Generator at 20s

To assess controller performance in the face of communication failures during system operation, simulations involving communication failure of agents controlling the generator G7310 at 20 seconds are conducted. The outcomes of both multi-agent DDPG and centralized DDPG are depicted in Figure 5.12 and Figure 5.13.

Upon a load change disturbance at 5 seconds, both multi-agent DDPG and centralized DDPG promptly respond to the voltage drop, providing system support that increases bus voltages. However, when communication failure is induced at 20 seconds, the centralized DDPG agent stops to output valid commands, causing the

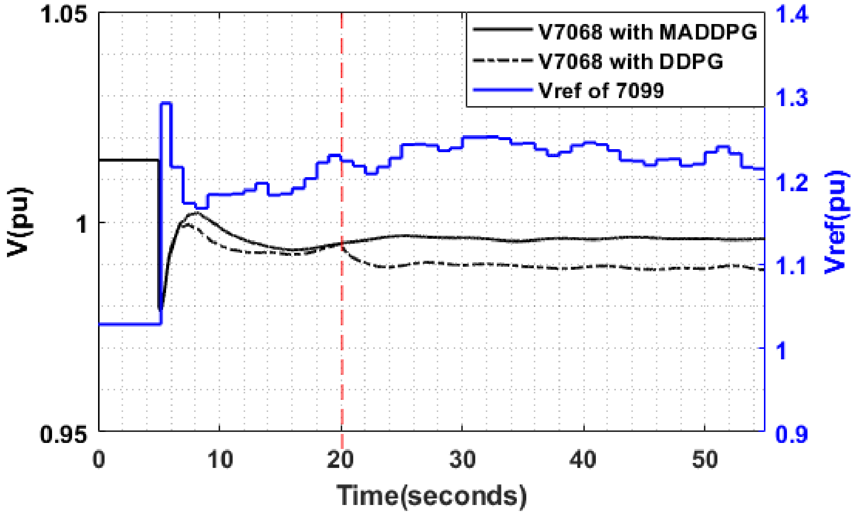

Figure 5.12: Case 3 of 2000-bus system: Voltage of bus 7068 and Vref command for Generator 7099.

voltage reference for G7310 to revert to its default value, the same with the system's initial setting, as shown in Figure 5.13. With only G7099 fully controlled by the DDPG agent, there is a modest drop in bus voltages due to limited voltage support.

In contrast, the multi-agent DDPG controller remains unaffected by the communication failure, continuously delivering sustained voltage support throughout the control process. It's noteworthy that even in the absence of communication failure, the multi-agent DDPG exhibits superior voltage support. The voltage level is higher than the centralized DDPG control from 7 seconds to 16 seconds, as evidenced by the results in both Figure 5.12 and Figure 5.13. This underscores the inherent robustness and efficacy of the multi-agent DDPG approach, not only in handling communication failures but also in consistently providing stronger voltage support under normal operating conditions.

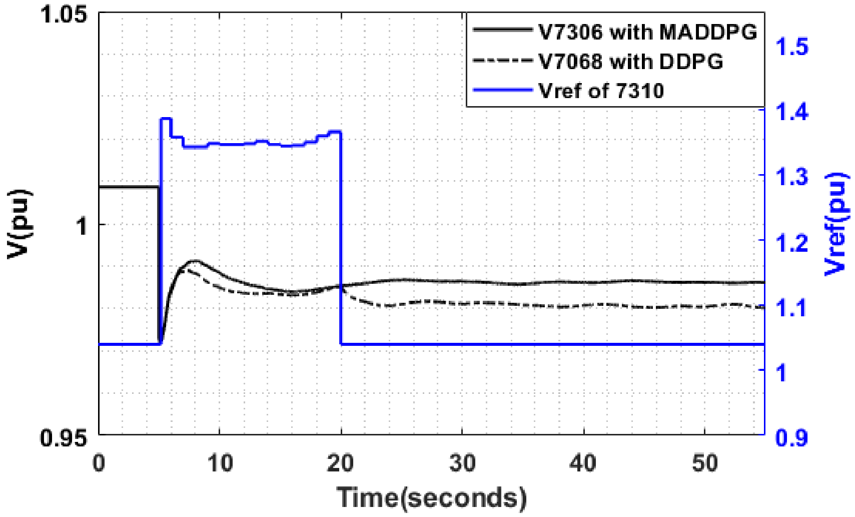

Figure 5.13: Case 3 of 2000-bus system: Voltage of bus 7306 and Vref command for Generator 7310.

5.5 Conclusions

This study introduces a dynamic voltage control method, leveraging the multi-agent DDPG algorithm, which operates on the principle of centralized training and decentralized execution. In this approach, each agent is equipped with independent actor neural networks responsible for generating generator control commands and critic neural networks that assess the performance of these commands. After training, each agent is capable of independently generating control commands using only local information. Simulation results underscore the effectiveness of the multi-agent DDPG controller, demonstrating its proficiency not only in offering voltage support but also in adeptly managing communication failures among distinct agents. This approach showcases the system's adaptability and robustness, emphasizing its potential for enhancing dynamic voltage control in power systems.

Chapter 6

CONCLUSION AND FUTURE WORK

This research work deals with voltage control problems both in steady state and dynamic control processes based on DRL method.

In the first approach, a voltage control strategy based on DDPG for managing changes in load within a power system is proposed. The approach takes into account multiple voltage control devices, and it seamlessly integrates both continuous and discrete actions to demonstrate its adaptability across various device types. Through extensive training, the DDPG-based agent exhibits remarkable resilience in rectifying voltage violations under diverse operating conditions. This novel approach effectively harnesses the various available reactive power resources, each with its unique response characteristics, to enhance the dependability of voltage support.

The second control method is used for dynamic simulation. An innovative data-driven excitation control approach based on Deep Reinforcement Learning (DRL) to achieve real-time voltage regulation is proposed. The voltage control challenge by formulating is addressed as a Markov Decision Process, incorporating historical voltage data, voltage rate of change information, voltage deviation, and regulation cost. This comprehensive approach enhances dynamic performance during voltage recovery following disturbances. To facilitate experimentation and evaluation across various system scales and control scenarios using DRL algorithms, we have developed a dynamic simulation training and testing platform. The results demonstrate that the proposed DRL-based dynamic voltage control method surpasses conventional methods in terms of both control speed and accuracy, bypassing the need for intricate system models. This method exhibits promising dynamic performance and holds the potential for widespread adoption in practical, real-time voltage control applications for large-scale power systems.

The third approach leverages the Multi-agent DDPG algorithm, which operates by centralized training and decentralized execution. In this approach, each agent is equipped with independent actor neural networks responsible for generating generator control commands and critic neural networks that assess the performance of these commands. After training, each agent is capable of independently generating control commands using only local information. Simulation results underscore the effectiveness of the Multi-agent DDPG controller, demonstrating its proficiency not only in offering voltage support but also in adeptly managing communication failures among distinct agents.

To further develop the ideas and approaches that have been presented in this work, some of the potential research areas could be improved as follows:

- The expansive potential of the deep reinforcement learning controller allows for the enlargement of both control scope and action diversity, facilitating the inclusion of a wider array of controlled devices. Moreover, the neural network's capabilities can be augmented, empowering it to effectively process and manage a more substantial volume of information, thereby enhancing its overall performance and adaptability. This opens up possibilities for a sophisticated and versatile system that can efficiently control an extensive range of devices.

- DDPG stands out as a sophisticated deep reinforcement learning algorithm characterized by its accommodation of continuous action spaces. However, it exhibits sensitivity to the training dataset, leading to increased training requirements.

  To mitigate this challenge, it's worthwhile to explore alternative deep reinforcement learning algorithms. Notably, algorithms like Proximal Policy Optimization (PPO) and Soft Actor-Critic (SAC) offer enhanced training stability and should be considered as promising alternatives.

REFERENCES

[1] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, “Multi-agent actor-critic for mixed cooperative-competitive environments,” *Advances in neural information processing systems*, vol. 30, 2017.

[2] Y. Wang, C. Chen, J. Wang, and R. Baldick, “Research on resilience of power systems under natural disasters—a review,” *IEEE Transactions on Power Systems*, vol. 31, no. 2, pp. 1604–1613, 2016.

[3] J. Jasiūnas, P. D. Lund, and J. Mikkola, “Energy system resilience – a review,” *Renewable and Sustainable Energy Reviews*, vol. 150, p. 111476, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032121007577

[4] M. Glavic, “(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives,” *Annual Reviews in Control*, vol. 48, pp. 22–35, 2019.

[5] P. S. Kundur and O. P. Malik, *Power System Stability and Control.* McGraw-Hill, 2022.

[6] J. Machowski, J. W. Bialek, and J. R. Bumby, *Power system dynamics and stability.* John Wiley & Sons, 1997.

[7] Z. Zhang and M. Wu, “Predicting real-time locational marginal prices: A gan-based video prediction approach,” 2020.

[8] L. Strezoski, H. Padullaparti, F. Ding, and M. Baggu, “Integration of utility distributed energy resource management system and aggregators for evolving distribution system operators,” *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 277–285, 2022.

[9] Z. Zhang and M. Wu, “Energy price prediction considering generation bids variation: A two-stage convolutional long short-term memory approach,” in *2022 IEEE Power Energy Society General Meeting (PESGM)*, 2022, pp. 1–5.

[10] M. Khan, “Robust load frequency control and integration of electric vehicles and renewable energy in the grid,” *IET Conference Proceedings*, pp. 141 (6 pp.)–141 (6 pp.)(1), January 2019. [Online]. Available: https://digital-library.theiet.org/content/conferences/10.1049/cp.2019.0397

[11] Z. Zhang and M. Wu, “Locational marginal price forecasting using convolutional long-short term memory-based generative adversarial network,” in *2021 IEEE Power Energy Society General Meeting (PESGM)*, 2021, pp. 1–5.

[12] S. Saha, M. Saleem, and T. Roy, “Impact of high penetration of renewable energy sources on grid frequency behaviour,” *International Journal of Electrical Power Energy Systems*, vol. 145, p. 108701, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061522006974

[13] B. Tan, J. Zhao, M. Netto, V. Krishnan, V. Terzija, and Y. Zhang, "Power system inertia estimation: Review of methods and the impacts of converter-interfaced generations," *International Journal of Electrical Power Energy Systems*, vol. 134, p. 107362, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061521006013

[14] R. Shah, N. Mithulananthan, R. Bansal, and V. Ramachandaramurthy, "A review of key power system stability challenges for large-scale pv integration," *Renewable and Sustainable Energy Reviews*, vol. 41, pp. 1423–1436, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032114008004

[15] M. Migliori, S. Lauria, L. Michi, G. Donnini, B. Aluisio, and C. Vergine, "Renewable sources integration using hvdc in parallel to ac traditional system: the adriatic project," in *2019 AEIT HVDC International Conference (AEIT HVDC)*, 2019, pp. 1–5.

[16] Y. Wang, K. li, D. Shao, Y. Xu, H. Sun, W. Zhang, X. Shen, and X. Tang, "Sensitivity based optimized voltage control strategy for power grid with uhvdc feed in," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 2698–2704.

[17] W. Zhang, D. Shao, Y. Xu, K. Li, H. Sun, Y. Wang, X. Shen, and X. Tang, "Extinction angle modulation in uhvdc systems for improving voltage stability of weak ac power system," in *2018 International Conference on Power System Technology (POWERCON)*, 2018, pp. 2705–2712.

[18] Y. Wang, D. Shao, Y. Xu, X. Shen, C. Duan, Y. Wang, and H. Sun, "Cooperated control strategy of generator re-dispatching and multi-hvdc modulation after ultra hvdc block," *The Journal of Engineering*, vol. 2019, no. 16, pp. 1299–1305, 2019. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/joe.2018.8767

[19] M. E. Initiative *et al.*, "Managing large-scale penetration of intermittent renewables."," 2012.

[20] Y. Wang, W. Zhang, H. Sun, Y. Xiang, D. Shi, and Z. Wang, "Research on fast response criterion of power grid distributed loads after hvdc block fault," *IET Generation, Transmission & Distribution*, vol. 14, no. 25, pp. 6230–6238, 2020. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/iet-gtd.2020.0852

[21] T. Krause, R. Ernst, B. Klaer, I. Hacker, and M. Henze, "Cybersecurity in power grids: Challenges and opportunities," *Sensors*, vol. 21, no. 18, 2021. [Online]. Available: https://www.mdpi.com/1424-8220/21/18/6225

[22] J. De La Ree, V. Centeno, J. S. Thorp, and A. G. Phadke, "Synchronized phasor measurement applications in power systems," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 20–27, 2010.

[23] H.-Y. Su and H.-H. Hong, “An intelligent data-driven learning approach to enhance online probabilistic voltage stability margin prediction,” *IEEE Transactions on Power Systems*, vol. 36, no. 4, pp. 3790–3793, 2021.

[24] T. Guo and J. V. Milanović, “Online identification of power system dynamic signature using pmu measurements and data mining,” *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1760–1768, 2016.

[25] S. J. S. J. Russell, P. Norvig, and E. Davis, *Artificial intelligence : a modern approach*, 3rd ed., ser. Prentice Hall series in artificial intelligence. Upper Saddle River, NJ: Prentice Hall, 2010 - 2010.

[26] Y. Wang, V. Vittal, X. Luo, S. Maslennikov, Q. Zhang, M. Hong, and S. Zhang, “Reinforcement learning based voltage control using multiple control devices,” in *2023 IEEE Power Energy Society General Meeting (PESGM)*, 2023, pp. 1–5.

[27] Y. Wang and V. Vittal, “Real-time excitation control-based voltage regulation using ddpg considering system dynamic performance,” *IEEE Open Access Journal of Power and Energy*, vol. 10, pp. 643–653, 2023.

[28] E. Alpaydin, *Introduction to machine learning*, ser. Adaptive computation and machine learning. Cambridge, Mass: MIT Press, 2004.

[29] “Scada: Supervisory control and data acquisition, 3d ed,” *SciTech Book News*, vol. 28, no. 3, pp. 135–, 2004.

[30] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, ser. Adaptive computation and machine learning. Cambridge, Massachusetts: The MIT Press, 2016 - 2016.

[31] S. Thomas, A. Oommen Philip, and N. Vishwanath, “Ml based data driven energy centered predictive maintenance,” in *2023 2nd International Conference on Edge Computing and Applications (ICECAA)*, 2023, pp. 994–1001.

[32] Z. Zhang and M. Wu, “Real-time locational marginal price forecasting using generative adversarial network,” in *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2020, pp. 1–6.

[33] M. Van Otterlo and M. Wiering, “Reinforcement learning and markov decision processes,” in *Reinforcement learning: State-of-the-art.* Springer, 2012, pp. 3–42.

[34] Y. Li, “Deep reinforcement learning: An overview,” *arXiv preprint arXiv:1701.07274*, 2017.

[35] “IEEE standard requirements for tap changers,” *IEEE Std C57.131-2012 (Revision of IEEE Std C57.131-1995)*, pp. 1–73, 2012.

[36] S. Sithole, N. Mbuli, and J. Pretorius, “Voltage regulation in the douglas area using shunt capacitor banks and controllable shunt reactors,” in *2013 13th International Conference on Environment and Electrical Engineering (EEEIC)*, 2013, pp. 85–90.

[37] H. Bourles, S. Peres, T. Margotin, and M. Houry, “Analysis and design of a robust coordinated avr/pss,” *IEEE Transactions on Power Systems*, vol. 13, no. 2, pp. 568–575, 1998.

[38] I. Hiskens and D. Hill, “Incorporation of svcs into energy function methods,” *IEEE Transactions on Power Systems*, vol. 7, no. 1, pp. 133–140, 1992.

[39] O. Gomis-Bellmunt, J. Sau-Bassols, E. Prieto-Araujo, and M. Cheah-Mane, “Flexible converters for meshed hvdc grids: From flexible ac transmission systems (facts) to flexible dc grids,” *IEEE Transactions on Power Delivery*, vol. 35, no. 1, pp. 2–15, 2020.

[40] H. Wang, L. Qu, and W. Qiao, “Adjustable-voltage-ratio magneto-electric transformer,” *IEEE Magnetics Letters*, vol. 6, pp. 1–4, 2015.

[41] B. A. Robbins, H. Zhu, and A. D. Domínguez-García, “Optimal tap setting of voltage regulation transformers in unbalanced distribution systems,” *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 256–267, 2016.

[42] M. Bazrafshan, N. Gatsis, and H. Zhu, “Optimal tap selection of step-voltage regulators in multi-phase distribution networks,” in *2018 Power Systems Computation Conference (PSCC)*, 2018, pp. 1–7.

[43] P. M. S. Carvalho, P. F. Correia, and L. A. F. M. Ferreira, “Distributed reactive power generation control for voltage rise mitigation in distribution networks,” *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 766–772, 2008.

[44] D. Tziouvaras, P. McLaren, G. Alexander, D. Dawson, J. Esztergalyos, C. Fromen, M. Glinkowski, I. Hasenwinkle, M. Kezunovic, L. Kojovic, B. Kotheimer, R. Kuffel, J. Nordstrom, and S. Zocholl, “Mathematical models for current, voltage, and coupling capacitor voltage transformers,” *IEEE Transactions on Power Delivery*, vol. 15, no. 1, pp. 62–72, 2000.

[45] M. Farivar, C. R. Clarke, S. H. Low, and K. M. Chandy, “Inverter var control for distribution systems with renewables,” in *2011 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, 2011, pp. 457–462.

[46] V. Kekatos, G. Wang, A. J. Conejo, and G. B. Giannakis, “Stochastic reactive power management in microgrids with renewables,” *IEEE Transactions on Power Systems*, vol. 30, no. 6, pp. 3386–3395, 2015.

[47] H. Zhu and H. J. Liu, “Fast local voltage control under limited reactive power: Optimality and stability analysis,” *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3794–3803, 2016.

[48] V. Kekatos, L. Zhang, G. B. Giannakis, and R. Baldick, "Voltage regulation algorithms for multiphase power distribution grids," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3913–3923, 2016.

[49] W. Lin, R. Thomas, and E. Bitar, "Real-time voltage regulation in distribution systems via decentralized pv inverter control," 01 2018.

[50] Y. Zhang, M. Hong, E. Dall'Anese, S. V. Dhople, and Z. Xu, "Distributed controllers seeking ac optimal power flow solutions using admm," *IEEE Transactions on Smart Grid*, vol. 9, no. 5, pp. 4525–4537, 2018.

[51] H. Li, Y. Xue, M. Chen, and Z. Zhang, "Analysis of voltage characteristics for single-phase line break fault in resonant grounding systems," *IEEE Transactions on Power Delivery*, vol. 38, no. 2, pp. 1416–1425, 2023.

[52] H. Lomei, D. Sutanto, K. M. Muttaqi, and A. Alfi, "An optimal robust excitation controller design considering the uncertainties in the exciter parameters," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4171–4179, 2017.

[53] T. F. Orchi, T. K. Roy, M. A. Mahmud, and A. M. T. Oo, "Feedback linearizing model predictive excitation controller design for multimachine power systems," *IEEE Access*, vol. 6, pp. 2310–2319, 2018.

[54] C. Zhu, R. Zhou, and Y. Wang, "A new decentralized nonlinear voltage controller for multimachine power systems," *IEEE Transactions on Power Systems*, vol. 13, no. 1, pp. 211–216, 1998.

[55] Y. Guo, D. Hill, and Y. Wang, "Global transient stability and voltage regulation for power systems," *IEEE Transactions on Power Systems*, vol. 16, no. 4, pp. 678–688, 2001.

[56] H. Liu, Z. Hu, and Y. Song, "Lyapunov-based decentralized excitation control for global asymptotic stability and voltage regulation of multi-machine power systems," *IEEE Transactions on Power Systems*, vol. 27, no. 4, pp. 2262–2270, 2012.

[57] H. Liu, J. Su, J. Qi, N. Wang, and C. Li, "Decentralized voltage and power control of multi-machine power systems with global asymptotic stability," *IEEE Access*, vol. 7, pp. 14 273–14 282, 2019.

[58] J. Duan, D. Shi, R. Diao, H. Li, Z. Wang, B. Zhang, D. Bian, and Z. Yi, "Deep-reinforcement-learning-based autonomous voltage control for power grid operations," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 814–817, 2020.

[59] S. Wang, J. Duan, D. Shi, C. Xu, H. Li, R. Diao, and Z. Wang, "A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4644–4654, 2020.

[60] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.

[61] X. S. Zhang, T. Yu, Z. N. Pan, B. Yang, and T. Bao, "Lifelong learning for complementary generation control of interconnected power grids with high-penetration renewables and evs," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4097–4110, 2018.

[62] Z. Zhang and M. Wu, "Real-time locational marginal price forecast: A decision transformer-based approach," in *2023 IEEE Power Energy Society General Meeting (PESGM)*, 2023, pp. 1–5.

[63] Z. Zhang and R. Yang, "High-resolution synthetic solar irradiance sequence generation: An lstm-based generative adversarial network," in *2023 IEEE Power Energy Society General Meeting (PESGM)*, 2023, pp. 1–5.

[64] Z. Zhang and M. Wu, "Predicting real-time locational marginal prices: A gan-based approach," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1286–1296, 2022.

[65] J. Vlachogiannis and N. Hatziargyriou, "Reinforcement learning for reactive power control," *IEEE Transactions on Power Systems*, vol. 19, no. 3, pp. 1317–1325, 2004.

[66] Y. Xu, W. Zhang, W. Liu, and F. Ferrese, "Multiagent-based reinforcement learning for optimal reactive power dispatch," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 6, pp. 1742–1751, 2012.

[67] D. O'Neill, M. Levorato, A. Goldsmith, and U. Mitra, "Residential demand response using reinforcement learning," in *2010 First IEEE International Conference on Smart Grid Communications*, 2010, pp. 409–414.

[68] Z. Wen, D. O'Neill, and H. Maei, "Optimal demand response using device-based reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2312–2324, 2015.

[69] H. Xu, A. D. Domínguez-García, and P. W. Sauer, "Optimal tap setting of voltage regulation transformers using batch reinforcement learning," *IEEE Transactions on Power Systems*, vol. 35, no. 3, pp. 1990–2001, 2020.

[70] "Reinforcement learning approach for congestion management and cascading failure prevention with experimental application," *Electric Power Systems Research*, vol. 141, pp. 179–190, 2016.

[71] Q. Yang, G. Wang, A. Sadeghi, G. Giannakis, and J. Sun, "Real-time voltage control using deep reinforcement learning," 04 2019.

[72] P. Gupta, A. Pal, and V. Vittal, “Coordinated wide-area damping control using deep neural networks and reinforcement learning,” *IEEE Transactions on Power Systems*, vol. 37, no. 1, pp. 365–376, 2022.

[73] R. Yousefian, R. Bhattarai, and S. Kamalasadan, “Transient stability enhancement of power grid with integrated wide area control of wind farms and synchronous generators,” *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4818–4831, 2017.

[74] R. Yousefian and S. Kamalasadan, “Energy function inspired value priority based global wide-area control of power grid,” *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 552–563, 2018.

[75] Q. Huang, R. Huang, W. Hao, J. Tan, R. Fan, and Z. Huang, “Adaptive power system emergency control using deep reinforcement learning,” *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1171–1182, 2020.

[76] C. C. Aggarwal, *Neural Networks and Deep Learning: A Textbook.* Springer; 2nd ed. 2023 edition.

[77] S. SHARMA. (2017) Activation functions in neural networks. [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[78] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. The MIT Press.

[79] T. Lillicrap and J. Hunt, “Continuous control with deep reinforcement learning,” *arXiv preprint*, 2015.

[80] Q. Yang, G. Wang, A. Sadeghi, G. Giannakis, and J. Sun, “Real-time voltage control using deep reinforcement learning,” 04 2019.

[81] M. Kamruzzaman, J. Duan, D. Shi, and M. Benidris, “A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources,” *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5525–5536, 2021.

[82] H. Yuan, R. S. Biswas, J. Tan, and Y. Zhang, “Developing a reduced 240-bus wecc dynamic model for frequency response study of high renewable integration,” in *2020 IEEE/PES Transmission and Distribution Conference and Exposition (TD)*, 2020, pp. 1–5.

[83] W. Shao and Z. Xu, “Excitation system parameter setting for power system planning,” in *IEEE Power Engineering Society Summer Meeting,*, vol. 1, 2002, pp. 541–546 vol.1.

[84] L. Jin, R. Kumar, and N. Elia, “Model predictive control-based real-time power system protection schemes,” *IEEE Transactions on Power Systems*, vol. 25, no. 2, pp. 988–998, 2010.

[85] J. Conto, *IEEE9 jconto.* [Online]. Available: https://drive.google.com/drive/folders/0B7uS9L2Woq_7fmd4YXVxMEZKT3dJV2FleGkzS2FzVmd1RHhBNVdUTGpvdldkMnl2bXRLM1kresourcekey=0-nuCqXu2XJ0_fxBzwHcmCGg

[86] A. B. Birchfield, T. Xu, K. M. Gegner, K. S. Shetye, and T. J. Overbye, "Grid structural characteristics as validation criteria for synthetic networks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3258–3265, 2017.

[87] A. B. Birchfield, K. M. Gegner, T. Xu, K. S. Shetye, and T. J. Overbye, "Statistical considerations in the creation of realistic synthetic power grids for geomagnetic disturbance studies," *IEEE Transactions on Power Systems*, vol. 32, no. 2, pp. 1502–1510, 2017.

[88] K. M. Gegner, A. B. Birchfield, T. Xu, K. S. Shetye, and T. J. Overbye, "A methodology for the creation of geographically realistic synthetic power flow models," in *2016 IEEE Power and Energy Conference at Illinois (PECI)*, 2016, pp. 1–6.