

University of Denver

Digital Commons @ DU

Electronic Theses and Dissertations

Graduate Studies

2023

Power System Dynamic Control and Performance Improvement Based on Reinforcement Learning

Wei Gao

University of Denver

Follow this and additional works at: <https://digitalcommons.du.edu/etd>



Part of the [Controls and Control Theory Commons](#), and the [Power and Energy Commons](#)

Recommended Citation

Gao, Wei, "Power System Dynamic Control and Performance Improvement Based on Reinforcement Learning" (2023). *Electronic Theses and Dissertations*. 2253.

<https://digitalcommons.du.edu/etd/2253>

This Dissertation is brought to you for free and open access by the Graduate Studies at Digital Commons @ DU. It has been accepted for inclusion in Electronic Theses and Dissertations by an authorized administrator of Digital Commons @ DU. For more information, please contact jennifer.cox@du.edu, dig-commons@du.edu.

Power System Dynamic Control and Performance Improvement Based on Reinforcement Learning

Abstract

This dissertation investigates the feasibility and effectiveness of using Reinforcement Learning (RL) techniques for power system dynamic control, particularly voltage and frequency control. The conventional control strategies used in power systems are complex and time-consuming due to the complicated high-order nonlinearities of the system. RL, which is a type of neural network-based technique, has shown promise in solving these complex problems by fitting any nonlinear system with the proper network structure.

The proposed RL algorithm, called Guided Surrogate Gradient-based Evolution Strategy (GSES) determines the weights of the policy (which generates the action for our control reference signal) without back-propagation process for gradient update using a simultaneous perturbation stochastic approximation approach comparing to many other RL algorithms, thus it achieves a much faster and more robust learning convergence. It is introduced and implemented in three different power system scenarios: High Voltage Direct Current (HVDC) based inter-area oscillation damping system, Doubly-fed Induction Generator (DFIG) based Fault-Ride-Through (FRT) system, and modified IEEE-39 Bus based frequency regulation system. In the case of the HVDC-based system, the proposed GSES-based power oscillation damping control approach overcomes the challenges of setting optimal controller parameters of the HVDC under various system transient events. This approach is also shown to be superior to conventional power oscillation damping methods. Further, the GSES algorithm is found to be effective in controlling the DFIG power and capacitor DC-link voltage, which helps prevent the rotor of DFIG from over-current risk and maintain the grid-connected operation. Finally, the proposed RL-based solution for frequency response in wind farms is tested on a modified IEEE-39 bus system and is found to reliably support the frequency of the power system and prevent unnecessary load shedding.

Overall, this dissertation shows the potential of RL-based techniques in power system dynamic control, particularly frequency control, and provides evidence for the effectiveness of the GSES algorithm in various power system scenarios. The use of RL in power systems could lead to more efficient and effective control strategies during contingencies, which is crucial in maintaining the stability of today's large, high-order nonlinear dynamic power systems.

Document Type

Dissertation

Degree Name

Ph.D.

Department

Electrical Engineering

First Advisor

Rui Fan

Second Advisor

David Wenzhong Gao

Third Advisor

Mohammad Matin

Keywords

Dynamic control, Fault ride through, Frequency regulation, Power system, Reinforcement learning, Renewable energy

Subject Categories

Controls and Control Theory | Electrical and Computer Engineering | Engineering | Power and Energy

Publication Statement

Copyright is held by the author. User is responsible for all copyright compliance.

POWER SYSTEM DYNAMIC CONTROL AND PERFORMANCE IMPROVEMENT
BASED ON REINFORCEMENT LEARNING

A Dissertation

Presented to

the Faculty of the Daniel Felix Ritchie School of Engineering and Computer Science

University of Denver

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

by

Wei Gao

June 2023

Advisor: Dr. Rui Fan

Author: Wei Gao

Title: POWER SYSTEM DYNAMIC CONTROL AND PERFORMANCE
IMPROVEMENT BASED ON REINFORCEMENT LEARNING

Advisor: Dr. Rui Fan

Degree Date: June 2023

ABSTRACT

This dissertation investigates the feasibility and effectiveness of using Reinforcement Learning (RL) techniques for power system dynamic control, particularly voltage and frequency control. The conventional control strategies used in power systems are complex and time-consuming due to the complicated high-order nonlinearities of the system. RL, which is a type of neural network-based technique, has shown promise in solving these complex problems by fitting any nonlinear system with the proper network structure.

The proposed RL algorithm, called Guided Surrogate Gradient-based Evolution Strategy (GSES) determines the weights of the policy (which generates the action for our control reference signal) without back-propagation process for gradient update using a simultaneous perturbation stochastic approximation approach comparing to many other RL algorithms, thus it achieves a much faster and more robust learning convergence. It is introduced and implemented in three different power system scenarios: High Voltage Direct Current (HVDC) based inter-area oscillation damping system, Doubly-fed Induction Generator (DFIG) based Fault-Ride-Through (FRT) system, and modified IEEE-39 Bus based frequency regulation system. In the case of the HVDC-based system, the proposed GSES-based power oscillation damping control approach overcomes the challenges of setting optimal controller parameters of the HVDC under various system transient events. This approach is also shown to be superior to conventional power oscillation damping methods. Further, the GSES algorithm is found to be effective in controlling the DFIG power and capacitor DC-link voltage, which helps prevent the rotor of DFIG from over-current

risk and maintain the grid-connected operation. Finally, the proposed RL-based solution for frequency response in wind farms is tested on a modified IEEE-39 bus system and is found to reliably support the frequency of the power system and prevent unnecessary load shedding.

Overall, this dissertation shows the potential of RL-based techniques in power system dynamic control, particularly frequency control, and provides evidence for the effectiveness of the GSES algorithm in various power system scenarios. The use of RL in power systems could lead to more efficient and effective control strategies during contingencies, which is crucial in maintaining the stability of today's large, high-order nonlinear dynamic power systems.

ACKNOWLEDGMENTS

I would like to take this opportunity to express my sincere gratitude towards the individuals who have helped me in the successful completion of my PhD's study.

First and foremost, I would like to extend my heartfelt appreciation to my advisor, Dr. Rui Fan, for his unwavering support, guidance, and encouragement throughout the research process. His valuable advice, suggestions, and feedback have been instrumental in shaping the direction of my research. I am also grateful to Dr. David Wenzhong Gao for his invaluable contributions, research collaborations, conference opportunities, and internship placements.

I would like to thank Dr. Yunbo Yi, the committee chair, and Dr. Mohammad Matin, the committee member, for their time, insightful feedback, and assistance during my dissertation and oral defense.

Furthermore, I am grateful to Dr. Margareta Stefanovic for those great control courses, and my lab mates Weihang Yan, Qiao Li, Abdullah Alharbi, Shruti Singh, Khawla Benjuma and Hongxia Wang for their support, constructive criticism, and collaborative efforts.

Last but not least, I would like to express my deepest appreciation to my family for their unwavering support, encouragement, and love throughout my academic journey. Their constant encouragement and motivation have been a source of inspiration and strength for me.

Once again, I would like to express my sincere gratitude to everyone who has helped me along the way. Without your support, this achievement would not have been possible.

TABLE OF CONTENTS

Acknowledgments	iv
List of Figures	viii
List of Acronyms	x
Chapter 1: Introduction	1
1.1 Literature Review	2
1.1.1 Motivation and Related Works	2
1.1.2 Role of Deep Learning	3
1.1.3 Comparison with Conventional Control Strategies	5
1.1.4 Guided Surrogate Gradient-based Evolution Strategy versus Existing Reinforcement Learning Methods	7
1.1.5 Conclusion	8
1.2 Power System Dynamic Control	9
1.2.1 HVDC Damping Control System	10
1.2.2 DFIG FRT System	13
1.2.3 DFIG based IEEE-39 Bus Frequency Regulation System	16
1.3 Expected Contribution	19
1.4 Outline of the Dissertation	20
Chapter 2: Fundamentals of NN, RL and GSES	22
2.1 Neural Network (NN)	22
2.1.1 Multilayer Perceptron (MLP)	22
2.1.2 Convolutional Neural Network (CNN)	25
2.2 Reinforcement Learning (RL)	27
2.3 RL with Guided Surrogate Gradient-based Evolution Strategy (GSES)	31
Chapter 3: GSES-based HVDC Damping Control System	35
3.1 Mathematical Modeling of HVDC	36
3.2 Controller	38
3.2.1 HVDC-based Oscillation Damping Controller	38
3.2.2 Limitations and Challenges of HVDC-based Oscillation Damping Controller	40

3.3	Design of GSES-based HVDC Damping Control Strategy	41
3.3.1	State Space	43
3.3.2	Action Space	44
3.3.3	Reward Function	44
3.4	GSES-based Damping Control Algorithm	45
3.5	Implementation Platform and Parallel Computation	47
3.6	Case Study and Discussion	49
3.6.1	Revised minniWECC System	49
3.6.2	Oscillation Scenarios and Damping Results	51
3.6.3	Case Studies and Discussion	52
3.6.4	Prony Analysis and Frequency Spectrum Analysis	53
Chapter 4:	GSES-based DFIG-FRT System	61
4.1	Mathematical Modeling of DFIG	62
4.1.1	Aerodynamics of wind turbine	62
4.1.2	Shaft system	62
4.1.3	Induction generator	63
4.2	Controller	64
4.2.1	Rotor Side Controller	64
4.2.2	Grid Side Controller	65
4.3	Design of GSES-DFIG-based Control Strategy	66
4.3.1	State Space	67
4.3.2	Action Space	68
4.3.3	Reward Function	68
4.4	GSES-based Control Algorithm	69
4.5	Implementation of Platform and Parallel Computation	70
4.6	Case Study and Discussion	72
4.6.1	Three-phase fault with 50% voltage drop	74
4.6.2	Three-phase fault with 100% voltage drop	77
4.6.3	Metallic single-line-to-ground fault	78
Chapter 5:	GSES-based IEEE-39 Bus Frequency Regulation System	80
5.1	Mathematical Modeling of DFIG	80
5.2	Controller	81
5.3	Pitch Angle based Deloading Controller	82
5.4	Design of GSES-based Control Strategy	82
5.4.1	State Space	82
5.4.2	Action Space	83
5.4.3	Reward Function	83
5.5	Implementation Platform and Parallel Computation	84
5.6	Case Study and Discussion	87

5.6.1	Loss of Generator	89
5.6.2	Sudden increase of Load	90
5.6.3	Decrease of Load	92
Chapter 6:	Conclusions and Future Work	94
6.1	Conclusions	94
6.2	Future Work	95
Publications	97
Appendices	99
Chapter A:	Comparison between BRS and GSES	99
A.1	Basic Random Search	99
A.2	Guided Surrogate Gradient-based Evolution Strategy	100
References	102

LIST OF FIGURES

1.1	General Control Strategy Platform	19
2.1	Structure of Example Multilayer Perceptron (MLP) [79]	25
2.2	Structure of Convolutional Neural Network (CNN) [79]	27
2.3	Diagram of RL: a_t is the action at time step t ; r_t is the reward at time step t ; s_{t+1} is the state at time step $t + 1$ [85]	28
3.1	Overview of an HVDC transmission system.	36
3.2	Reliable HVDC-based oscillation damping approach proposed by Schoenwald [102, 92]	39
3.3	Proposed GSES-Based HVDC Oscillation Damping Control Approach [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]	43
3.4	Flowchart of proposed GSES-based HVDC damping control approach [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]	48
3.5	An example of the reward of training the GSES-based controller [92]	49
3.6	Revised minniWECC system with PDCI HVDC transmission [107, 92]	56
3.7	Small-signal stability studies of the minniWECC system [92]	57
3.8	Inter-area oscillation damping for (i) Scenario I, (ii) Scenario II [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]	58
3.9	Inter-area oscillation damping for (i) Scenario III, (ii) Scenario IV [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]	59
3.10	Frequency spectrum analysis of the oscillations [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]	60
4.1	Typical DFIG grid connection structure [109, 85]	61
4.2	Typical DFIG RSC control structure [109, 85]	65
4.3	Typical DFIG GSC control structure [109, 85]	66
4.4	Overall GSES-based DFIG control framework: <i>reward_ep</i> is the episode reward; <i>ob_std</i> is the state covariance matrix; <i>ob_mean</i> is the state mean vector; <i>ob_array</i> is the trajectory of the states; <i>weights</i> is the weights of DNN policy [85].	71
4.5	An example of rewards during the training process [85]	72

4.6	Grid-connected DFIG system topology with fault [85]	73
4.7	Result of 3ϕ fault with 50% voltage drop	74
4.8	Result of 3ϕ fault with 100% voltage drop	75
4.9	Result of single-line-to-ground fault	75
5.1	DFIG grid connection structure for frequency response [85]	81
5.2	History of GSES training rewards	85
5.3	Overall GSES-based DFIG control framework: <i>reward_ep</i> is the episode reward; <i>ob_std</i> is the state covariance matrix; <i>ob_mean</i> is the state mean vector; <i>ob_array</i> is the trajectory of the states; <i>weights</i> is the weights of DNN policy. [85]	86
5.4	Modified IEEE 39-Bus test system for case studies	88
5.5	Result of frequency response for a loss of generator	90
5.6	Result of frequency response for a sudden increase of load	91
5.7	Result of frequency response for a sudden decrease of load	92

LIST Of ACRONYMS

AI	Artificial Intelligence.
API	Application Programming Interface.
BRS	Basic Random Search.
CNN	Convolutional Neural Network.
DFIG	Doubly-fed Induction Generator.
DL	Deep Learning.
DDPG	Deep Deterministic Policy Gradient.
DQN	Deep Q-Network.
DRL	Deep Reinforcement Learning.
ESS	Energy Storage System.
ES	Evolution Strategy.
FRT	Fault-Ride-Through.
FACTS	Flexible AC Transmission System.
GSES	Guided Surrogate Gradient-based Evolution Strategy.
GAE	Generalized Advantage Estimation.
GD	Gradient Descent.
GSC	Grid-side Converter.
IG	Induction Generator.
MLP	Multilayer Perceptron.
MPPT	Maximum Power Point Tracking.
MPP	Maximum Power Point.
MSE	Mean Square Error.
MDP	Markov Decision Process.
NN	Neural Network.

PMU	Phasor Measurement Unit.
PPO	Proximal Policy Optimization.
PSS	Power System Stabilizer.
PCC	Point of Common Coupling.
PDCI	Pacific DC Intertie.
RL	Reinforcement Learning.
RES	Renewable Energy Source.
ReLU	Rectified Linear Unit.
RSC	Rotor-side Converter.
SHL	Shallow Learning.
SAC	Soft Actor-Critic.
SVC	Static Var Compensator.
SGD	Standard Gradient Descent.
SLG	Single Line to Ground.
TRPO	Trust Region Policy Optimization.
TCSC	Thyristor Controlled Series Compensator.
TSR	Tip-speed Ratio.
VFC	Variable Frequency Converter.
WAMS	Wide Area Measurement Systems.
WADC	Wide Area Damping Control.
WPCS	Wind Power Conversion System.

CHAPTER 1

INTRODUCTION

Electricity has become a fundamental necessity in modern society, and power system control and stability are essential to ensure its reliable delivery. One of the primary objectives of power system control is to maintain a balance between generation and load to ensure voltage and frequency stability. However, with the increasing complexity of modern power systems, controlling them has become more challenging, especially during contingencies.

Today's power systems are large high-order nonlinear dynamic systems that pose significant challenges to conventional control strategies. Most traditional control strategies rely on system modeling, which is time-consuming and challenging in modern large-scale power systems. However, the advent of neural networks has revolutionized the field of control engineering by offering a powerful tool for fitting any nonlinear system. As such, they are a promising technology for power system control.

This dissertation aims to explore the feasibility and effectiveness of neural network-based technology in solving complex problems in today's power systems. The study will simulate various systems and scenarios, including Doubly-fed Induction Generator (DFIG)-based systems, High Voltage Direct Current (HVDC)-based systems, and DFIG-based IEEE-39 Bus Frequency Regulation (FR) systems. The research will evaluate the performance of neural network-based control strategies in comparison to conventional control

strategies, with a particular focus on the ability to maintain voltage and frequency stability during contingencies.

The research conducted in this dissertation will provide valuable insights into the potential of neural network-based control strategies in modern power systems. By demonstrating the feasibility and effectiveness of these strategies, the study could pave the way for the development of more efficient and effective control technologies for power systems, contributing to the sustainable and reliable delivery of electricity for generations to come.

1.1 Literature Review

1.1.1 Motivation and Related Works [1]

In the 21st century, the power system becomes more and more complex. The power system's complexity is a result of the integration of renewable energy resources, flexible active loads, and coupling with other energy forms. First, the integration of renewable energy resources, such as wind and solar power, presents challenges to frequency control due to the variability of these energy sources. Then, flexible active loads, such as electric vehicles and smart homes, introduce additional complexity due to their variability and unpredictability. Last, coupling with other energy forms, such as gas and heat, introduces new challenges due to the need for coordinated control. Especially, the complexity of the power system poses challenges for frequency analysis and control due to high-order nonlinearities introduced by the highly complex mechanism of frequency dynamics and stability and the diversity of frequency control means. In this dissertation, we will explore the role of Deep Learning (DL) in addressing the challenges associated with frequency analysis and control in complex power systems.

1.1.2 Role of Deep Learning

Deep learning is a popular computational technology nowadays because it has shown great potential in addressing the challenges associated with frequency analysis and control in complex power systems due to the following aspects:

- The power system is evolving into a cyber-physical system that combines smart grid technology with measurement, communication, and external systems [2]. New measuring devices, such as Phasor Measurement Units (PMUs) and Wide Area Measurement Systems (WAMS), are becoming more popular for power system security, stability assessment, disturbance monitoring, and position [3, 4]. The authors discuss and vision the future distributed energy systems and control technologies in [5, 6]. In [7], an energy consumption planing approach based on game theory is proposed for future smart grid at distribution level. [8, 9] demonstrate the potential of reinforcement learning for building energy control. Future pricing mechanism for smart distributional system is also discussed in [10, 11]. The informatization of the power system provides opportunities for analyzing and controlling the power system and requires a rethinking of frequency research methodology. Massive sample data-based data mining and analysis are essential for leveraging the advantages of power system informatization. In particular, DL, a powerful big data analysis method, is useful for data classification and feature extraction in complex mechanism environments, allowing fast online calculation and adaptive adjustment of frequency regulation and control strategy in power system informatization.
- The integration of Renewable Energy Sources (RES), flexible active loads, and large-scale regional interconnection is increasing system dimensions and decreasing system stability margins, making the power system more complex [12] as mentioned in subsection 1.1.1. This complexity affects the dynamic behavior of power system

frequency due to the uncertainty of RES, electronic devices, and multi-time scale coupling with other energy forms. The low inertia and randomness of the output of RES units reduce the ability to resist disturbance and regulation, leading to frequency stability issues. Additionally, the multi-scale, wide-band electromagnetic dynamics between converters and conventional grid elements increase the difficulty of frequency stability analysis and control [13]. However, DL can explore potentially related factors of frequency problems and reveal the physical characteristics of frequency problems from a data perspective, which is beneficial for studying the dynamic behavior of frequency, predicting its trend, and assessing its stability. DL also offers data-driven fast regulation capability and new opportunities for frequency issues [14]. Thus, DL is a powerful tool for frequency stability analysis and control in complex power systems, driving changes in frequency analysis theory and methods [15].

- Traditional offline prediction-based frequency stability control strategies face challenges due to the high uncertainty and unpredictability of RES and loads [16]. The expansion of system scale and enhancement of uncertainty in operation modes increase the number of stability control strategies required and enlarge control target dimensions, making control difficult due to unclear mechanisms [17, 18]. Simple pattern-based approaches are insufficient for stability control under the impact of complex and large capacity active power. Therefore, online computational aspects of frequency control are crucial, and data-driven modeling and control techniques are important for future research [18]. Intelligent control based on Artificial Intelligence (AI) can effectively use vast amounts of data from PMU and WAMS, take advantage of DL in nonlinear function approximation and learning, and fit data-driven mapping relationships between stability control strategy and frequency stability constraints.

Adaptive analysis on system operation mode and complex fault scenarios can enable online calculation and rapid adjustment of control strategy, enabling precise frequency control. Thus, data-driven modeling and control techniques, as well as intelligent control based on AI, are necessary for effective frequency stability control in modern power grids. These techniques can enable online computation and adapt to complex scenarios, ensuring the optimal operation of the power system.

In a word, DL algorithms can analyze large amounts of data to identify patterns and relationships, enabling accurate and timely decision-making. In addition, it can also learn from past experiences to improve its performance and adapt to changing conditions. These features succeed to bring the DL technology into power system to solve the difficulty of high-order nonlinearities by the increased complexity as we mentioned in subsection 1.1.1. Therefore, there are plenty of different DL-based applications in power systems, but a few about the frequency analysis and control. Thus, we would like to fill the gap in this area. Typically, DL can be used for various tasks related to frequency analysis and control, such as predicting frequency deviations, identifying the causes of frequency deviations, and designing control strategies to maintain frequency stability. DL algorithms can also be used to optimize the operation of the power system by predicting the power output of renewable energy resources and adjusting the operation of flexible active loads. In this dissertation, we will use RL, one type of machine learning which utilizes DL's advantages-based algorithm to design a general control strategy for power system dynamic control.

1.1.3 Comparison with Conventional Control Strategies

Power system dynamic control techniques such as time-domain simulation, equivalent models, and intelligent methods have been widely used in power systems [19]. However, these traditional methods have limitations when dealing with the frequency and voltage

problems of modern power systems, which are high-dimensional, time-varying, nonlinear, and involve multi-source information interaction as summarized in follows [1]:

- Time-domain simulation is advantageous because it provides high accuracy and comprehensive information. However, it has some disadvantages, such as a large amount of calculation, time-consuming, and modeling difficulty. Time-domain simulation is applicable to disturbances of unlimited size and scale.
- Mathematical equivalent model is a fast approach that offers visuality, rigorous derivation, and clear logic. However, it has low calculation accuracy and ignores some factors, and obtaining parameters is difficult. Mathematical equivalent model is applicable to small disturbances and scale.
- Artificial intelligence based on Shallow Learning (SHL) has a fast training speed and strong mathematical theory support. However, its ability of feature extraction is limited, it depends on manual experience, and has poor generalization. AI based on SHL is applicable to large disturbances and unlimited scale.
- Artificial intelligence based on DL offers high precision, high speed, and good generalization. However, it has poor interpretability and depends on the data volume. AI based on DL is applicable to large disturbances and unlimited scale.

Thanks to the development of WAMS techniques and data processing capabilities, information collection and processing are no longer obstacles to power system online operation and control [20]. In particular, intelligent methods that rely on frequency data samples for learning and online analysis are suitable for frequency analysis under large disturbances. Among these methods, DL stands out as a promising approach to address complex frequency issues in modern power systems. DL can adapt to the nonlinear and

complex features of the power system and identify the potential causes of frequency problems. Moreover, DL's ability to learn from massive data and identify subtle patterns can help capture complex relationships between input and output variables, making it a suitable method for frequency analysis under large disturbances. Therefore, the application of intelligent methods, especially DL, in frequency analysis can enhance the effectiveness and efficiency of power system frequency analysis and control, which is crucial for maintaining the stability and security of power systems.

However, DL requires large amount of labeled data in advance, which is one of the major disadvantages of DL. Reinforcement Learning (RL), which does not require those data prepared in advance, instead, it collects the data through exploring the environment by itself during the training process. In addition, it's policy function can be based on deep neural network for taking the advantage of DL's characteristics. Thus, it is popular for optimal decision-making problems. It is also a good fit of our power system dynamic control strategy problem as we can make RL make decision or control signal of our dynamical power system. We will explore more about RL in section 2.2.

1.1.4 Guided Surrogate Gradient-based Evolution Strategy versus Existing Reinforcement Learning Methods

Continuous control with RL methods is a common practice in simulation environments using standard benchmark suites. Among the most popular RL algorithms, asynchronous parallelization of actor-critic methods is used for fast policy training in Atari video games and MuJoCo (A physics engine for model-based control [21]) models [22]. The Generalized Advantage Estimation (GAE) method, which reduces variance with less bias than previous techniques, is widely used for advantage estimation [23]. The Trust Region Policy Optimization (TRPO) algorithm maximizes an approximate average reward objective and is regularized by a KL-divergence penalty [24]. The Proximal Policy Optimization

(PPO) algorithm is a successor to TRPO and has better sample complexity and easier implementation [25]. The Deep Deterministic Policy Gradient (DDPG) method integrates Deep Q-Network (DQN) [26] and actor-critic framework for deterministic policy learning [27]. On the other hand, the Soft Actor-Critic (SAC) algorithm is an off-policy actor-critic method that maximizes both the expected reward and the entropy of a stochastic policy. Linear policies can be used to simplify the search space, and natural gradients can be utilized to train linear policies for the MuJoCo locomotion tasks.

However, the major disadvantage of most existing RL algorithms is less robustness due to too many hyper-parameters to be tuned. The Guided Surrogate Gradient-based Evolution Strategy (GSES) is a type of RL algorithm that takes a different approach to gradient descent computation than other algorithms. Unlike other RL algorithms, GSES does not require a backpropagation process to determine the gradient descent. Instead, it uses an evolutionary strategy that involves perturbing multiple stochastic approximations and selecting the action direction with the highest estimated reward. This approach makes GSES more robust and easier to train compared to other RL algorithms that rely heavily on hyper-parameters. Moreover, a surrogate gradient method is used during the training process to further reduce the variance and guide the evolutionary direction more effectively. To speed up the computational process, parallel computation techniques are implemented, which allow GSES to engage with multiple workers simultaneously during the training process. This approach results in faster and more efficient training. The use of parallel computation techniques, along with the surrogate gradient method, has made GSES a highly effective and efficient RL algorithm.

1.1.5 Conclusion

In conclusion, the integration of renewable energy resources, flexible active loads, and coupling with other energy forms has made the power system more complex, posing chal-

lenges for frequency analysis and control. DL has shown great potential in addressing these challenges by leveraging the advantages of power system informatization, exploring potentially related factors of frequency problems, and offering data-driven fast regulation capabilities. Therefore, DL (more concretely, RL) algorithms can be used for various tasks related to frequency analysis and control, such as predicting frequency deviations, identifying the causes of frequency deviations, and designing control strategies to maintain frequency stability. Compared to traditional offline prediction-based frequency stability control strategies, data-driven modeling and control techniques, as well as intelligent control based on AI, are necessary for effective frequency stability control in modern power grids. These techniques can enable online computation and adapt to complex scenarios, ensuring the optimal operation of the power system.

1.2 Power System Dynamic Control

Power system dynamic control are essential for the stable operation of the power system. Especially, in the field of frequency analysis and control, the power system's in-depth informatization of monitoring, highly complex mechanism of frequency dynamics and stability, and diversity of frequency control means are presenting new features that require DL. The use of DL in frequency analysis and control can improve the accuracy and speed of frequency analysis and control and enable real-time decision-making. There are plenty of DL-based application in power system, but not many for the power system dynamic frequency analysis and control. Therefore, we would like to fill the gap in the area using several typical power systems including HVDC damping control system, DFIG FRT system and DFIG based IEEE-39 bus frequency regulation system.

1.2.1 HVDC Damping Control System

Inter-area oscillations in large-scale power systems are a type of low-frequency oscillation that can occur due to the coupling of power systems in different areas. These oscillations typically have a frequency of 0.1 Hz to 1 Hz and can cause power fluctuations and instability. They are often caused by power system disturbances, such as generator or transmission line outages, and can be exacerbated by the high penetration of RES. The high variability of RES can cause sudden changes in power generation, leading to inter-area oscillations. These oscillations can lead to significant power fluctuations and instability, causing economic and social losses [28].

Wide Area Damping Control

To address these issues, Wide Area Damping Control (WADC) methods have been widely used to damp inter-area oscillations. These methods involve using Power System Stabilizer (PSS) and Flexible AC Transmission System (FACTS) devices as actuators to control the power flow and reduce the oscillations [29, 30] based on the corresponding oscillation information collected by PMU [31]. PSS devices are used to stabilize the generator output and reduce the oscillations by adjusting the excitation voltage. FACTS devices, such as Thyristor Controlled Series Compensator (TCSC) [32, 33] and Static Var Compensator (SVC) [34], can adjust the transmission line impedance and voltage to control the power flow and reduce the oscillations.

However, the requirement of a linearized system model for controller design poses a problem. The linearized model assumes that the system is operating at a specific operating point and that small disturbances can be modeled as linear perturbations around this operating point. However, in large-scale power systems, the operating conditions can vary significantly, making it challenging to use a linearized model for controller design. More-

over, the nonlinear dynamics of the power system can lead to unexpected behavior, making it difficult to design a robust controller.

Therefore, this article discusses the challenges of inter-area oscillations in large-scale power systems and the use of WADC methods to address these issues. It also explores HVDC transmission as an excellent solution and proposes a novel approach that uses RL to dynamically determine the parameters of the HVDC controller based on system states to address the challenges.

HVDC Transmission based Damping Control

In recent years, HVDC transmission [35, 36] has been proposed as an excellent solution to address the challenges of inter-area oscillations in large-scale power systems [37, 38]. HVDC transmission can modulate the active power directly and has the capability to modulate area power flows to damp inter-area oscillations. Moreover, HVDC transmission can provide fast response times and high controllability, making it a suitable solution for controlling power system dynamics.

However, most existing HVDC-based oscillation damping approaches also require a linearized or reduced-order power system model, which poses a problem for large-scale power systems. The use of a linearized or reduced-order model can lead to inaccurate control and can be difficult to tune, leading to poor performance. In addition, controllers with fixed parameters would face difficulty in achieving optimal performance under various transient events such as faults, change of loads or loss of generations which could trigger different inter-area oscillations. This poses a challenge as the parameters of the existing HVDC-based oscillation damping controllers would be fixed once they come into service.

RL based HVDC Damping Control

To address the challenges of HVDC-based oscillation damping, a novel approach that uses RL to dynamically determine the parameters of the HVDC controller based on system states has been proposed. RL is a type of machine learning that involves training an agent to take actions in an environment to maximize a reward signal. In the context of power systems, the agent can be trained to determine the optimal parameters for the HVDC controller based on the current system states to damp inter-area oscillations.

The RL-based HVDC control approach involves training an RL agent to determine the optimal parameters for the HVDC controller based on the current system states. The RL agent receives feedback in the form of a reward signal, which reflects the performance of the controller. The agent uses this feedback to adjust its parameters and improve its performance over time.

One of the benefits of using RL-based HVDC control is that it does not require a linearized or reduced-order model of the power system. Instead, the RL agent can learn from the nonlinear dynamics of the power system and adapt its control strategy to changing operating conditions. This can lead to more accurate and robust control, even in large-scale power systems with significant variations in operating conditions.

The RL (specifically, GSES)-based HVDC control approach has been tested on a benchmark power system model with inter-area oscillations. The results showed that the RL-based approach outperformed existing HVDC-based oscillation damping approaches in terms of damping inter-area oscillations. Moreover, the RL-based approach was able to adapt to changing operating conditions and maintain stable performance over time.

1.2.2 DFIG FRT System

The integration of RES into the power grid has gained significant attention in recent years. Among the different types of RES, wind power is one of the most promising due to its high availability, low cost, and environment-friendly nature [39]. DFIGs are the most commonly used technology for Wind Power Conversion Systems (WPCS) due to their cost-effectiveness and ability to control power flow [40, 41, 42]. In general, DFIG can generate the maximum power using Maximum Power Point Tracking (MPPT) technique under normal conditions [43]. However, during grid faults, maintaining the connection of a DFIG to the power grid becomes challenging due to over-current situations and increased capacitor DC-link voltage [44, 45]. To prevent the unnecessary costs of stopping and restarting wind generators, it is required by grid codes that these generators have sufficient capability to remain connected to the grid during faults for a specific period of time [46]. This capability, known as fault ride-through capability, ensures that wind generators can continue to provide power to the grid during disturbances, such as short circuits or voltage dips. The duration of the fault ride-through capability varies according to the specific regulations outlined in the grid code, but typically lasts for a few seconds to a few minutes. Therefore, to overcome these challenges, different hardware and advanced control strategies have been proposed. This article discusses these strategies and proposes the use of RL algorithms to improve DFIG performance during grid faults.

DFIG Performance during Grid Faults

DFIGs are designed to operate under nominal conditions with a constant frequency and voltage. During grid faults, the voltage and frequency at the Point of Common Coupling (PCC) change, which affects the performance of DFIG. The performance of DFIG during grid faults is evaluated based on the following factors:

- **Over-Current Protection:** During grid faults, an over-current situation can occur, leading to equipment damage or tripping of the system. Over-current protection is required to prevent such situations.
- **Capacitor DC-Link Voltage:** The capacitor DC-link voltage increases during grid faults, which can cause over-voltage situations and damage the DFIG.
- **Power Quality:** The power quality of the system is affected during grid faults, leading to voltage dips and harmonics.

Hardware Control Strategies based DFIG FRT

Different hardware strategies have been proposed to improve the performance of DFIG during grid faults. Some of these strategies are:

- **Energy Storage System (ESS):** An ESS can be used to regulate the power flow during grid faults. During a fault, the ESS can supply power to the DFIG, maintaining its operation [47, 48].
- **Crowbar Protection:** A crowbar protection system can be used to protect the DFIG from over-voltage situations. When the DC-link voltage exceeds a certain threshold, the crowbar protection system operates, short-circuiting the stator windings and reducing the DC-link voltage [49, 50].
- **Rotor Resistance Control:** Rotor resistance control can be used to limit the over-current situation during grid faults. By increasing the rotor resistance, the current flow through the stator windings can be reduced, preventing over-current situations [51].

Advanced Control Strategies based DFIG FRT

Advanced control strategies can be used to improve the performance of DFIG during grid faults. Some of these strategies are:

- **Demagnetizing Current Control:** Demagnetizing current control can be used to limit the over-current situation during grid faults. By controlling the demagnetizing current, the current flow through the stator windings can be reduced, preventing over-current situations [52].
- **Virtual Resistance Control:** Virtual resistance control can be used to maintain the DC-link voltage during grid faults. By introducing a virtual resistance in the control loop, the DC-link voltage can be regulated [53, 54, 55, 56].

RL based DFIG FRT

RL is a type of machine learning that is used to make decisions in an uncertain environment and RL algorithms learn by interacting with the environment and receiving feedback in the form of rewards or penalties [57]. Therefore, RL algorithms can be used to dynamically regulate the reference signal of the controllers inside DFIG during grid faults to improve performance. Moreover, RL algorithms can also be used to learn an adaptive signal dispatch strategy that collects necessary system states and learns the best behaviors during grid faults.

The RL algorithm collects the system's state and learns the best actions to take during grid faults. The RL algorithm consists of an actor-critic network and a replay buffer. The actor network takes the system's state as input and outputs the control action. The critic network evaluates the output of the actor network and provides feedback in the form of a reward signal. The replay buffer stores the experiences of the RL algorithm, which are used to train the actor-critic network.

The proposed RL algorithm based adaptive signal dispatch strategy is evaluated using a simulation model of a wind turbine system. The simulation model consists of a wind turbine, a DFIG, and a power grid. The RL algorithm is used to control the DFIG during grid faults. The results show that the proposed RL algorithm outperforms the traditional control strategies in terms of over-current protection and DC-link voltage regulation.

The dissertation also proposes the use of the GSES algorithm [58] to improve DFIG performance during grid faults. The GSES algorithm is used due to its advantages over other RL algorithms including DQN, DDPG, and PPO [59]. The GSES algorithm is a model-free RL algorithm that is suitable for continuous control problems. The GSES algorithm uses surrogate-gradient-based optimization to update the actor network and evolution strategies to update the critic network.

The proposed GSES RL algorithm is evaluated using a simulation model of a wind turbine system. The simulation model consists of a wind turbine, a DFIG, and a power grid. The GSES algorithm is used to control the DFIG during grid faults. The results show that the proposed GSES RL algorithm outperforms the traditional control strategies and the RL algorithm based adaptive signal dispatch strategy in terms of over-current protection and DC-link voltage regulation.

1.2.3 DFIG based IEEE-39 Bus Frequency Regulation System

Frequency response is a critical function in power systems that ensures the frequency of the system stays within an acceptable range. Traditional power systems use synchronous generators to provide frequency response [60]. However, the increasing integration of RES such as wind turbines has led to the exploration of alternative solutions [61]. Wind turbines have the potential for cost savings and fast response times, making them an attractive option for frequency response. The use of DFIGs is preferred in the renewable energy generation industry due to their advantages, such as variable speed operation options and

low power-rating inverter requirements [62]. This dissertation discusses the use of wind turbines, specifically DFIGs, for frequency response, and proposes a RL-based approach to dynamically adjust the pitch angle of wind turbine blades.

DFIG Pitch Control based Frequency Regulation

Despite the challenges posed by the variability and uncertainty of wind power [63], the potential cost savings and fast response time of wind turbines compared to conventional generators have made them an attractive option for frequency response in recent years [64].

Pitch control is one of the most common approaches to utilizing wind turbines for frequency response [65]. This approach involves adjusting the pitch angle of the wind turbine blades to regulate the power output and maintain the frequency within an acceptable range. When the frequency of the power system decreases, the pitch angle of the blades is increased to reduce the power output, which helps to increase the frequency [66, 67, 68, 69, 70, 71, 72]. Conversely, when the frequency of the power system increases, the pitch angle of the blades is decreased to increase the power output and decrease the frequency. Utilizing wind turbines for frequency response involves more than just pitch control. In fact, other methods have been suggested, including combining energy storage systems with wind turbines and solar panels [73].

The use of pitch control for frequency response requires a reference point for the pitch angle. During normal operations, wind turbines usually operate in the Maximum Power Point (MPP) mode to output as much power as possible [43]. However, power system operators may require certain wind farms to operate in deloaded modes [74] during emergencies to fulfill auxiliary grid services such as load regulation, spin and non-spin reserve, and frequency support. In these cases, the reference point for the pitch angle needs to be adjusted to maintain the frequency within an acceptable range.

The area of research that involves using wind turbines for frequency response shows promise. Nonetheless, further investigations are required to determine the most effective control strategies and system configurations that can be utilized for different power system scenarios in using wind turbines for frequency response.

DFIG RL based Frequency Regulation

RL is a machine learning technique that enables an agent to learn optimal decision-making strategies by receiving feedback from the environment. It has been shown to be highly effective in solving complex nonlinear optimization problems in power systems [75, 76, 77, 78].

In this study, we propose a RL-based method that utilizes a new evolutionary RL-based adaptive signal modulation algorithm, GSES [58]. The goal of our approach is to dynamically adjust the reference point of the pitch angle of a wind turbine to regulate the system frequency while maintaining its power output. In our case, the wind turbine serves as the agent, and the power system represents the environment. By utilizing the GSES algorithm, we can find the optimal pitch angle that can effectively regulate the system frequency while maintaining the power output of the wind turbine.

To evaluate the performance of our proposed approach, we tested it on a modified IEEE-39 bus system, which is a widely accepted benchmark system in power system analysis. We replaced one synchronous generator in the test system with a DFIG operating in de-loaded mode to provide capacity reserve for frequency response. To compare our method's performance with traditional frequency response approaches, we conducted simulations. The outcomes demonstrated that our approach can efficiently support the power system's frequency and prevent unnecessary load shedding.

1.3 Expected Contribution

The aim of this dissertation is to create and implement a general RL-based control strategy for modern power systems, with a focus on HVDC Damping Control System, DFIG FRT System and DFIG-based IEEE-39 Bus Frequency Regulation System. By developing a platform using the general GSES based control strategy, we could easily extend to some other typical power systems as shown in Figure 1.1. The following contributions are highlighted in this dissertation:

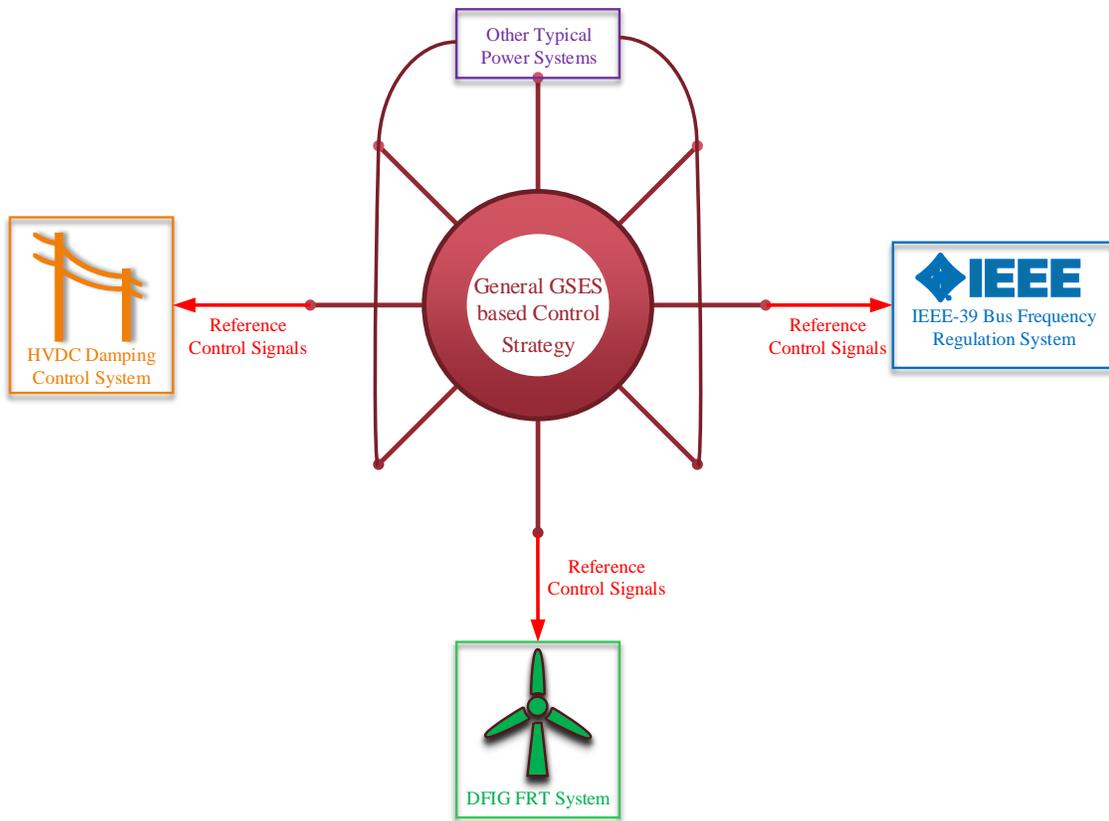


Figure 1.1: General Control Strategy Platform

- A novel RL-based control strategy for different subsystems of large power systems that outperforms conventional methods.

- The RL-based controller is capable of adapting to various system scenarios triggered by different transient events through dynamic parameter changes.
- A model-free controller that doesn't require a linearized or reduced-order system model.
- The GSES-based RL algorithm used in the controller requires fewer hyper-parameters, making it easier to extend the control approach to other power systems.
- The proposed approach doesn't require a back-propagation process and can engage with multiple agents simultaneously, and the use of parallel computation techniques improves the training speed significantly.

1.4 Outline of the Dissertation

The structure of this dissertation is as follows:

In Chapter 2, a comprehensive literature review is conducted, where fundamental knowledge of neural networks, RL, and a specialized RL algorithm, GSES are introduced. The neural network structure and mathematical foundations of RL and GSES are discussed in-depth.

Moving on to Chapter 3, the mathematical model of HVDC and its typical control scheme are discussed, followed by the step-by-step illustration of the design of an RL-based control strategy for improving performance during inter-area oscillation. The corresponding algorithm is built and tested, and various case studies demonstrate the effectiveness of the RL-based control strategy.

Similarly, in Chapter 4, the mathematical model of DFIG and its typical control scheme are discussed. Then, a step-by-step illustration of the design of an RL-based control strategy for improving performance during faults is provided, along with the corresponding

algorithm and testing platform. Finally, the effectiveness of the RL-based control strategy is demonstrated through various case studies.

In Chapter 5, the modified IEEE 39-bus system with one synchronous generator replaced by a DFIG subsystem and its typical control scheme are discussed. Then, a step-by-step illustration of the design of an RL-based control strategy for improving performance during contingencies is provided, along with the corresponding algorithm and testing platform. Finally, the effectiveness of the RL-based control strategy is demonstrated through various case studies.

Chapter 6 concludes the research work presented in this dissertation, and several recommendations for future studies are made.

CHAPTER 2

FUNDAMENTALS OF NN, RL AND GSES

2.1 Neural Network (NN) [79]

In recent years, Neural Networks (NN) have become increasingly popular in various research fields due to the rapid evolution of computing technology. They have shown great potential in solving complex or even impossible problems that cannot be tackled by conventional methods. Two common neural network structures, namely the Multilayer Perceptron (MLP) and Convolutional Neural Network (CNN), have received significant attention and success in various applications.

2.1.1 Multilayer Perceptron (MLP)

The multilayer perceptron is the basic deep neural network structure that consists of an input layer, output layer, and multiple hidden layers, as shown in Figure 2.1. Each hidden layer comprises several neurons, and the neurons in the same layer share the same activation function. MLP is powerful because it has the ability to learn complex nonlinear relationships between input and output data by adjusting the weights between the neurons. The basic operation of a neuron is to compute a weighted sum of the input signals followed by a nonlinear activation function. The mathematical representation of a neuron can be

expressed as:

$$f(x_1, \dots, x_n) = \sum_{i=1}^n w_i * x_i + b \quad (2.1)$$

where n is the total number of input dimensions, w_i is the weight of the i^{th} input neuron, x_i is the input signal, b is the bias, and f is the activation function.

Various activation functions have been proposed and used in MLP, such as Sigmoid, Tanh, and Rectified Linear Unit (ReLU), which are shown in (Equation 2.2), (Equation 2.3), (Equation 2.4), respectively.

$$Sigmoid(x) = \frac{1}{1 + e^{-x}} \quad (2.2)$$

$$Tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (2.3)$$

$$ReLU(x) = \max(0, x) \quad (2.4)$$

These activation functions introduce nonlinearity into the neural network, allowing it to model complex nonlinear relationships between input and output data.

In the example MLP shown in Figure 2.1, the input layer has 2 neurons, hidden layer 1 has 4 neurons, hidden layer 2 has 5 neurons, and the output layer has 2 neurons. Each neuron has a linear function and a nonlinear activation function, and all the neurons will generate output and transfer it to the next layer, which is also called forward propagation. The output of the neural network can be expressed as:

$$\hat{y}_i = Sigmoid\left(\sum_{i=1}^{n_{k-1}} w_i * \dots ReLU\left(\sum_{i=1}^{n_2} w_i * \left(ReLU\left(\sum_{i=1}^{n_1} w_i * x_i + b_1\right) + b_2\right) \dots + b_{k-1}\right)\right) \quad (2.5)$$

where k is the number of layers, n_{k-1} is the number of neurons in the previous layer, w_i is the weight of the i th connection, x_i is the input value of the i th neuron, b_j is the bias of the j th layer, and *ReLU* is the Rectified Linear Unit activation function for the hidden layers, while the sigmoid function is used for the output layer.

To train an MLP, we need to determine the weights that minimize the difference between the predicted output and the actual output. One commonly used loss function is the Mean Square Error (MSE), which measures the average of the squared differences between the predicted output \hat{y}_i and actual output y_i . The MSE is defined as:

$$L(y_i, \hat{y}_i) = \frac{1}{m} \sum_{j=1}^m (y_{ij} - \hat{y}_{ij})^2 \quad (2.6)$$

where m is the number of sample data. Based on the loss, various optimization methods can be applied to update the weights of the neural network. One of the most widely used optimization methods is the Gradient Descent (GD) algorithm, which aims to minimize the loss function by iteratively updating the weights in the opposite direction of the gradient of the loss function. The Standard Gradient Descent (SGD) algorithm can be expressed as:

$$\nabla L = \partial \frac{L}{\partial x_i} \nabla x_i \quad (2.7)$$

This process of updating the weights using the gradient of the loss function is known as backpropagation. By combining the forward propagation and backpropagation, the neural network iteratively learns the optimal weights to minimize the loss function and fit the training data.

Neural networks have been applied in various fields, such as image classification, natural language processing, speech recognition, and recommendation systems and achieve a great success.

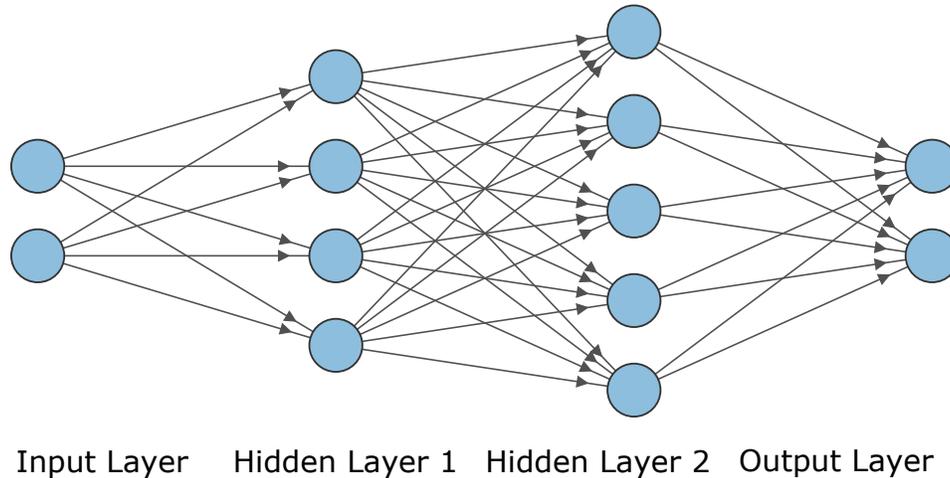


Figure 2.1: Structure of Example Multilayer Perceptron (MLP) [79]

In conclusion, the MLP is a fundamental type of neural network that can model non-linear relationships between input and output data. By applying forward propagation and back propagation, the MLP can learn from large amounts of training data and achieve high accuracy in various applications. With the development of optimization algorithms and the emergence of new architectures, neural networks have become a powerful tool in machine learning and have shown remarkable performance in many real-world applications.

2.1.2 Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) [80] is a type of neural network that is primarily used for image preprocessing, but it has found widespread use in many other fields, including audio signal processing, natural language processing, and medical image analysis. The architecture of a CNN consists of several layers, including convolutional, pooling, and fully connected layers as shown in Figure 2.2. The convolutional layers are the key component of a CNN, as they perform feature extraction from the input image by convolving a set of filters with the input data. The pooling layers are used for reducing the dimensional-

ity of the output of the convolutional layers, while the fully connected layers are used for classification or regression tasks.

The basic idea behind a convolutional layer is to convolve a set of learnable filters with the input image to produce a set of feature maps. Each filter is a small matrix that is slid over the input image, computing a dot product between the filter and the local patch of the input image. The output of this convolution operation is a new matrix called a feature map, which contains the responses of the filter to different parts of the input image. By learning a set of filters that can detect different features of the input image, the CNN can automatically extract meaningful features from the input data.

After each convolutional layer, a pooling layer is typically used to reduce the spatial dimensionality of the output feature maps, while retaining the most salient information. This is typically done using a max-pooling or average-pooling operation, which computes the maximum or average value of a small sub-region of the feature map. The pooling operation helps to reduce the number of parameters in the network, making it less prone to overfitting.

One of the recent applications of CNNs is in power system analysis [81], where 1-D CNNs have been shown to be effective in extracting features from 1-D waveforms, which are commonly used in power system analysis. By using a 1-D convolutional layer, the network can learn to extract important features from the time-domain waveform, such as harmonic content, transient behavior, and steady-state oscillations. This approach has been applied to a variety of power system analysis tasks, including fault detection, load forecasting, and power quality analysis [82].

In conclusion, CNNs are a powerful tool for feature extraction and classification in a wide range of applications. The architecture of a CNN is flexible and can be adapted to different types of input data, including images, audio signals, and time-series data. The

use of CNNs in power system analysis has shown promising results, and further research is needed to explore the full potential of this approach.

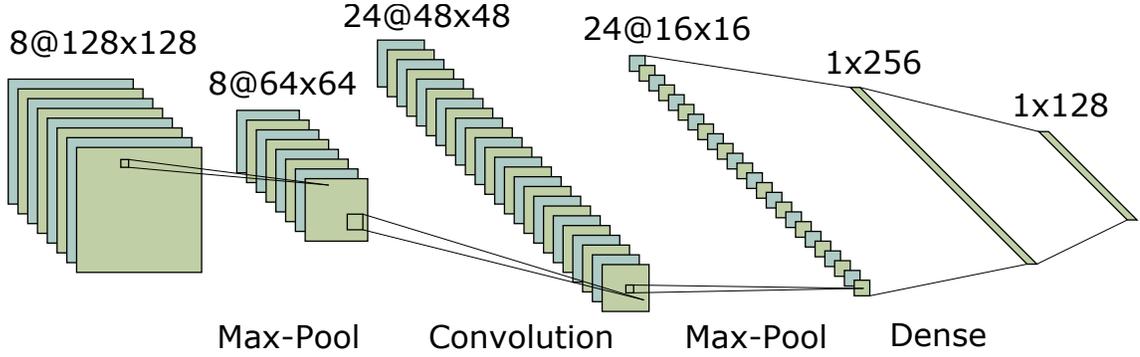


Figure 2.2: Structure of Convolutional Neural Network (CNN) [79]

2.2 Reinforcement Learning (RL)

Reinforcement Learning (RL) has been widely applied in various domains such as robotics, finance, and control systems due to its capability of learning optimal behavior from interactions with the environment. A typical RL problem can be described as a finite Markov Decision Process (MDP) $(\mathcal{S}, \mathcal{A}, P_{s,s'}^a, r_{s,s'}^a)$ [83], where \mathcal{S} is a set of states; \mathcal{A} is a set of actions; $P_{s,s'}^a$ represents the probability that action a results in transiting from state s to the other state s' ; $r_{s,s'}^a$ represents the received reward after transition from state s to state s' . The objective of RL is to learn a policy $\pi(a_t, |, s_t) : \mathcal{S} \rightarrow \mathcal{A}$, such that it maximizes the expected accumulated discounted reward $\mathcal{J}(\pi)$ with respect to the policy π over time under the state transition function p [57].

In practice, the policy function $\pi(a_t, |, s_t)$ is usually parameterized by a neural network with weights θ and the policy is denoted as π_θ . The neural network-based policy has shown great potential in solving complex RL problems with high-dimensional state and action spaces [26, 84]. As show in Figure 2.3, the policy function π_θ generates action signals a_t at current time step t , and obtains a reward r_t based on the user-designed

reward function r_{s_t, a_t} . At the next time step $t + 1$, the state s_t changes to s_{t+1} according to the transition function $p(s_{t+1}, |, s_t, a_t)$. Therefore, a finite MDP can be summarized as a trajectory $(s_t, a_t, r_t; , s_{t+1}, a_{t+1}, r_{t+1}; , \dots; , s_T, a_T, r_T)$, where T is the total time step of the whole process. State transition function $p(s_{t+1}, |, s_t, a_t)$ may be random depending on the environment, where capital S_t is a random variable; lower-case s_t is the specific state.

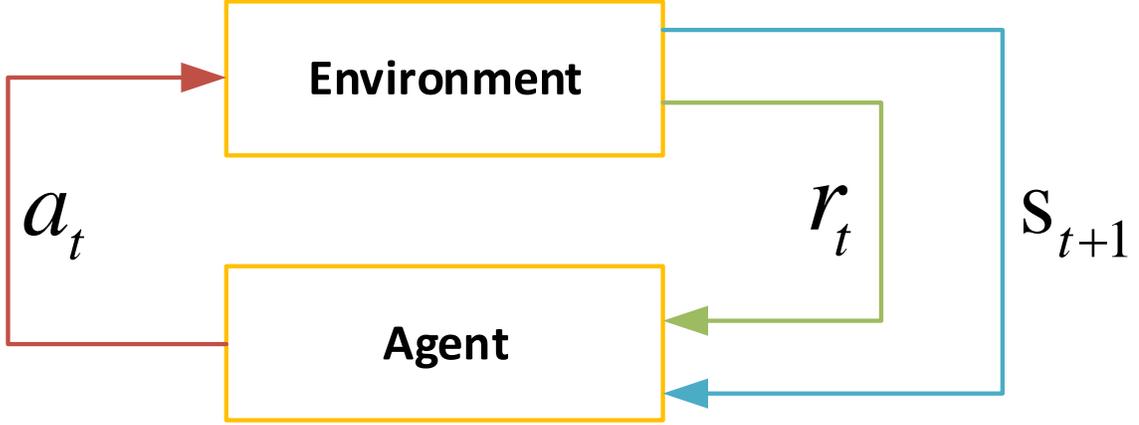


Figure 2.3: Diagram of RL: a_t is the action at time step t ; r_t is the reward at time step t ; s_{t+1} is the state at time step $t + 1$ [85]

The reward is the key criterion to measure whether an agent learns meaningful behavior from experienced interactions with the environment. More concretely, total reward of the whole trajectory in one episode is more significant than single time step reward as we focus more on the final result. As a result, the goal of RL is to maximize the expected accumulated discounted reward $\mathcal{J}(\pi)$ with respect to the policy π over time. The expected accumulated discounted reward $\mathcal{J}(\pi)$ can be calculated as follows:

$$\mathcal{J}(\pi) = \mathbb{E}_{s_0, a_0, s_1, a_1, \dots, s_T, a_T} \sum_{t=0}^T \gamma^t r(s_t, a_t) \quad (2.8)$$

where γ is a discount factor, $a_t \sim \pi(a_t | s_t)$ and $s_{t+1} \sim p(s_{t+1} | s_t, a_t)$, and T is the maximum end time. In power system optimization problems, the typical objective is to minimize the cost function. This objective is similar to the accumulated discounted reward maximiza-

tion in Reinforcement Learning (RL). In Deep Reinforcement Learning (DRL), the policy is denoted as π_θ and is usually parameterized by a neural network with weights θ .

Therefore, return defined as cumulative future reward is computed as:

$$U_t = R_t + R_{t+1} + R_{t+2} + R_{t+3} + \dots \quad (2.9)$$

However, future reward is not equally valued with current reward and thus modified discounted return defined as cumulative discounted future reward is computed as:

$$U_t = R_t + \gamma * R_{t+1} + \gamma^2 * R_{t+2} + \gamma^3 * R_{t+3} + \dots \quad (2.10)$$

where γ is called discount rate between $[0,1]$; U_t is also a random variable depends on the future random variables $A_t, A_{t+1}, A_{t+2}, \dots$ and $S_t, S_{t+1}, S_{t+2}, \dots$. To evaluate the random variable U_t , expectation can be computed to remove its randomness as follows:

$$Q_\pi(s_t, a_t) = \mathbb{E}[U_t | S_t = s_t, A_t = a_t] \quad (2.11)$$

where $Q_\pi(s_t, a_t)$ is called action-value function for policy π , which can describe the value of state s_t and action a_t at current time step t given policy π so as to determine what action is best in the set of actions under current situation. Further, to find the best policy, we can maximize the $Q_\pi(s_t, a_t)$ as:

$$Q^*(s_t, a_t) = \max_{\pi} Q_\pi(s_t, a_t) \quad (2.12)$$

where $Q^*(s_t, a_t)$ is called optimal action-value function. Meanwhile, we can also evaluate the state at time step t as follows:

$$\begin{aligned}
 V_\pi(s_t) &= \mathbb{E}_A[Q_\pi(s_t, A)] \\
 &= \begin{cases} \sum_a \pi(a|s_t) * Q_\pi(s_t, a), & a \text{ is discrete} \\ \int_a \pi(a|s_t) * Q_\pi(s_t, a), & a \text{ is continuous} \end{cases} \quad (2.13)
 \end{aligned}$$

where $V_\pi(s_t)$ is called state-value function, which can describe the value of current state s_t for policy π .

To solve the RL problem, various algorithms have been proposed, including Q-learning [86], policy gradient methods [87], actor-critic methods [88], and deep reinforcement learning [89]. Q-learning is a model-free RL algorithm that learns the action-value function $Q(s, a)$ directly, while policy gradient methods aim to optimize the policy by directly computing gradients of the expected reward with respect to the policy parameters. Actor-critic methods combine both value-based and policy-based methods and use a critic to estimate the value function and an actor to select actions. DRL extends traditional RL methods by using deep neural networks to represent the policy or value function.

One of the challenges in RL is the exploration-exploitation trade-off, where the agent needs to balance between taking actions that maximize the expected reward and exploring new actions that may lead to better long-term performance. Various exploration strategies have been proposed, including epsilon-greedy, softmax, and Upper Confidence Bound (UCB) methods. Another challenge is the curse of dimensionality, where the number of states and actions grows exponentially with the number of system variables, making it difficult to learn the optimal policy for large-scale systems. To address this challenge, various dimensionality reduction techniques, such as function approximation, coarse-graining, and hierarchical RL, have been proposed.

RL has been applied to various power system problems, including unit commitment, economic dispatch, power system scheduling, and demand response. RL has been shown to outperform traditional optimization methods in some cases, especially for problems with complex dynamics or uncertain parameters. However, RL also has limitations, including the need for extensive training data, the risk of overfitting, and the lack of interpretability of the learned policy.

In summary, RL is a powerful tool for learning optimal policies for complex systems, including power systems. RL is a model-free, data-driven approach that can learn from experience and adapt to changing environments. Various RL algorithms and exploration strategies have been proposed, and RL has been applied to various power system problems with promising results. However, RL also has limitations and challenges that need to be addressed, such as the curse of dimensionality, the exploration-exploitation trade-off, and the lack of interpretability. Further research is needed to improve the scalability, robustness, and interpretability of RL for power system applications.

2.3 RL with Guided Surrogate Gradient-based Evolution Strategy (GSES)

The major drawback of most existing RL algorithms is that increasing complexity leads to less robustness due to too many hyper-parameters. For the consideration of robustness and easy adaption, we proposed to use an easy-to-train and robust RL algorithm called Guided Surrogate Gradient-based Evolution Strategy (GSES) [58]. Different from existing model-free DRL algorithms that use action-space exploration, GSES performs policy parameter-space exploration, and estimates the gradient of the returns using simultaneous perturbation stochastic approximation, thus back-propagation is not needed.

The Evolution Strategy (ES) [90] targets at minimizing the opposite of the expected accumulated discounted reward $\mathcal{J}(\theta) = -E\{\sum_{t=0}^T \gamma^t r_t(s_t, \pi(s_t|\theta))\}$ with respect to the

parameters θ of the policy π by obtaining some estimate of the (smoothed) gradient of $\mathcal{J}(\theta)$ that provides a good parameter update direction for θ . The ES first applies the Gaussian smoothing to obtain a smoothed function $\mathcal{J}_\sigma(\theta)$ as:

$$\mathcal{J}_\sigma(\theta) = \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)} \{ \mathcal{J}(\theta + \sigma\epsilon)\epsilon \}, \quad (2.14)$$

where σ is the modulating parameter that controls the size of the smoothing area and $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is the standard n -dimensional Gaussian distribution with zero-mean and identity matrix \mathbf{I} as variance matrix. The gradient of $\mathcal{J}_\sigma(\theta)$ with respect to parameters θ is given by

$$\nabla \mathcal{J}_\sigma(\theta) = \frac{1}{\sigma} \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I)} \{ \mathcal{J}(\theta + \sigma\epsilon)\epsilon \}, \quad (2.15)$$

which can be sampled by a Monte Carlo estimator, see [90], [91]. Often stochastic finite differences [90], [91] is used as the antithetic ES method to estimate the gradient as

$$g_{ES} = \frac{\beta}{2\sigma^2 N} \sum_{i=1}^N \epsilon_i [\mathcal{J}(\theta + \sigma\epsilon_i) - \mathcal{J}(\theta - \sigma\epsilon_i)], \quad (2.16)$$

where ϵ_i are independently sampled from $\mathcal{N}(0, I)$ for $i \in \{1, \dots, N\}$. The overall scale of the estimate β and variance of the perturbations σ^2 are constants, to be chosen as hyperparameters. This antithetic ES estimator solely relies on computing $2N$ function evaluations. However, it tends to have high variance, thus requiring a large number of samples to be practical, and scales poorly with the dimension n [58], [90].

The key idea of the proposed guided surrogate-gradient-based ES is to keep track of a low dimensional subspace, defined by the recent history of surrogate gradients during optimization, which is called the guiding subspace. A finite difference random search (as in ES) is then performed preferentially within this subspace. By concentrating the search samples in a low-dimensional subspace where the true gradient has non-negative support,

the proposed guided ES approach dramatically reduce the variance of the search direction. The guided ES algorithm takes advantage of the surrogate gradient in the following way [58]: suppose we can get a vector of surrogate gradient for the policy parameters at each iteration, then by collecting the surrogate gradients from the previous k iterations, we can generate a subspace $U^T U = I_k$, where U is an $n \times k$ orthogonal basis for this subspace, and n is the dimension of the policy parameters. The gradient information can be further embedded in the ES algorithm by changing the distribution of the perturbation ϵ_i from $\mathcal{N}(0, I)$ to $\mathcal{N}(0, \Sigma)$, where Σ is calculated as follows:

$$\Sigma = \alpha^2 I_n + (1 - \alpha)^2 U U^T \quad (2.17)$$

In (Equation 2.17), α is a weight factor that makes a trade-off between the random search (exploration) and the guided search with surrogate gradient (exploitation). With the modified distribution, the perturbation direction ϵ_i can be calculated as follows:

$$\epsilon_i = \alpha \epsilon' + (1 - \alpha) \epsilon'' \quad (2.18)$$

where $\epsilon' \sim \mathcal{N}(0, I_n)$, and $\epsilon'' \sim \mathcal{N}(0, I_k)$.

Setting $\alpha = 1$ will lead to a perturbation direction $\epsilon_i = \epsilon'$, which is exactly the original ES algorithm. As a result, the search will be in the full space, which reduces the training bias but with increased variance. Similarly, setting $\alpha = 0$ will lead to a perturbation direction $\epsilon_i = \epsilon''$, which means the search will be only in the guided subspace. As a consequence, the training variance can be reduced with a sacrifice in the bias. Therefore, balancing the bias and variance is very important. The squared norm of the bias of the

GSES gradient is

$$\begin{aligned}
|Bias|_2^2 &= (\mathbb{E}[g_{ES}] - \nabla \mathcal{J}_\sigma(\theta))^T (\mathbb{E}[g_{ES}] - \nabla \mathcal{J}_\sigma(\theta)) \\
&= \nabla \mathcal{J}_\sigma(\theta)^T (\beta \Sigma - I)^2 \nabla \mathcal{J}_\sigma(\theta)
\end{aligned} \tag{2.19}$$

Similarly, the variance of the GSES gradient is

$$\begin{aligned}
Var &= \mathbb{E}[g_{ES} g_{ES}^T] - \mathbb{E}[g_{ES}] \mathbb{E}[g_{ES}]^T \\
&= \mathbb{E}[g_{ES}^T g_{ES}] - \mathbb{E}[g_{ES}]^T \mathbb{E}[g_{ES}] \\
&= \beta^2 \nabla \mathcal{J}_\sigma(\theta)^T \mathbb{E}[\epsilon \epsilon^T] \nabla \mathcal{J}_\sigma(\theta) \\
&\quad - \beta^2 \nabla \mathcal{J}_\sigma(\theta)^T \Sigma^T \Sigma \nabla \mathcal{J}_\sigma(\theta)
\end{aligned} \tag{2.20}$$

The above two (Equation 2.19) and (Equation 2.20) can be used to further estimate the normalized bias \tilde{B} and variance \tilde{V} of the GSES gradient as

$$\tilde{B} = (\beta\alpha - 1)^2 + (\beta^2\alpha^2 + 2\beta(1 - \alpha)(\beta\alpha - 1))|\rho|_2^2 \tag{2.21}$$

$$\tilde{V} = \beta^2(\alpha^2 + \alpha) + \beta^2(1 - \alpha)(2 + \alpha)|\rho|_2^2 \tag{2.22}$$

where ρ is correlation coefficient:

$$\rho = \frac{\nabla \mathcal{J}_\sigma(\theta)^T U}{|\nabla \mathcal{J}_\sigma(\theta)|} \tag{2.23}$$

We can see that both the bias and variance consist of two parts: the first part that depends on the search of the full space, and the second part that depends on the guided subspace. In this research, we set α to 0.5 to balance the search on the full and guided subspace, which is similar to the balance between the exploration and exploitation.

CHAPTER 3

GSES-BASED HVDC DAMPING CONTROL SYSTEM [92, 93, 94]

The use of High Voltage Direct Current (HVDC) transmission has gained significant attention in recent years due to its advantages over conventional AC transmission systems in terms of reduced losses, increased power transfer capacity, and improved system stability [95, 96]. HVDC transmission systems are typically used for long-distance power transmission, interconnection of asynchronous power systems, and connection of offshore wind farms to the onshore grid [97] and damping inter-area oscillations in our research [98, 38].

In addition, control of HVDC systems is essential for maintaining the stability and reliability of the interconnected AC system. The control can be centralized or decentralized. In centralized control, all HVDC links are controlled by a single control center, which is responsible for monitoring and controlling the power flow through the entire network. In decentralized control, each HVDC link is controlled by a local control center, which communicates with neighboring control centers to coordinate the power flow in the network. The choice of control strategy depends on various factors such as the size of the network, the complexity of the system, and the availability of communication channels.

3.1 Mathematical Modeling of HVDC

The basic structure of an HVDC transmission system is shown in Figure 3.1. The electric power flows from Area 1 to Area 2. The power electronic rectifier is at the HVDC terminal in Area 1, converting the AC power to DC power, while the inverter is at the HVDC terminal of Area 2, converting the DC power back to AC power. The DC line is modeled using three differential equations that describe the behavior of the DC link voltage and currents at the rectifier and inverter terminals. The equations are given by [99, 100]:

$$L_{dc} \frac{dI_{rec}}{dt} = V_{rec} - R_{dc}I_{rec} - V_{dc}, \quad (3.1)$$

$$L_{dc} \frac{dI_{inv}}{dt} = -V_{inv} - R_{dc}I_{inv} + V_{dc}, \quad (3.2)$$

$$C_{dc} \frac{dV_{dc}}{dt} = I_{rec} - I_{inv}, \quad (3.3)$$

where R, L, C_{dc} are the DC line parameters, I_{rec} and I_{inv} are the currents at the rectifier and inverter terminals, V_{dc}, V_{rec} and V_{inv} are the DC link, rectifier, and inverter voltages, respectively.

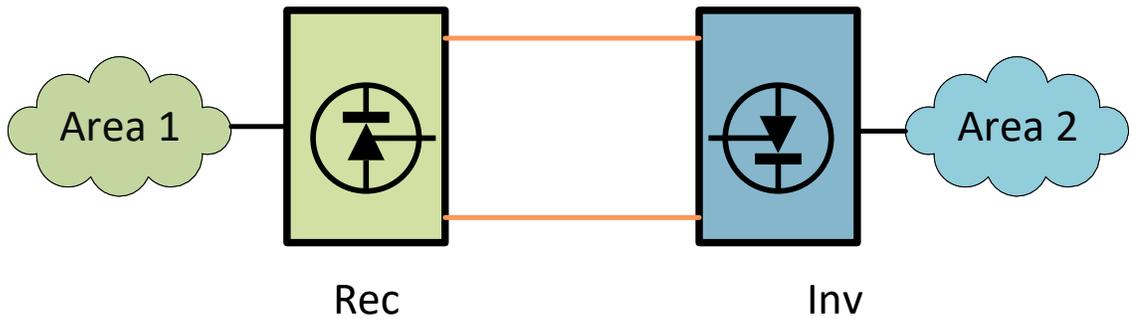


Figure 3.1: Overview of an HVDC transmission system.

The active and reactive power transferred through the HVDC transmission are given by the following equations:

$$P_{rec} = V_{rec}I_{rec} \quad (3.4)$$

$$Q_{rec} = P_{rec} \times \tan \left[\cos^{-1} \left(\frac{\cos(\alpha) + \cos(\pi - \beta)}{2} \right) \right] \quad (3.5)$$

where P_{rec} and Q_{rec} are the active and reactive power, respectively, V_{rec} and I_{rec} are the voltage and current at the rectifier terminal, and α and β are the firing and extinction angles of the converters. The power transfer capability of an HVDC transmission system can be modulated by changing the converter angles.

As a general representation, the voltage V_{dc} and current I_{dc} of a HVDC line can be calculated as [101] and in used in this research:

$$\begin{aligned} V_{dc} &= \frac{3\sqrt{2}}{\pi} N_{rec} V_{rec} \cos\alpha - R_{rec} I_{dc} \\ I_{dc} &= \frac{\frac{3\sqrt{2}}{\pi} N_{rec} V_{rec} \cos\alpha - \frac{3\sqrt{2}}{\pi} N_{inv} V_{inv} \cos\beta}{R_{rec} - R_{inv} + R_{dc}} \end{aligned} \quad (3.6)$$

where V_{rec} and V_{inv} are the AC transformers secondary side voltages of the HVDC rectifier and inverter, N_{rec} and N_{inv} are the number of rectifier and inverter bridges, R_{rec} and R_{inv} are the equivalent commutation resistances of the rectifier and inverter, R_{dc} is the DC line resistance, α and β are the firing angle of the rectifier and the extinction angle of the inverter. Note that we have to guarantee $180^\circ - \alpha - \beta > \phi$ to avoid commutation failures, where ϕ is a positive value that stands for overlap angle.

3.2 Controller

3.2.1 HVDC-based Oscillation Damping Controller

The control of the HVDC transmission system is critical for ensuring reliable and efficient operation. The control system includes a number of subsystems such as the rectifier and inverter controls, DC link voltage control, and active and reactive power controls. The rectifier and inverter controls are responsible for maintaining the desired DC voltage and current levels and for regulating the converter angles to control the power transfer. The DC link voltage control regulates the DC voltage to the desired level, while the active and reactive power controls adjust the active and reactive power transfer to meet the system requirements. For example, if the firing angle is increased, the DC voltage will increase, which increases the power flow in the DC line, but decreases the power transfer capability of the AC system. Similarly, decreasing the firing angle will decrease the DC voltage and the power flow in the DC line, but increase the power transfer capability of the AC system. On the other hand, varying the extinction angle will affect the reactive power flow, which can be used to control the voltage profile of the system.

Modulating the transferred DC power over the HVDC line is an effective means of controlling inter-area oscillations caused by unbalanced power flows. The power transferred over the line can be expressed as $P_{dc} = V_{dc}I_{dc}$ and can be modulated by adjusting the firing angle α of the rectifier or the extinction angle β of the inverter. Typically, the rectifier controls the DC current, while the inverter maintains the DC voltage. As such, adjusting the value of α provides a simple control method for regulating the HVDC power and mitigating inter-area oscillations.

Figure 3.2 illustrates a reliable power oscillation damping approach proposed by Schoenwald [102] using HVDC technology. The proposed approach calculates the frequency difference Δf between the phase angles at the rectifier and inverter terminals using a filter.

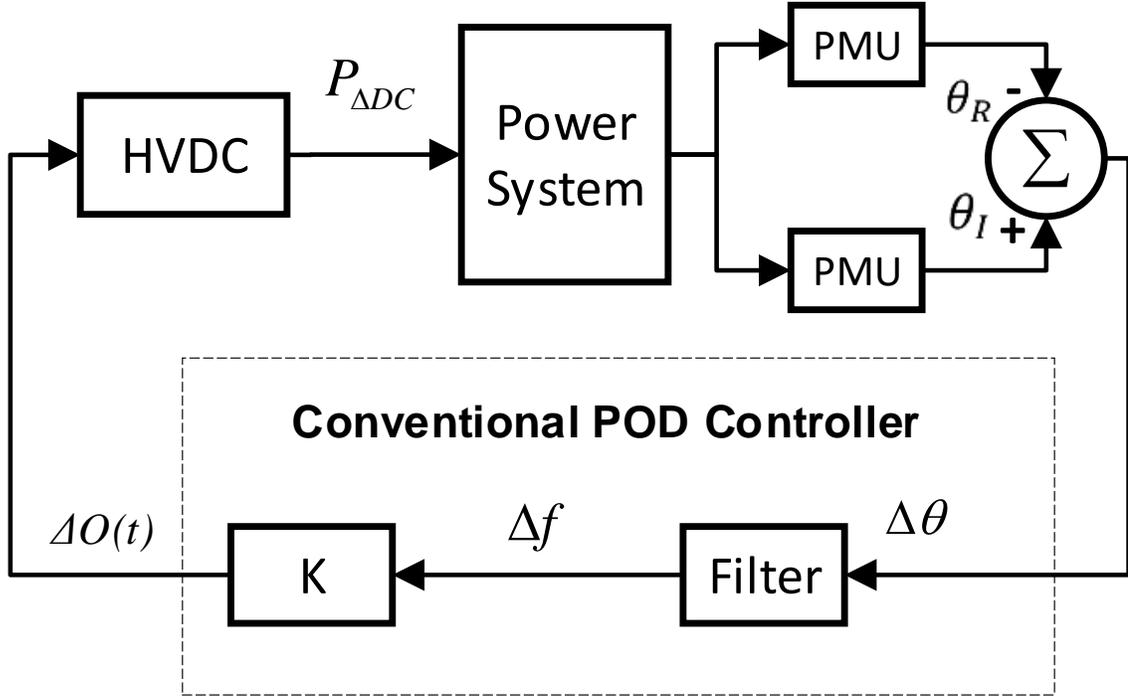


Figure 3.2: Reliable HVDC-based oscillation damping approach proposed by Schoenwald [102, 92]

The frequency difference Δf is then amplified by a proportional gain K to derive the oscillation signal $\Delta O(t)$, as shown in (Equation 3.7). Here, f_{rec} and f_{inv} denote the frequencies at the rectifier and inverter terminals, respectively. Finally, the oscillation signal $\Delta O(t)$ is fed to the rectifier to modulate the change of DC power $P_{\Delta dc}$, which effectively damps inter-area oscillations.

$$\Delta O(t) = K\Delta f = K(f_r - f_i) \quad (3.7)$$

Although the proposed approach is a straightforward and effective power oscillation damping controller that captures the inter-area oscillation information from the frequency difference Δf , it has limitations in terms of optimal damping performance.

3.2.2 Limitations and Challenges of HVDC-based Oscillation Damping Controller

Difficulty in determining K

Finding the optimal value of K is a challenging task, as it requires a small-signal stability analysis to determine it. This analysis is based on the linearized power system model, which is represented as:

$$\begin{bmatrix} \Delta \dot{x} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \quad (3.8)$$

where x is the n -dimensional state vector and y is the m -dimensional measurement vector. The system matrices are given by $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{n \times m}$, $\mathbf{C} \in \mathbb{R}^{m \times n}$, and $\mathbf{D} \in \mathbb{R}^{m \times m}$, respectively, for the linearized small-signal state-space equations. One can eliminate Δy from (Equation 3.8) by rewriting it in terms of the state-space matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , and \mathbf{D} as follows:

$$\Delta \dot{x} = \mathbf{\Lambda} \Delta x \quad (3.9)$$

where $\mathbf{\Lambda} = \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}$. The eigenvalue analysis of the matrix $\mathbf{\Lambda}$ can roughly indicate the oscillation modes of a system.

The value of K must be tuned according to the specific oscillation modes. However, the small-signal stability analysis is based on a linearized power system model, which is often difficult to obtain. Moreover, the power system model is nonlinear in nature, and the small-signal analysis ignores its nonlinear characteristics. This can cause issues during the implementation of the HVDC-based inter-area oscillation damping controller.

Limitation of fixed K

Despite being able to determine the value of K , the effectiveness of the oscillation damping controller cannot be guaranteed at all times. This is because various transient events, such as short circuit faults, line losses, sudden increases in loads, or generator losses, can trigger inter-area oscillations. These events can result in different oscillation modes or levels of inter-area oscillations. Because distinct oscillations require different damping reactions, the performance of the oscillation damping controller with a fixed K value may function effectively for certain oscillation scenarios but not for others. Indeed, our research found that the optimal K for some oscillation scenarios may even lead to instability issues in other oscillation scenarios. The fixed value of K limits its ability to adapt to diverse transient events.

Insufficiency of $\Delta O(t) = K\Delta f$

Although using the frequency difference of the two HVDC terminals Δf may aid in stabilizing the power system via the damping controller, the application of a single frequency difference-based expression $\Delta O(t) = K\Delta f$ is not optimal. This is due to the fact that the frequency difference Δf is susceptible to various sources of noise and may contain a combination of multiple oscillation modes. The utilization of $\Delta O(t) = K\Delta f$ may inadvertently amplify the noise or negatively impact non-dominant oscillation modes. Therefore, a more sophisticated approach is necessary, such as the incorporation of more wide-area system information in the design of $\Delta O(t)$.

3.3 Design of GSES-based HVDC Damping Control Strategy

The conventional controllers face several limitations and challenges, and to overcome these, we propose a novel approach based on GSES for inter-area oscillation damping. In

this work, we introduce the GSES algorithm, which is an improved version of the Basic Random Search (BRS) algorithm (see Appendix A). Our proposed approach is to develop a model-free controller that can dynamically adjust its parameters to achieve optimal damping performance during various transient events. To achieve this, we utilize the GSES-based RL method for adjusting the parameters of the HVDC damping controller. Furthermore, we aim to address the insufficiency issue of $\Delta O(t) = K\Delta f$ by introducing an improved oscillation signal extraction method inspired by our previous work in [103][104]. We propose to calculate the oscillation signal $\Delta O(t)$ using the bus frequencies at multiple locations, as follows:

$$\Delta O(t) = \sum^n K_i f_i = \begin{bmatrix} K_1 & K_2 & \dots & K_n \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \quad (3.10)$$

where K_1, K_2, \dots, K_n are the parameters of the proposed controller, and f_1, f_2, \dots, f_n are the corresponding frequencies at the oscillation regions. It is worth noting that the equation presented can be expressed as $\Delta O(t) = K\Delta f$ when $n = 2$ and $K_1 = -K_2$. In our previous work [103], we have demonstrated that the proposed controller is advantageous in terms of its ability to significantly reduce noise and negative damping effects. Furthermore, the use of multiple frequencies f_1, f_2, \dots, f_n provides the damping controller with comprehensive information about inter-area oscillations, which enhances its performance beyond that of using only the frequency difference at the HVDC two terminals.

Figure 3.3 shows the proposed GSES-based HVDC oscillation damping controller. As a RL-based control approach, the GSES-based damping controller also requires a *state space* and an *action space* and *reward function*.

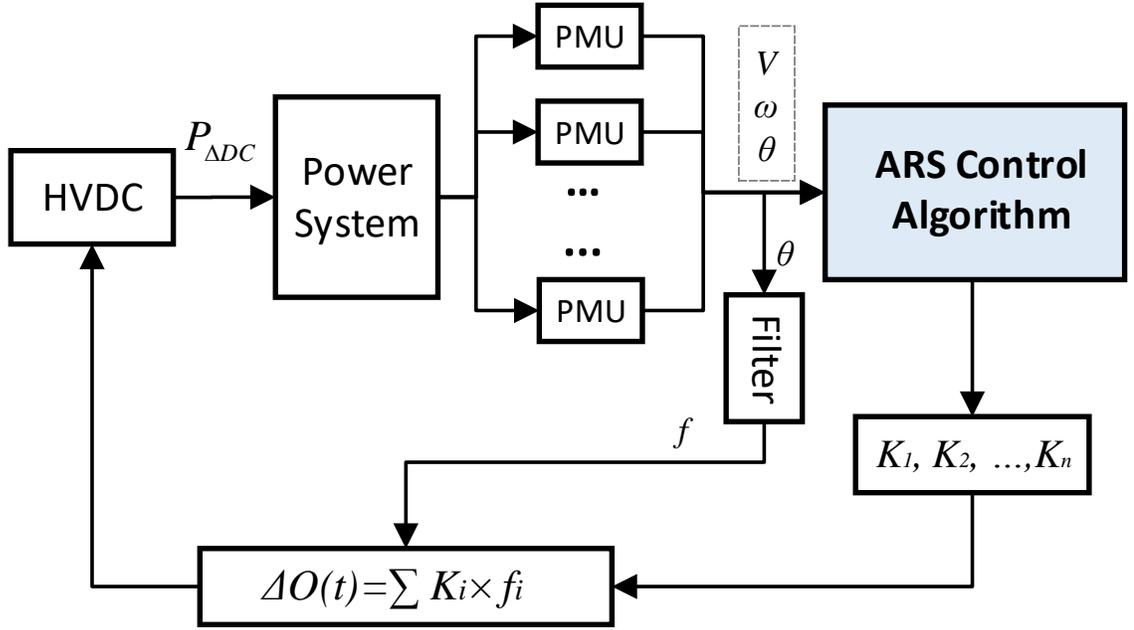


Figure 3.3: Proposed GSES-Based HVDC Oscillation Damping Control Approach [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]

3.3.1 State Space

In order to enhance the performance of the HVDC Damping Control under different scenarios using the GSES algorithm, certain variables that represent the oscillation system's operation status need to be extracted. To achieve the objective of reducing system frequency fluctuation, the bus frequency f_{osc} , the machine speeds ω_{osc} at oscillation areas. The wide-area system voltage magnitudes v_{osc} , phase angles θ_{osc} are also included to better deficit the oscillation situation. Therefore, the state space that contains the most useful information from the inter-area oscillation system is expressed as follows:

$$\mathcal{S} = \{f_{osc}, \omega_{osc}, v_{osc}, \theta_{osc}\} \quad (3.11)$$

Please note that all the elements in the state space may represent a vector which containing multiple variables at different buses in the system.

3.3.2 Action Space

The action space is designed to dynamically regulate the power of HVDC line through the damping controller, thus the controller parameters K_1, K_2, \dots, K_n are included in the action space as follows:

$$\mathcal{A} = \{K_1, K_2, \dots, K_n\} \quad (3.12)$$

3.3.3 Reward Function

To ensure the GSES algorithm converges quickly and identifies the optimal policy π_{θ}^* , it is crucial to have a well-designed reward function. In our research, we analyzed the inter-area oscillation characteristics and selected two parameters, namely the machine speed difference and the HVDC terminal frequency difference, for creating the reward function. During normal operating conditions, the generators' machine speeds are synchronized, resulting in a zero difference. However, during inter-area oscillations, the machine speeds of generators in one area would oscillate against those in the other area, resulting in a significant difference. Similarly, the frequency difference (Δf) at the HVDC two terminals is usually zero under normal conditions but could provide important information about inter-area oscillations during transient events. Therefore, we selected the machine speed differences and the frequency difference at the HVDC terminals as indicators of inter-area oscillations. Moreover, the magnitudes of these indicators would increase with the severity of the inter-area oscillations. After considering these factors, we developed a reward function that could effectively train the GSES algorithm as follows:

$$r_t = \lambda_1 \sum_i^D |\omega_{A,i}(t) - \omega_{B,i}(t)|^2 + \lambda_2 |\Delta f(t)|^2 \quad (3.13)$$

where area A consists of D selected generators with machine speeds $\omega_{A,i}$, while Area B includes D selected generators with machine speeds $\omega_{B,i}$. The generators in both areas oscillate against each other. The negative reward coefficients λ_1 and λ_2 are used to balance the contribution of oscillation rewards in machine speed and frequency difference. Proper setting of λ_1 and λ_2 is crucial to achieve the desired balance.

3.4 GSES-based Damping Control Algorithm

The simulation of the HVDC-based inter-area system is performed using the MATLAB/PST [105] toolbox environment. The predefined states s_t in (Equation 3.11) are collected by the RL agent implemented as a user-defined function in MATLAB/PST at each time step t . The reward function in (Equation 3.13) is utilized to calculate the reward r_t at each time step. The proposed GSES algorithm uses a neural network with parameters θ to represent the policy π_θ . The output action is determined based on the trained policy π_θ and the values of the states, as defined in (Equation 3.12). After each simulation episode, the episode reward is calculated by accumulating the discounted reward, as defined in (Equation 2.8). The accumulated rewards are used to estimate the gradient of the network parameters and update the policy.

To address the issue of unequal weights on different states due to their varying scales, a state normalization process is performed in the GSES algorithm. The mean and covariance of the corresponding state are used for state normalization, as shown in (Equation 3.14):

$$s_{t,norm} = \frac{1}{\sqrt{\text{diag}(\sigma)}} (s_t - \mu) \quad (3.14)$$

Here, σ and μ are the covariance matrix and mean vector of the states s in the previous episode, respectively. The algorithm explores the parameter space of the policy for multiple directions $\epsilon_{j,+}$ and $\epsilon_{j,-}$ for $j \in 1, 2, \dots, N$ in opposite directions. The m directions with the highest average episode reward are selected as the guided subspace direction. The full search direction and guided subspace are balanced by calculating the distribution of perturbation and updating the policy, following the GSES algorithm described in section 2.3.

The algorithm of GSES is outlined in Algorithm 1. Firstly, the GSES policy initializes the neural network and perturbs the parameter θ with N random samples $\delta_1, \delta_2, \dots, \delta_N$. The neural networks correspond to N MATLAB/PST instances, and Ray software [106] is employed to control the N PST simulators in parallel. Each PST simulator performs a power system simulation independently. The system states such as bus voltage magnitudes V , phase angles δ , and machine speeds ω are used as inputs to the neural network during the simulation and training. The state normalization in Eq (Equation 3.14) is carried out to improve the neural network's training. The inferred parameters of the controllers K_1, \dots, K_n are the outputs of the neural network. Based on the actions of the RL agent and the chosen frequencies f_1, \dots, f_n , the oscillation signal $\Delta O(t)$ is calculated, and the HVDC-based oscillation damping controller adjusts the transferred power to damp the inter-area oscillations. Note that during the simulation, the parameters of the controllers K_1, \dots, K_n are dynamically adjusted based on the system states V, ω, δ . After all the N MATLAB/PST simulations are completed, the rewards $R(\pi)$ are gathered and sorted in descending order. The top M pairs of neural networks are used to calculate the parameter gradient $\Delta\theta$ as per (Equation 3.15).

$$\Delta\theta = \frac{1}{M\sigma_M} \sum_{k=1}^M [R(\pi'_{\theta+\epsilon\delta_k}) - R(\pi'_{\theta-\epsilon\delta_k})] \delta_k \quad (3.15)$$

Once the parameter θ is updated, the GSES perturbs the neural network once more to generate N new random samples $\delta_1, \delta_2, \dots, \delta_N$, and the new simulations with N MATLAB/PST instances begin again. The GSES training process repeats until convergence is achieved.

Algorithm 1 GSES-based HVDC Oscillation Damping Control [92]

Require: α : learning rate; reward functions; N : parallel worker number.

Ensure: Trained parameters θ .

- 1: Initialize θ_0 .
 - 2: **while** training not converged **do**
 - 3: Generate $\delta_1, \delta_2, \dots, \delta_N$ random samples.
 - 4: Perturb the neural network to generate N pairs of $\pi_{\theta+\epsilon\delta}$ and $\pi_{\theta-\epsilon\delta}$.
 - 5: Calculate the GSES-based network policy:
 - 6: $\pi'_{\theta+\epsilon\delta_k}(s) = \pi_{\theta+\epsilon\delta_k}(\text{diag}(\Sigma)^{-1/2}(s - \mu))$
 - 7: $\pi'_{\theta-\epsilon\delta_k}(s) = \pi_{\theta-\epsilon\delta_k}(\text{diag}(\Sigma)^{-1/2}(s - \mu))$
 - 8: Run the simulation until it finishes at T and collect the N pairs of rewards $R(\pi'_{\theta+\epsilon\delta})$ and $R(\pi'_{\theta-\epsilon\delta})$.
 - 9: Sort the directions δ_k by $\max \{R(\pi'_{\theta+\epsilon\delta}), R(\pi'_{\theta-\epsilon\delta})\}$ and select the top- M pairs.
 - 10: Calculate the gradient:
 - 11: $\Delta\theta = \frac{1}{M\sigma_M} \sum_{k=1}^M [R(\pi'_{\theta+\epsilon\delta_k}) - R(\pi'_{\theta-\epsilon\delta_k})] \delta_k$
 - 12: Update the parameter θ : $\theta \leftarrow \theta + \alpha\Delta\theta$.
 - 13: **end while**
-

3.5 Implementation Platform and Parallel Computation

The power system simulation toolbox PST V3.0 [105] is used to implement the proposed approach based on the GSES algorithm. It is worth noting that the PST simulator is based on MATLAB while the GSES algorithm is implemented in Python. To address the communication issue, we established an Application Programming Interface (API) to facilitate the interaction between the GSES and PST simulators. Figure 3.4 illustrates the flowchart of the proposed GSES-based HVDC damping control approach.

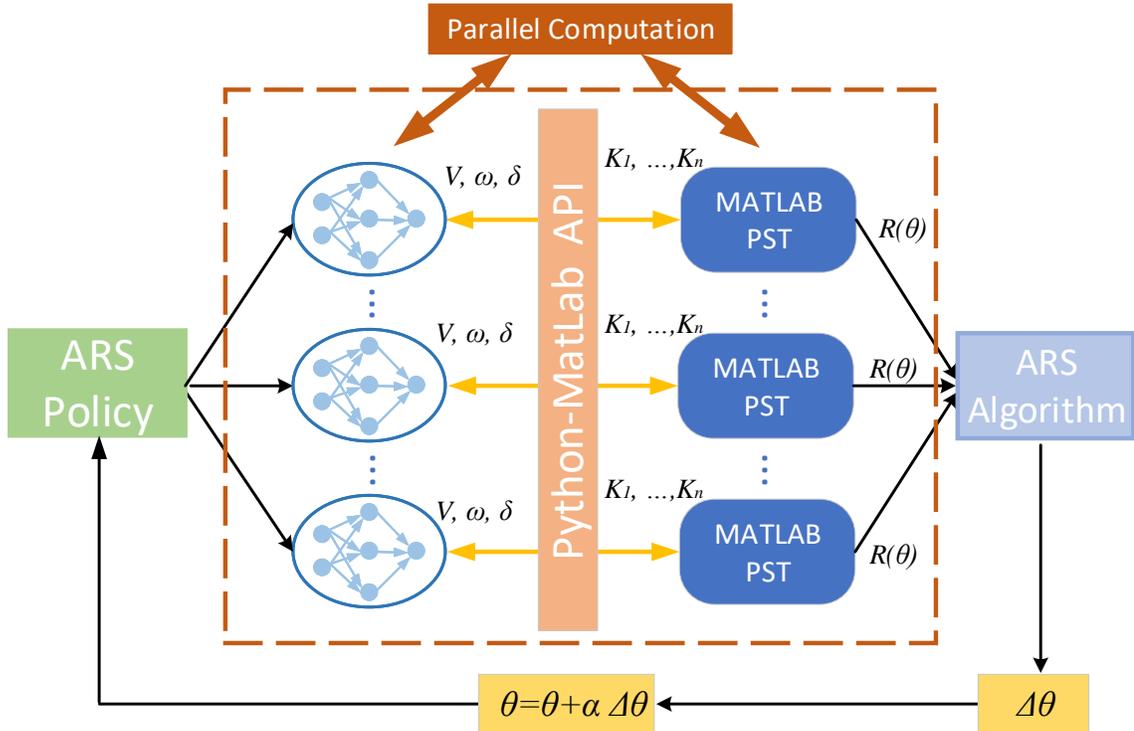


Figure 3.4: Flowchart of proposed GSES-based HVDC damping control approach [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]

To demonstrate the convergence of the GSES-based HVDC damping controller algorithm, we can focus on the reward improvement criterion. Figure 3.5 illustrates an example of the reward obtained during the training process.

The reward curve represents the performance of the algorithm over iterations. The GSES algorithm aims to improve the controller's performance by leveraging the top- M perturbed neural networks to update its parameters iteratively. As training progresses, the algorithm explores different perturbation directions, evaluates their corresponding policies, and collects rewards for each policy. By comparing these rewards, the algorithm identifies the most promising directions and adjusts the parameters accordingly.

Examining the reward curve, we observe a notable improvement in performance. Initially, as the algorithm explores various perturbation directions, the rewards may fluctuate. However, as the training continues, the algorithm learns to select the perturbation directions that lead to higher rewards. This results in a smoother reward curve with a positive upward trend.

The convergence of the algorithm can be inferred from the reward curve. After about 60 iterations, the reward reaches a plateau over iterations, indicating that the algorithm is converging towards an optimal solution.

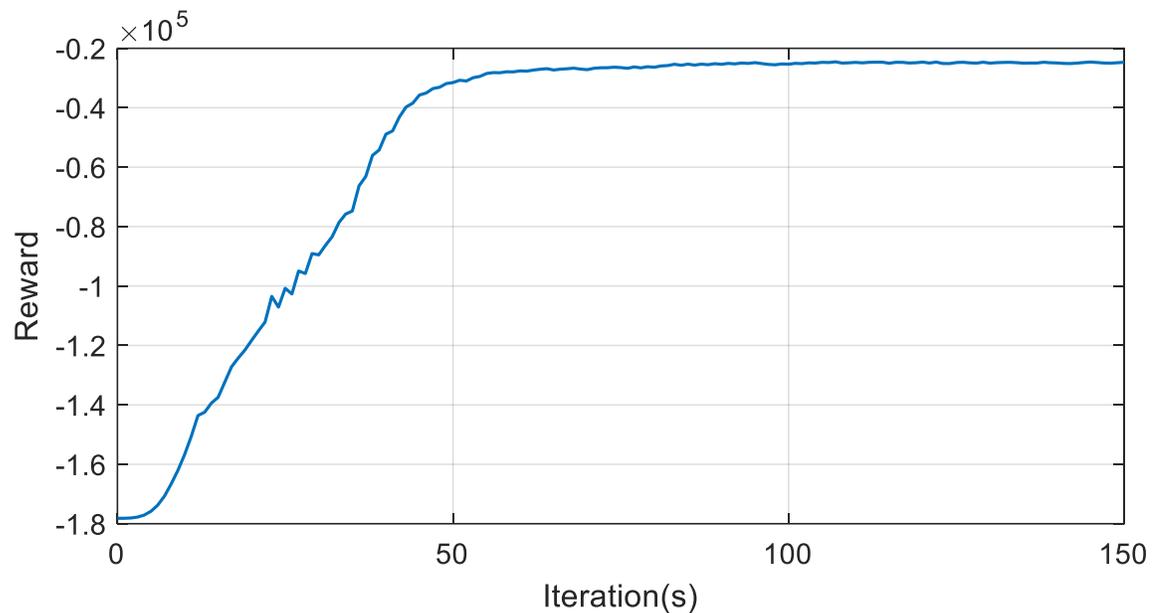


Figure 3.5: An example of the reward of training the GSES-based controller [92]

3.6 Case Study and Discussion

3.6.1 Revised minniWECC System

In order to validate the efficacy of the new GSES-based HVDC oscillation damping control strategy, we have applied the proposed controller to the minniWECC system

[107]. The minniWECC system is a simplified version of the Western Interconnection system located in North America, which encompasses numerous generators, buses, and lines. Despite being a simplified model, the minniWECC system comprises 34 generators, 120 buses, and 171 transmission lines, and it encompasses the most significant oscillation modes of the authentic Western Interconnection system, making it an excellent testbed for the proposed oscillation damping control techniques.

The revised minniWECC system incorporates an HVDC transmission (represented by a dashed line) that connects Bus 24 (rectifier) to Bus 49 (inverter) as depicted in Figure 3.6. This transmission line represents the Pacific DC Intertie (PDCI) in reality. The PDCI is a ± 500 kV DC transmission system that can handle up to 3 GW of power.

To suppress inter-area oscillations, we have implemented the proposed GSES-based approach on the PDCI transmission line. The proposed approach has been compared to a conventional HVDC oscillation damping controller described in [102]. We conducted a small-signal stability study of the minniWECC system, as described in subsection 3.2.2, to establish the appropriate parameters for the conventional controller. The outcomes of the study are presented in Figure 3.7.

Figure 3.7 shows multiple inter-area oscillation modes, with the 0.3 Hz Alberta mode and 0.65 Hz BC mode being the dominant modes. The Alberta and BC modes correspond to oscillations between generators in the north and south regions. It is possible to effectively dampen these inter-area oscillations by regulating the power flow through the PDCI, which connects the north and south regions.

Small-signal studies, as depicted in Figure 3.7, indicate that the conventional controller was appropriately tuned to enable a fair comparison with the proposed GSES-based controller. The GSES-based controller utilized a multi-layer neural network with two hidden layers comprising 64 and 32 neurons, respectively. The GSES-based controller was trained

using an 8-core CPU with a learning rate of 0.02. Additionally, parallel training was implemented by engaging eight pairs of perturbed policies simultaneously.

The reward for the GSES-based controller was computed using (Equation 3.13), where the machine speeds of Generators 1, 2, 3, 5, 7, 8, 10, and 34 in the north region, and Generators 19, 21, 22, 23, 25, 26, 27, and 28 in the south region were selected. In contrast, the frequencies at Bus 24, 18, 79, 49, 61, and 66 were chosen as input signals for the damping controller in the proposed GSES-based method. For the conventional control methods described in [102], (Equation 3.7) was utilized to generate oscillation damping signals, where f_r and f_i corresponded to the frequencies at Bus 24 (rectifier) and 49 (inverter), respectively. The proportional gain K was set to 4,760, and the value of K was tuned using genetic algorithms to achieve optimal performance for the conventional control method. The two reward coefficients were $-50,000$ and $-1,000$.

3.6.2 Oscillation Scenarios and Damping Results

This research examines four inter-area oscillation scenarios that result from various transient events:

- In *Scenario I: loss of line*, the 500 kV transmission line connecting Bus 99 and 115 is abruptly cut off at time $t=1.0$ sec. This can occur due to breaker malfunction or cyberattacks.
- In *Scenario II: single-phase fault*, a Single Line to Ground (SLG) fault occurs on the transmission line between Bus 89 and 110 at time $t=1.0$ sec. The fault is cleared by tripping the Line 89-110 off 5 cycles after the fault.
- In *Scenario III: loss of generator*, Generator 23 is suddenly disconnected at time $t=1.0$ sec. This can occur due to overloading, breaker/relay malfunction, or cyberattacks.

- In *Scenario IV: three-phase fault*, a three-phase fault occurs on the transmission line between Bus 44 and 90 at time $t=1.0$ sec. The fault is cleared by tripping the line off 4.8 cycles after the fault.

All four scenarios result in different inter-area oscillations in the minniWECC system.

The conventional HVDC oscillation damping controller and the proposed GSES-based controller were both used to damp the inter-area oscillations. The results of the damping performance for each control option were presented in Figure 3.8-Figure 3.9. The oscillations in the machine speed differences and HVDC terminal frequency differences reveal that multiple inter-area oscillation modes were triggered in different scenarios. The dominant mode observed was the 0.3 Hz Alberta mode. Without any oscillation damping controller, the inter-area oscillations were severe and remained undamped until the end of the simulation. Both the conventional and GSES-based controllers effectively dampened the inter-area oscillations. However, the GSES-based controller was able to damp the oscillations much quicker than the conventional controllers in all four scenarios.

3.6.3 Case Studies and Discussion

Table 3.1 summarizes the total rewards obtained for each control method in the four scenarios. It can be observed that both the conventional controller and the GSES-based controller were successful in enhancing the total rewards:

- Specifically, in *Scenario I*, the reward was -9,465 without any damping control, whereas the conventional control improved it to -3,197, and the GSES-based controller further improved it to -1,324.
- Likewise, in *Scenario II*, the reward was -919,698 without any damping control, and the conventional control improved it to -244,543, and the GSES-based controller further improved it to -82,614.

- Moreover, in *Scenario III*, the reward was -37,859 without any damping control, and the conventional control improved it to -9,148, and the GSES-based controller further improved it to -1,645.
- Lastly, in *Scenario IV*, the reward was -2,017,542 without any damping control, and the conventional control improved it to -103,681, and the GSES-based controller further improved it to -21,339.

These findings support our hypothesis that the GSES-based controller is more effective than the conventional controller in mitigating inter-area oscillations.

Table 3.1: Comparison of total reward between different controllers in different scenarios [92]

	No Ctrl	Conv-Ctrl	GSES-Ctrl
Scenario I	-9,465	-3,197	-1,324
Scenario II	-919,698	-244,543	-82,614
Scenario III	-37,859	-9,148	-1,645
Scenario IV	2,017,542	-103,681	-21,339

The rewards' magnitudes in *Scenario IV: three-phase fault* stood out as significantly larger than those of other scenarios. This is due to the absence of any applied oscillation damping control method in this scenario, resulting in the system's gradual loss of stability as the oscillation magnitudes in speed or frequency differences increase exponentially. Conversely, the conventional or proposed oscillation damping control approaches can restore the system's stability.

3.6.4 Prony Analysis and Frequency Spectrum Analysis

In this study, oscillation signals were analyzed using Prony analysis and frequency spectrum analysis [108]. Table 3.2 presents a comparison of dominant oscillation modes in

Table 3.2: Comparison of dominant oscillation modes in different scenarios via Prony Analysis [92]

	No Ctrl	Conv-Ctrl	GSES-Ctrl	Frequency
S I Damp Ratio	3.85%	7.81%	16.73%	0.307 Hz
S II Damp Ratio	2.24%	7.20%	11.34%	0.294 Hz
S III Damp Ratio	3.71%	8.34%	15.48%	0.298 Hz
S IV Damp Ratio	-1.43%	7.01%	12.31%	0.301 Hz

different scenarios. The results indicate that the 0.3 Hz Alberta mode is the dominant mode in all four scenarios. Without an oscillation damping method, the inter-area oscillations exhibit a low damping ratio of less than 4%. Conventional control can improve the damping ratio to approximately 8%, while the GSES-based control can further improve it to above 11%.

In Scenario IV, the damping ratio of the inter-area oscillation is negative, which indicates that the system will lose its stability if no oscillation damping control method is implemented in time. Fortunately, both the conventional and proposed approaches can stabilize the system. Table 3.2 shows that the proposed GSES-based approach is more effective in improving the inter-area oscillation damping performance with HVDC transmission.

Figure 3.10 shows the results of the frequency spectrum analysis for all four scenarios. For a better visualization of the damping performance of the controllers, the maximum oscillation magnitude was set to 1.0 and the other oscillation magnitudes were normalized accordingly. The 0.3 Hz Alberta mode was the dominant mode, and it had the highest magnitude in all scenarios. The second dominant mode was the 0.65 Hz BC mode, which was observed in Scenarios I, II, and IV. Interestingly, only the 0.3 Hz Alberta mode was triggered in Scenario III, while Scenario I exhibited two modes besides the dominant modes.

The spectrum analysis presented in Figure 3.10 further validates the effectiveness of the proposed GSES-based approach in damping various inter-area oscillations.

The feasibility of the proposed GSES-based approach for practical application in real-world scenarios can be attributed to various factors. Firstly, it is an adaptive and reproducible technique, meaning it can adjust to changes in the system's states and produce consistent results. Secondly, once the GSES is trained, the computational burden is significantly low, as the approach does not require a back-propagation process and is model-free, making it highly scalable with few hyper-parameters. Consequently, the proposed approach can adapt to various inter-area oscillations caused by different transient events. Although training the GSES takes a few hours, it can be done offline using parallel computing techniques. Once the training is completed, the well-trained GSES can instantly output the actions based on the system states. Moreover, updating the actions is required only every 0.01 seconds, reducing the computational burden further.

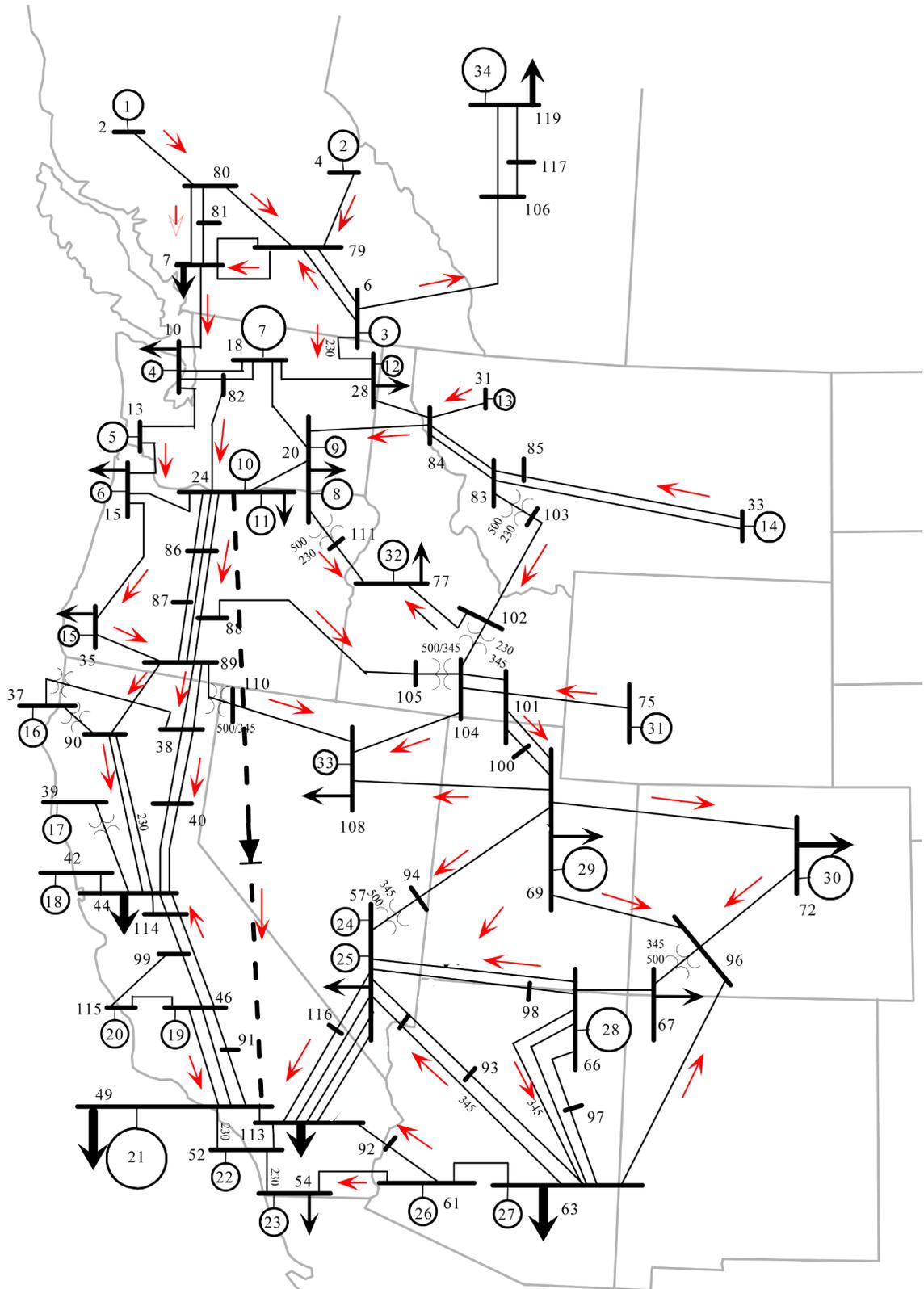


Figure 3.6: Revised miniWECC system with PDCI HVDC transmission [107, 92]

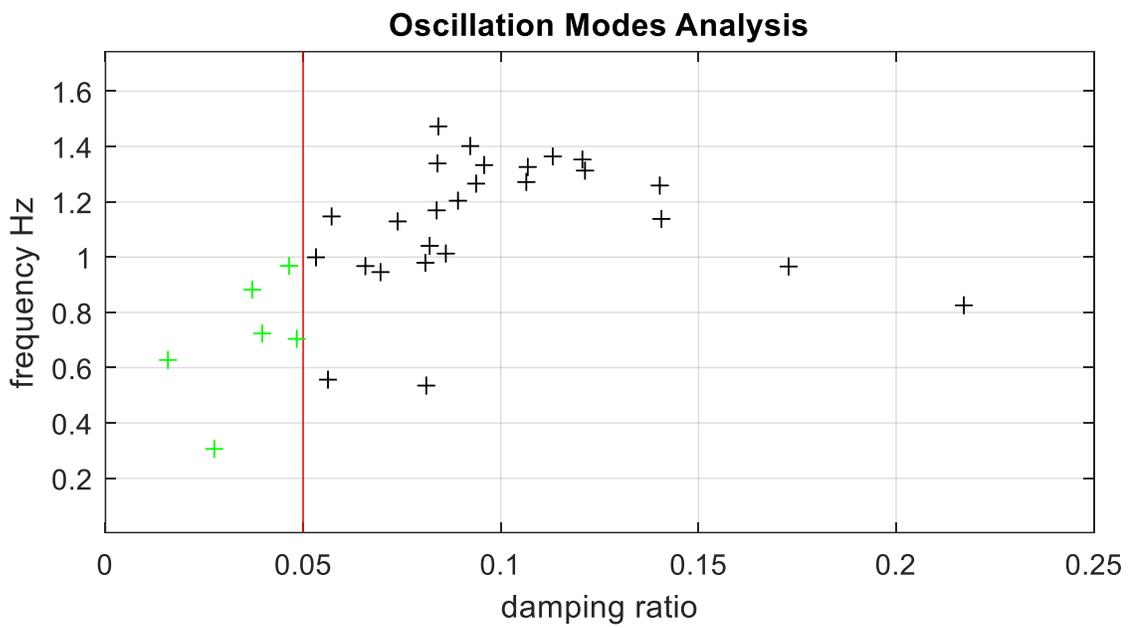


Figure 3.7: Small-signal stability studies of the minniWECC system [92]

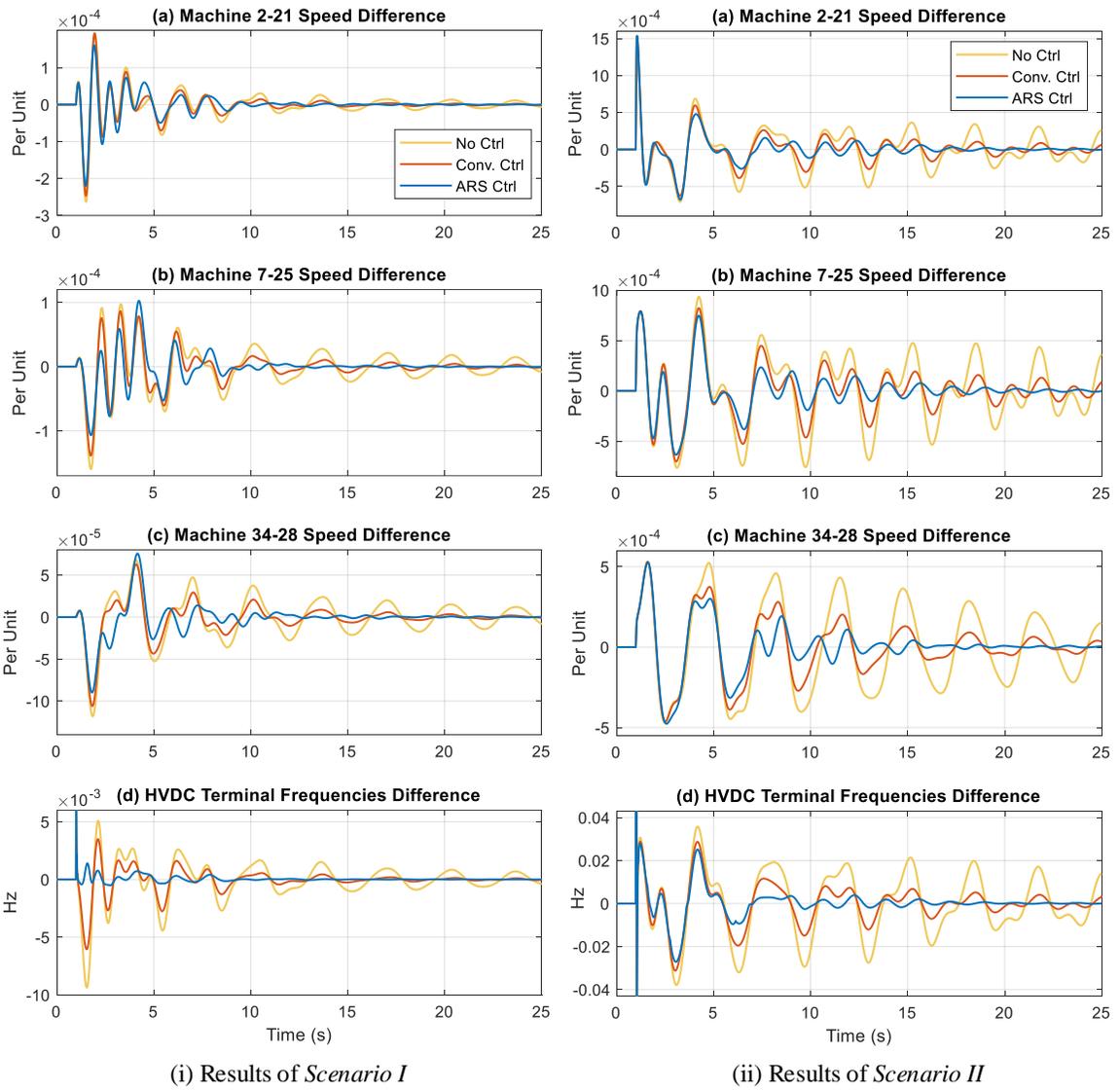


Figure 3.8: Inter-area oscillation damping for (i) Scenario I, (ii) Scenario II [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]

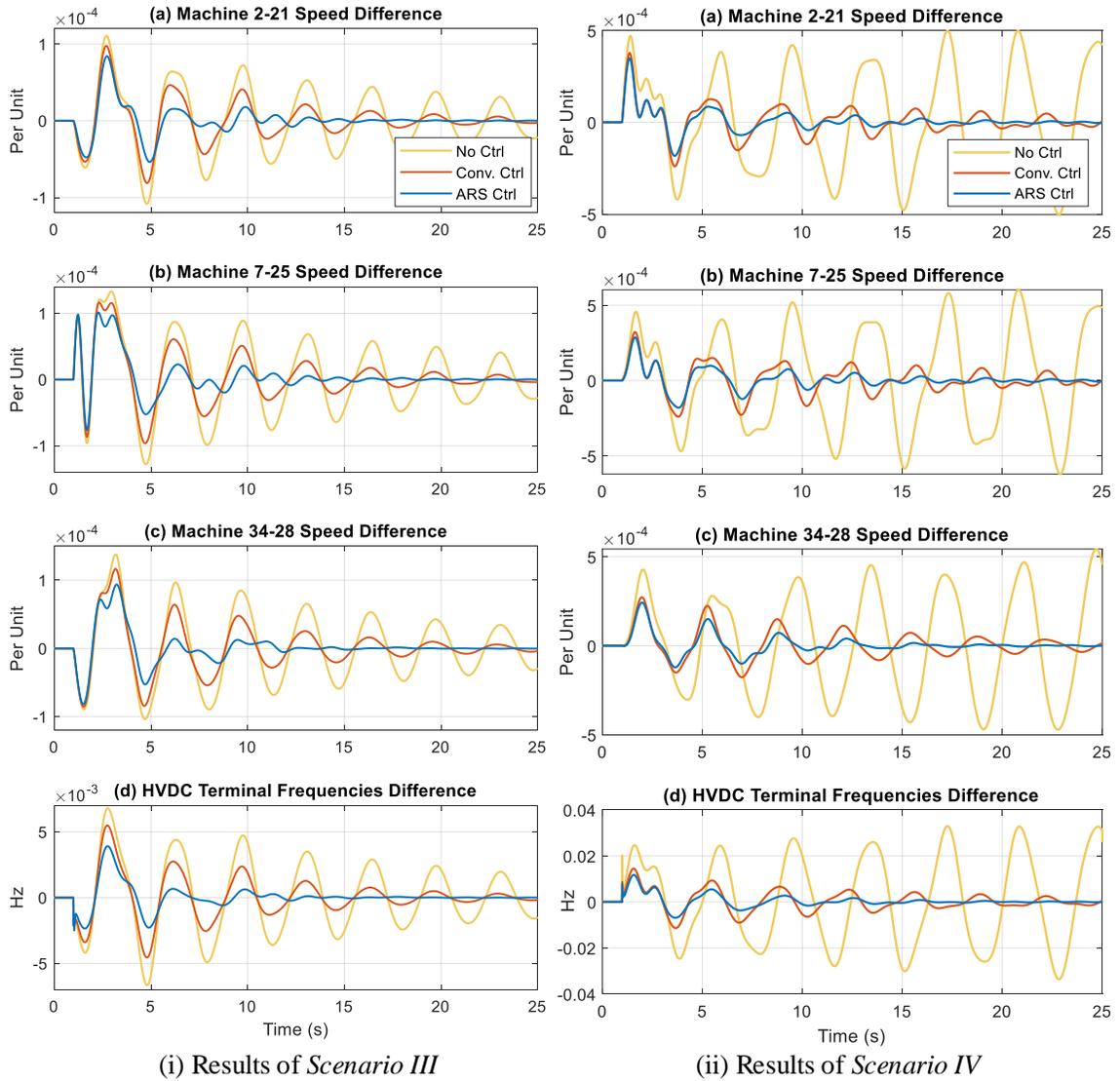


Figure 3.9: Inter-area oscillation damping for (i) Scenario III, (ii) Scenario IV [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]

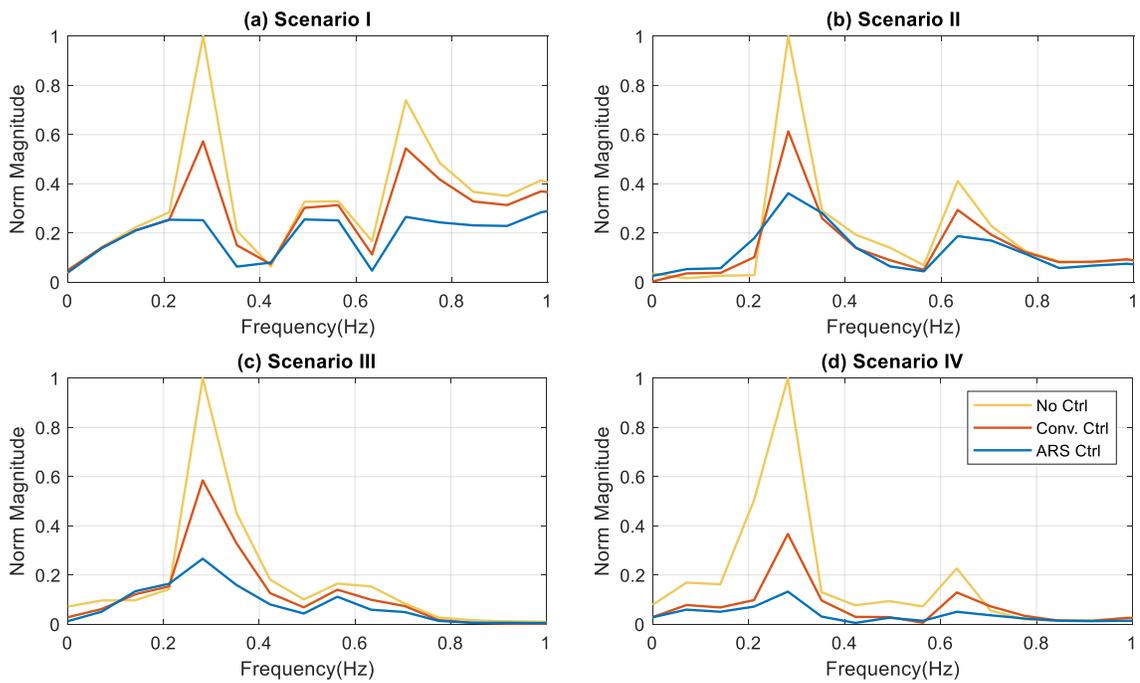


Figure 3.10: Frequency spectrum analysis of the oscillations [92]. Note. ARS here stands for Advanced Random Search [91] and is almost equivalent to GSES [58]

CHAPTER 4

GSES-BASED DFIG-FRT SYSTEM [85]

A typical grid-connected DFIG system configuration is shown in Figure 4.1. Energy from wind turbine rotor is transferred to DFIG through a mechanical shaft system including a gearbox. Rotor of the DFIG is connected to the power grid through a Variable Frequency Converter (VFC) and stator of the DFIG is directly connected to the power grid. To maintain the stable power delivery between DFIG and power grid, the VFC consisting of a Rotor-side Converter (RSC), a Grid-side Converter (GSC) and a DC-link is necessary to be controlled through both RSC controller and GSC controller.

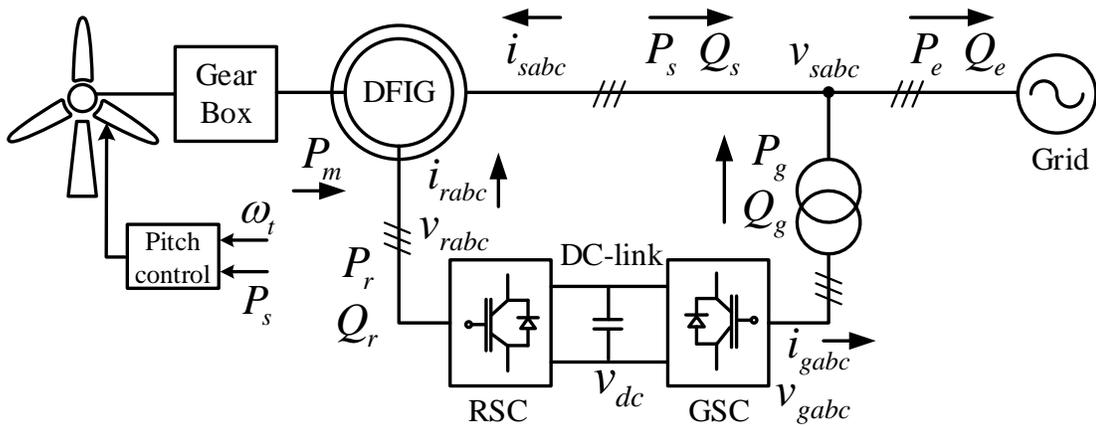


Figure 4.1: Typical DFIG grid connection structure [109, 85]

4.1 Mathematical Modeling of DFIG

4.1.1 Aerodynamics of wind turbine

The power harvested by a wind turbine from the wind can be calculated using the following equation:

$$P_m = \frac{1}{2}(\rho S_w v_w) v_w^2 C_p = \frac{1}{2}(\rho S_w) v_w^3 C_p \quad (4.1)$$

Here, ρ denotes air density, $S_w = \pi R^2$ represents the sweep area of the wind turbine blades with R as the blade radius, v_w is the wind speed, and C_p is the wind power coefficient that is dependent on the Tip-speed Ratio (TSR) $\lambda = \omega_t R / v_w$, the pitch angle of the wind turbine β , and the rotating speed of the wind turbine ω_t . The optimal TSR is denoted as λ_{opt} and the maximum wind power coefficient as $C_{p,max}$. A Maximum Power Point Tracking (MPPT) controller is used to determine the active power reference for the RSC controller to extract the maximum wind power [110, 111]. This can be represented as:

$$P_{m,max} = K \omega_t^3 = \frac{1}{2} \rho R^5 \frac{C_{p,max}}{\lambda_{opt}^3} \quad (4.2)$$

where λ_{opt} is the optimal TSR and $C_{p,max}$ is the maximum wind power coefficient. The MPPT controller is activated when the wind speed is below the rated value, and the pitch angle β is fixed. However, if the wind speed is above the rated value, a pitch angle controller (usually PI controller) is activated to avoid damage to the wind turbine.

4.1.2 Shaft system

The mechanical dynamics of the wind turbine and the DFIG can be modeled using a swing equation:

$$2H_w\dot{\omega}_t = P_m - P_e - D_w\omega_t \quad (4.3)$$

Here, H_w represents the aggregated inertia constant of the shaft system that links the wind turbine and the DFIG, P_e denotes the electromagnetic power of the DFIG, and D_w represents the aggregated damping constant of the shaft system.

4.1.3 Induction generator

The dynamic stator and rotor equations for the Induction Generator (IG) can be represented as [109]:

$$v_{sabc} = R_s i_{sabc} + \dot{\psi}_{sabc} \quad (4.4)$$

$$v_{rabc} = R_r i_{rabc} + \dot{\psi}_{rabc} \quad (4.5)$$

The flux linkages of the stator and rotor under abc -axis are denoted by ψ_{sabc} and ψ_{rabc} respectively. For decoupled controller design, these variables are typically transformed to dq -axis, leading to the following equations:

$$v_{sd} = R_s i_{sd} - \omega_s \psi_{sq} + \dot{\psi}_{sd} \quad (4.6)$$

$$v_{sq} = R_s i_{sq} - \omega_s \psi_{sd} + \dot{\psi}_{sq} \quad (4.7)$$

$$v_{rd} = R_r i_{dr} - (\omega_s - \omega_r) \psi_{rq} + \dot{\psi}_{rd} \quad (4.8)$$

$$v_{rq} = R_r i_{rq} - (\omega_s - \omega_r) \psi_{rd} + \dot{\psi}_{rq} \quad (4.9)$$

where ω_s is the synchronous speed of dq -axis and flux linkages are determined by the stator self-inductance, rotor self-inductance and mutual-inductance between them [109].

Then, the stator and rotor output power can be calculated according to the instantaneous power theory as follows:

$$P_s = \frac{3}{2}(v_{sd}i_{sd} + v_{sq}i_{sq}) \quad (4.10)$$

$$Q_s = \frac{3}{2}(v_{sq}i_{sd} - v_{sd}i_{sq}) \quad (4.11)$$

$$P_r = \frac{3}{2}(v_{rd}i_{rd} + v_{rq}i_{rq}) \quad (4.12)$$

$$Q_r = \frac{3}{2}(v_{rq}i_{rd} - v_{rd}i_{rq}) \quad (4.13)$$

Using the per-unit system, the electromagnetic power can be determined as:

$$P_e = \psi_{sd}i_{sq} - \psi_{sq}i_{sd} = \psi_{rq}i_{rd} - \psi_{rd}i_{rq} \quad (4.14)$$

where ψ_{sd} and ψ_{sq} denote the stator d- and q-axis flux linkages, respectively, and ψ_{rd} and ψ_{rq} denote the rotor d- and q-axis flux linkages, respectively.

4.2 Controller

4.2.1 Rotor Side Controller

In DFIG system, the RSC controller usually guarantees the frequency to be stable, which consists of two feedback loops (inner current loop and outer power or frequency loop) as shown in Figure 4.2. The inner current loop has faster dynamics than the outer loop and is able to regulate the decoupled rotor dq-axis currents i_{rd} and i_{rq} . The outer loop can be either power loop which determines the DFIG electromagnetic output power or the frequency loop (DFIG rotating speed). We choose the power control as the outer loop as the benchmark in this research. The active power reference P_e^* can either be given by MPPT controller or a constant value (usually the rated power of the DFIG). The reactive

power reference Q_e^* can either be zero or a constant non-zero value as a reactive power compensator if there is reactive power demand. With the two cascaded control loops, the output signal v_{rabc}^* will be converted to a 6-channel signal by the PWM generator to drive the IGBTs in the RSC.

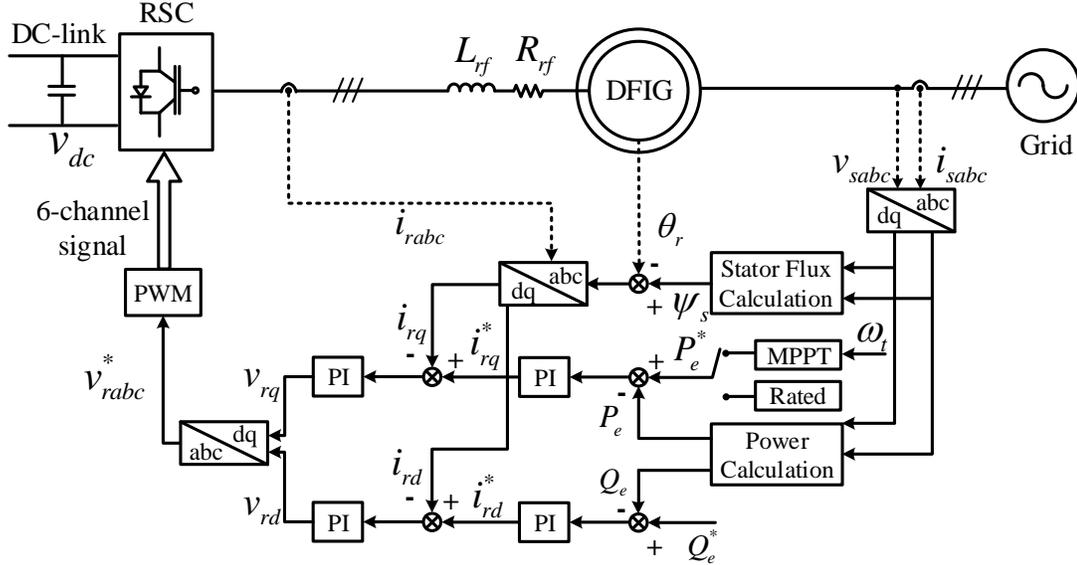


Figure 4.2: Typical DFIG RSC control structure [109, 85]

4.2.2 Grid Side Controller

Different from RSC controller, the control objective of GSC is usually to maintain the DC-link voltage between RSC and GSC, which also consists of two feedback loops (inner current loop and outer DC-link voltage loop) as shown in Figure 4.3. Likewise, the inner current loop has faster dynamics than the outer loop and is able to regulate the decoupled grid dq-axis currents i_{gd} and i_{gq} . The outer loop determines the DC-link voltage and reactive power flow between DFIG and power grid. With the two cascaded control loops, the output signal v_{gabc}^* will be converted to a 6-channel signal by the PWM generator to drive the IGBTs in the GSC.

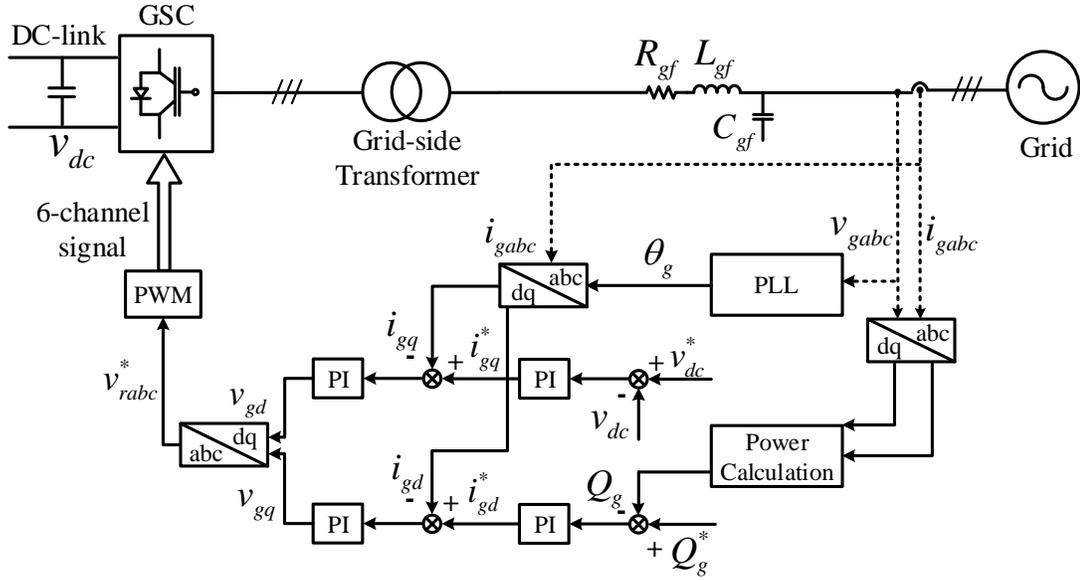


Figure 4.3: Typical DFIG GSC control structure [109, 85]

4.3 Design of GSES-DFIG-based Control Strategy

The grid-connected DFIG system mentioned above can transmit power to the power grid with a stable frequency and voltage under well-designed RSC and GSC controllers during normal operational conditions. However, when a fault occurs at the Point of Common Coupling (PCC) between the DFIG terminal and the power grid, the situation becomes unsatisfactory. In the event of a symmetric or asymmetric fault in the power grid, the DFIG terminal voltage drops suddenly, followed by a grid voltage depression. The mechanical part of the wind turbine has a lower response dynamics due to its inertia, resulting in the original output power. The rotor and stator currents of the DFIG increase very quickly, particularly for the rotor, when the terminal voltage is low. The rotor of the DFIG may be damaged due to the over-current situation. Such faults typically last for approximately 12 cycles (0.2s for a 60-Hz system). Cutting off the DFIG from the power grid can prevent potential damage, but restarting the entire DFIG system is also costly. As a result, an efficient and intelligent method for assisting DFIG FRT is critical.

To avoid over-current situations, we propose regulating the electric power reference signal. Dynamically determining the optimal value of the reference signal is challenging, as it is a complex optimization problem with a lot of non-linearities and coupled effects. In such situations, the RL-based algorithm is appropriate for the problem. We use the GSES algorithm with the following information:

4.3.1 State Space

In order to enhance the performance of the DFIG during fault scenarios using the GSES algorithm, certain variables that represent the DFIG system's operation status need to be extracted. To achieve the objective of reducing rotor currents during faults, both the rotor current i_r and the grid-side current i_g are selected as part of the state space. The reference signals i_{rd} and i_{rq} shown in Figure 4.2 are associated with the active power P_e and reactive power Q_e , respectively. Hence, P_e and Q_e are also included in the state space. The DFIG rotor rotating speed ω_r reflects the power flow exchange between the stator and rotor of DFIG during operation, and the DC-link voltage v_{dc} is included to prevent it from over-voltage during faults. Additionally, the terminal voltage v_g of DFIG is selected to illustrate the fluctuation of the system voltage at the grid connection point during faults. The rotor side voltage is also included to monitor whether overvoltage would happen. Therefore, the state space that contains the most useful information from the DFIG system is expressed as follows:

$$\mathcal{S} = \{i_r, i_g, v_{dc}, v_g, v_s, P_e, Q_e, \omega_r\} \quad (4.15)$$

4.3.2 Action Space

The action space is designed to dynamically regulate the power and DC-link voltage reference values of the RSC and GSC controllers, respectively, as shown in Figure 4.2 and Figure 4.3. Therefore, the action space can be expressed as follows:

$$\mathcal{A} = \{P_e^*, v_{dc}^*\} \quad (4.16)$$

Although control of the reactive power Q_e can support the grid voltage and benefit the DFIG during faults, this research only focuses on the reduction of rotor current and DC-link voltage, and therefore, voltage supporting is not discussed in this dissertation. Thus, the reference for the reactive power is simply set to zero.

4.3.3 Reward Function

To design a well-motivated reward function for the agent to learn desired behaviors from its experiences and explore the environment in a more reasonable direction, the objectives are to prevent rotor current and DC-link voltage from over-loading. Hence, the rotor current magnitude $i_r = \sqrt{i_{rd}^2 + i_{rq}^2}$ and DC-link voltage magnitude v_{dc} are used in designing the reward function, and for consistency, only penalty is applied to the reward function for both partial rewards associated with i_r and v_{dc} , but different coefficients are used:

$$R_{i_r} = \begin{cases} \alpha_{1,i_r}(i_{r,rated} - i_r)^2, & i_r \leq i_{r,rated} \\ \alpha_{2,i_r}(i_{r,rated} - i_r)^2, & i_r > i_{r,rated} \end{cases} \quad (4.17)$$

$$R_{v_{dc}} = \begin{cases} \alpha_{1,v_{dc}}(v_{dc,rated} - v_{dc})^2, & v_{dc} \leq v_{dc,rated} \\ \alpha_{2,v_{dc}}(v_{dc,rated} - v_{dc})^2, & v_{dc} > v_{dc,rated} \end{cases} \quad (4.18)$$

$$R = R_{i_r} + R_{v_{dc}} \quad (4.19)$$

where R_{i_r} , $R_{v_{dc}}$ and R are rewards for rotor current, DC-link voltage and aggregation of both, respectively; α_{1,i_r} , α_{2,i_r} , $\alpha_{1,v_{dc}}$ and $\alpha_{2,v_{dc}}$ are negative penalty coefficients.

4.4 GSES-based Control Algorithm

The simulation of the DFIG system is performed using the PSCAD/EMTDC software environment. The predefined states s_t in (Equation 4.15) are collected by the RL agent implemented as a user-defined component in PSCAD/EMTDC at each time step t . The reward function in (Equation 4.17),(Equation 4.18),(Equation 4.19) is utilized to calculate the reward r_t at each time step. The proposed GSES algorithm uses a neural network with parameters θ to represent the policy π_θ . The output action is determined based on the trained policy π_θ and the values of the states, as defined in (Equation 4.16). After each simulation episode, the episode reward is calculated by accumulating the discounted reward, as defined in (Equation 2.8). The accumulated rewards are used to estimate the gradient of the network parameters and update the policy.

To address the issue of unequal weights on different states due to their varying scales, a state normalization process is performed in the GSES algorithm. The mean and covariance of the corresponding state are used for state normalization, as shown in (Equation 4.20):

$$s_{t,norm} = \frac{1}{\sqrt{diag(\sigma)}}(s_t - \mu) \quad (4.20)$$

Here, σ and μ are the covariance matrix and mean vector of the states s in the previous episode, respectively. The algorithm explores the parameter space of the policy for multiple directions $\epsilon_{j,+}$ and $\epsilon_{j,-}$ for $j \in 1, 2, \dots, N$ in opposite directions. The m directions with the highest average episode reward are selected as the guided subspace direction. The full

search direction and guided subspace are balanced by calculating the distribution of perturbation and updating the policy, following the GSES algorithm described in section 2.3.

The complete GSES algorithm is presented in Algorithm 2. It involves simulating the DFIG system, collecting predefined states, calculating the reward at each time step, training the policy using a neural network, performing state normalization, and exploring the parameter space for multiple directions. The algorithm balances the full search direction and guided subspace and updates the policy by calculating the distribution of perturbation.

4.5 Implementation of Platform and Parallel Computation

To simulate the grid-connected DFIG system in PSCAD-EMTDC software and exchange the simulation data with the GSES algorithm, a solution is proposed in this work. Instead of implementing GSES completely in PSCAD-EMTDC using FORTRAN programming language, which could result in a long development time and difficulty in future maintenance, the GSES algorithm is implemented in Python, a mature and easy-to-use programming language in the field of AI. The overall GSES-based DFIG platform is presented in Figure 4.4. Although a neural network is built in PSCAD-EMTDC through FORTRAN for fast computation speed, PSCAD-EMTDC does not provide an API for exchanging real-time data with Python directly. To overcome this issue, the authors developed their own API, as depicted in Figure 4.4. The simulation data, including states, actions, rewards, and parameters, can be stored in a local database that is accessible to both PSCAD-EMTDC and the Python program.

To exploit the characteristic of the GSES algorithm, which can explore the policy parameter in multiple directions independently and simultaneously, parallel computation techniques are utilized to speed up the training process. Specifically, the distributed AI framework Ray [112] is employed on an 8-core Intel Core i7-10700 CPU to drive 8 PSCAD-

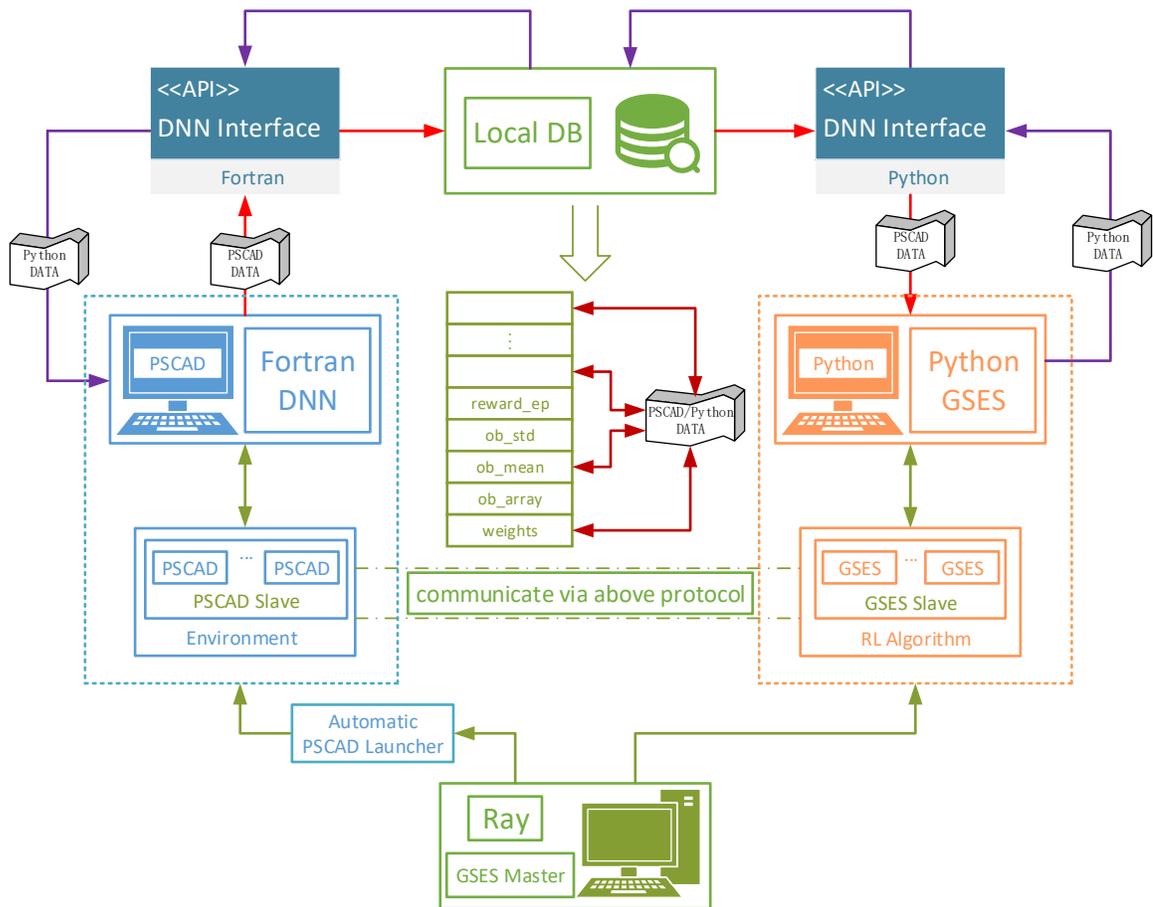


Figure 4.4: Overall GSES-based DFIG control framework: $reward_ep$ is the episode reward; ob_std is the state covariance matrix; ob_mean is the state mean vector; ob_array is the trajectory of the states; $weights$ is the weights of DNN policy [85].

EMTDC instances concurrently. After each iteration, the GSES algorithm coded on the Ray framework collects all the aggregated simulation data and updates the policy parameters. This process repeats until the end of the training. Figure 4.5 shows an example of the accumulated reward of the proposed GSES-based approach during the training process. The GSES algorithm reached convergence within 120 iterations, demonstrating the efficiency of the GSES algorithm under parallel computation.

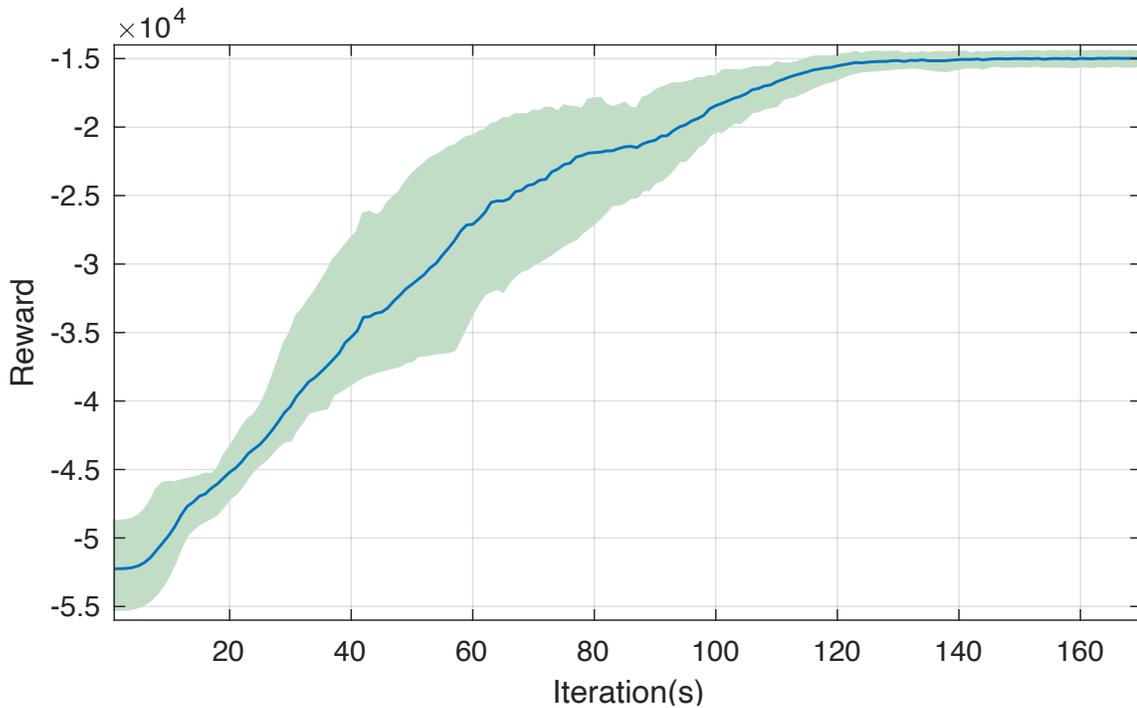


Figure 4.5: An example of rewards during the training process [85]

4.6 Case Study and Discussion

In this study, a PSCAD/EMTDC model of a grid-connected DFIG system with fault configurations is developed, as depicted in Figure 4.6. The DFIG has a rated power of 3.6 MW and a rated voltage of 4.16 kV [113]. The energy is transmitted to the power grid through a 4.16/34.5 kV step-up transformer. The wind speed is set to 11 m/s, and the DFIG

output is 2.88 MW. The proposed GSES algorithm utilizes a policy neural network, which is a three-hidden-layer MLP network with 128, 128, and 62 neurons, respectively. The output layer of the neural network has two neurons, representing the actions of P_e^* and v_{dc}^* . The continuous ranges of action space for both P_e^* and v_{dc}^* are selected from (Equation 4.16) as $P_e^* \in [0, 3.6]MW$ and $v_{dc}^* \in [3.0, 4.0]kV$, respectively. The coefficients of the reward function are chosen as follows: $\alpha_{1,i_r} = -2000$, $\alpha_{2,i_r} = -10000$, $\alpha_{1,v_{dc}} = -500$, and $\alpha_{2,v_{dc}} = -1000$. Here, $|\alpha_{2,i_r}|$ is greater than $|\alpha_{1,i_r}|$ because over-current of the rotor is more critical than over-controlled under-current. Similarly, $|\alpha_{1,i_r}|$ and $|\alpha_{2,i_r}|$ are greater than $|\alpha_{1,v_{dc}}|$ and $|\alpha_{2,v_{dc}}|$ since suppressing over-current of the rotor is the main priority for a DFIG under fault scenarios. It is important to note that the proposed GSES-based approach is not intended to replace the traditional hardware crowbar protection. Instead, it provides an additional software solution. Similar to the fact that the crowbar can work with any other software solution, the crowbar protection can also work along with the proposed method to provide further safety margin. However, in this study, we only focused on the performance of the software solution, and the crowbar was deactivated in all case studies. To evaluate the performance of the trained RL algorithm, three different scenarios were considered.

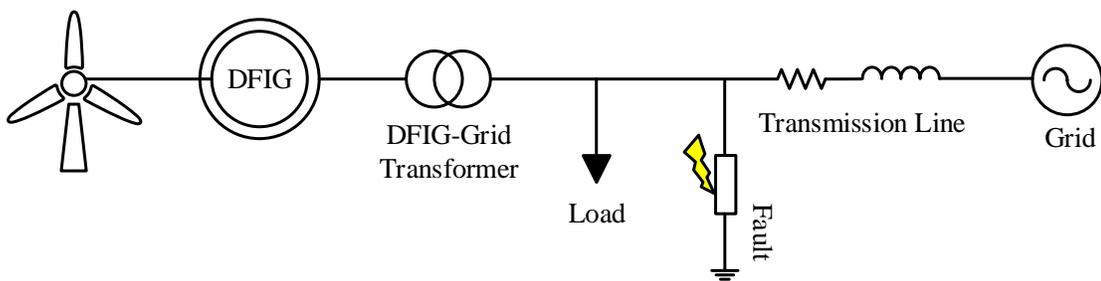


Figure 4.6: Grid-connected DFIG system topology with fault [85]

4.6.1 Three-phase fault with 50% voltage drop

In this case, a three-phase fault with 50% voltage drop occurred at time $t = 10$ s, and it was cleared after 0.2 s. To evaluate the trained policy of GSES, the episode reward R_{i_r} , $R_{v_{dc}}$ of rotor current i_r , DC-link voltage v_{dc} with and without GSES were listed in Table 4.1a. Both increased R_{i_r} and $R_{v_{dc}}$ with GSES algorithm verified that the agent learned the effective policy to maximize the reward successfully.

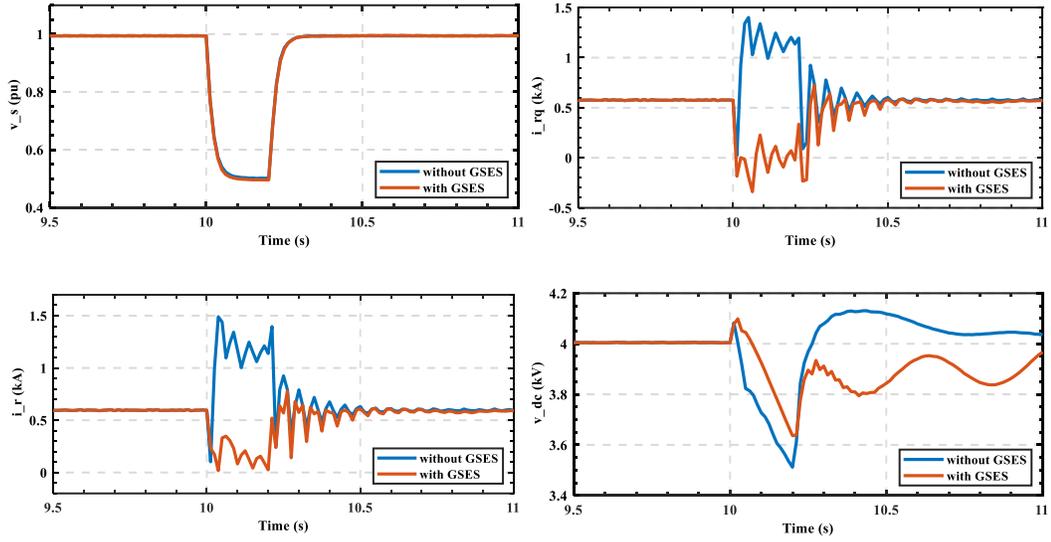


Figure 4.7: Result of 3ϕ fault with 50% voltage drop

Figure 4.7 shows results of the DFIG system states. The DFIG terminal voltage (grid voltage v_g) v_s dropped about 50% during the fault and then recovered when the fault was cleared. The rotor d -axis current i_{rd} did not change too much; however, it can be seen that when the GSES algorithm was activated, the rotor q -axis current i_{rq} decreased dramatically as i_{rq} is directly regulated based on P_e^* (as shown in Figure 4.2). As a result, the aggregated rotor current i_r was obviously reduced by the GSES-based control. Without GSES, the rotor current during fault could reach as high as 1.5 kA. Therefore, it proved that the GSES algorithm is very effective in suppressing over-current at the DFIG rotor during the fault. In addition, DC-link voltage is also reduced below 4.0 kV to prevent from over-voltage during

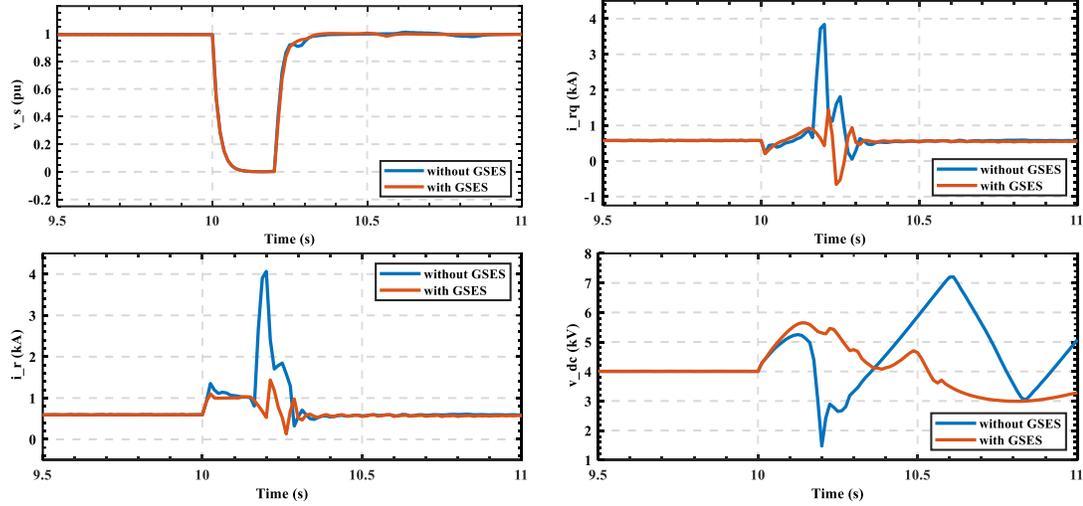


Figure 4.8: Result of 3ϕ fault with 100% voltage drop

the fault. The GSES algorithm is able to increase the rotor-current related reward R_{i_r} from -75,785 to -6,905, increase the DC-link related reward $R_{v_{dc}}$ from -1,595 to -1,199, and the overall reward from -77,380 to -8,104. The increased total reward, reduced rotor current and DC-link voltage have proved that the GSES-based control could effectively improve the DFIG performance during the three-phase fault with 50% voltage drop.

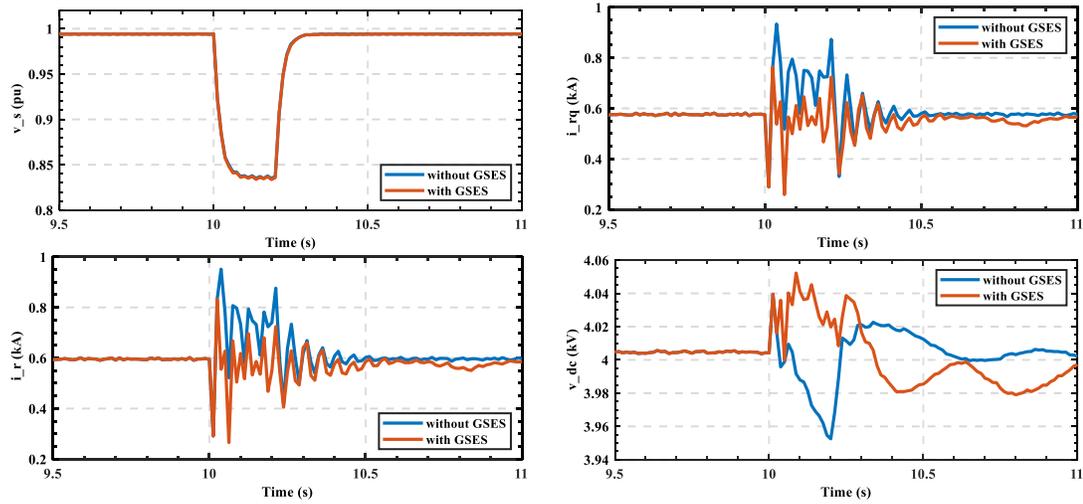


Figure 4.9: Result of single-line-to-ground fault

Table 4.1: Comparison of episode reward between with GSES and without GSES in terms of rotor current and DC-link voltage [85]

	Without GSES	With GSES
R_{i_r}	-75,785	-6,905
$R_{v_{dc}}$	-1,595	-1,199
Total R	-77,380	-8,104

(a) Three-phase fault with 50% voltage drop

	Without GSES	With GSES
R_{i_r}	-491,070	-48,757
$R_{v_{dc}}$	-236,970	-70,780
Total R	-728,040	-119,540

(b) Three-phase fault with 100% voltage drop

	Without GSES	With GSES
R_{i_r}	-7,694.3	-933.9
$R_{v_{dc}}$	-20.4	-40.5
Total R	-7,714.7	-974.4

(c) Single-line-to-ground fault

4.6.2 Three-phase fault with 100% voltage drop

In this case, we tested the performance of proposed GSES-based control during a severer fault scenario. A three-phase fault with 100% voltage drop occurred at time $t = 10$ s, and it was cleared after 0.2 s. The episode reward R_{i_r} , $R_{v_{dc}}$ of rotor current i_r , DC-link voltage v_{dc} with and without GSES during fault were listed in Table 4.1b. It was noticed that because of the severe fault, the absolute values of these rewards with or without GSES-based control were dramatically increased compared to the previous case of three-phase fault with 50% voltage drop. However, when comparing the rewards for situations with and without GSES-based control, we can clearly see that the GSES-based control increased the rewards R_{i_r} and $R_{v_{dc}}$, which proves that the GSES algorithm learned the effective policy to maximize the reward successfully.

Figure 4.8 shows results of the DFIG system states in this three-phase fault with 100% voltage drop. The DFIG terminal voltage (grid voltage v_g) v_s dropped to zero when the fault occurred and it lasted for 0.2 s. The results of this severe fault had higher DFIG rotor current and DC-link voltage. For the situations without the GSES algorithm, the maximum rotor side current was 4 kA, compared to the 1.5 kA in the previous case when the voltage drop was only 50%. Correspondingly, the maximum DC-link voltage was over 7 kV, compared to the 4.2 kV in the previous case. These large over-current and over-voltage resulted in the large negative rewards as shown in Table 4.1b. Compared with the situations with and without GSES algorithm, we can see a significant improvement in the DFIG performance when the GSES was activated. With GSES algorithm, the maximum fault current was reduced from 4 kA to 1.4 kA during the severe three-phase fault, resulting in a change in the rotor current related reward R_{i_r} from -491,070 to -48,757. Similarly, the maximum DC-link voltage was reduced from 7 kV to 5.6 kV when GSES algorithm was implemented, resulting in a change in the DC-link voltage related reward $R_{v_{dc}}$ from

-236,970 to -70,780. The total reward was increased from -728,040 to -119,540. The increased total reward, reduced rotor current and DC-link voltage have proved that the GSES-based control could effectively improve the DFIG performance during the three-phase fault with 100% voltage drop.

4.6.3 Metallic single-line-to-ground fault

A Single-line-to-Ground (SLG) fault was the most common fault in a power system. In this research, we also tested the performance of the proposed GSES algorithm for the SLG fault. A metallic SLG fault (fault resistance was set to 0.001Ω) occurred at time $t = 10 \text{ s}$, and it was cleared after 0.2 s . The episode reward R_{i_r} , $R_{v_{dc}}$ of rotor current i_r , DC-link voltage v_{dc} with and without GSES during fault were listed in Table 4.1c. It was noticed that the absolute values of these rewards with or without GSES-based control were much smaller compared to the previous two three-phase fault cases. However, when comparing the rewards for situations with and without GSES-based control, we can clearly see that the GSES-based control increased the rewards R_{i_r} ($R_{v_{dc}}$ was little changed), which proves that the GSES algorithm learned the effective policy to maximize the reward successfully.

Figure 4.9 shows results of the DFIG system states in this SLG fault. The DFIG terminal voltage v_s dropped about 17% during the fault and then recovered as fault was cleared. The over-current or over-voltage situation was not as severe as the previous two cases. However, we could still see that the GSES algorithm improved the DFIG performance. Compared with the situations without the GSES algorithm, the maximum fault current was reduced from 0.95 kA to 0.85 kA when the GSES algorithm was implemented, resulting in a change in the rotor current related reward R_{i_r} from -7,694.3 to -933.9. The DC-link voltage was little changed as it stayed around 4 kV all the time. The increased total reward, reduced rotor current have proved that the GSES-based control could effectively improve the DFIG performance during the metallic SLG fault.

Algorithm 2 GSES-based DFIG FRT Control [85]

Require: α : learning rate; γ : discount rate; N : number of exploration directions; m : number of exploration directions with highest reward; σ : standard deviation of exploration noise; β : the scale factor; k : the number of surrogate gradients to use.

Ensure: π_{θ^*} : NN-based policy with optimal parameters.

1: initialize $\theta_0 = \mathbf{0} \in \mathbb{R}^{w \times w}$, $\mu_0 = \mathbf{0} \in \mathbb{R}^x$, $\sigma_0 = \mathbf{I} \in \mathbb{R}^{x \times x}$, $i = 0$, where w is the dimension of NN; x is the dimension of states; i is the iteration number.

2: **repeat**

3: Sample N perturbation directions $\epsilon_i, \dots, \epsilon_N$ from $\mathcal{N}(\mathbf{0}, \Sigma)$

4: Assign $2N$ policy parameters:

$$\theta_{i,j,+} = \theta_i + \sigma \epsilon_j$$

$$\theta_{i,j,-} = \theta_i - \sigma \epsilon_j$$

for $j \in \{1, 2, \dots, N\}$.

5: **repeat**

6: Receive states $s_{i,j,t}$ from simulation of PSCAD/EMTDC DFIG system at time step t and calculate the corresponding normalized states $s_{i,j,t,norm}$ as shown in (Equation 4.20).

7: Output actions by the $2N$ policy parameters at time step t :

$$a_{i,j,t} = \pi_{\theta_{i,j,+}}(s_{i,j,t,norm})$$

8: **until** (one episode simulation is finished)

9: Collect $2N$ trajectories of states and the corresponding episode rewards.

10: Obtain the $2m$ policy parameter exploration directions with highest rewards.

11: Update the θ with surrogate gradient:

$$g = \frac{\beta}{2\sigma N} \sum_{i=1}^N \epsilon_i [r(\theta_t + \sigma \epsilon_i) - r(\theta_t - \sigma \epsilon_i)]$$

$$\theta_{t+1} = \theta_t + \alpha g$$

12: Store g to the buffer B and update the surrogate gradient subspace U and perturbation distribution Σ with (Equation 2.17) and (Equation 2.18)

13: $i \leftarrow i + 1$

14: **until** (convergence)

CHAPTER 5

GSES-BASED IEEE-39 BUS FREQUENCY REGULATION SYSTEM

A Doubly-fed Induction Generator (DFIG) is a popular technology for converting wind energy into electrical energy in wind power systems. The DFIG consists of a rotor and a stator, which are connected to the grid through two separate sets of windings. The rotor windings are connected to the grid through converters, whereas the stator windings are directly connected to the main grid through transformers. The configuration of a typical grid-connected DFIG system is illustrated in Figure 5.1. The mechanical shaft system, which includes a gearbox, transfers wind energy from the turbine rotor to the DFIG. The rotor of the DFIG is driven by the turbine rotor and rotates at a speed that is slightly higher than the synchronous speed, while the stator is connected to the grid and rotates at the synchronous speed. The difference between the rotor speed and the synchronous speed is called the slip, and it determines the amount of power that can be transferred between the rotor and the grid (through transformers).

5.1 Mathematical Modeling of DFIG

The mathematical modeling of DFIG including aerodynamics of wind turbine, shaft system and induction generator is same as demonstrated in section 4.1 previously. However, in this typical system, we replace the traditional pitch controller with the RL-based pitch controller to stabilize the system frequency at 60 Hz during different events.

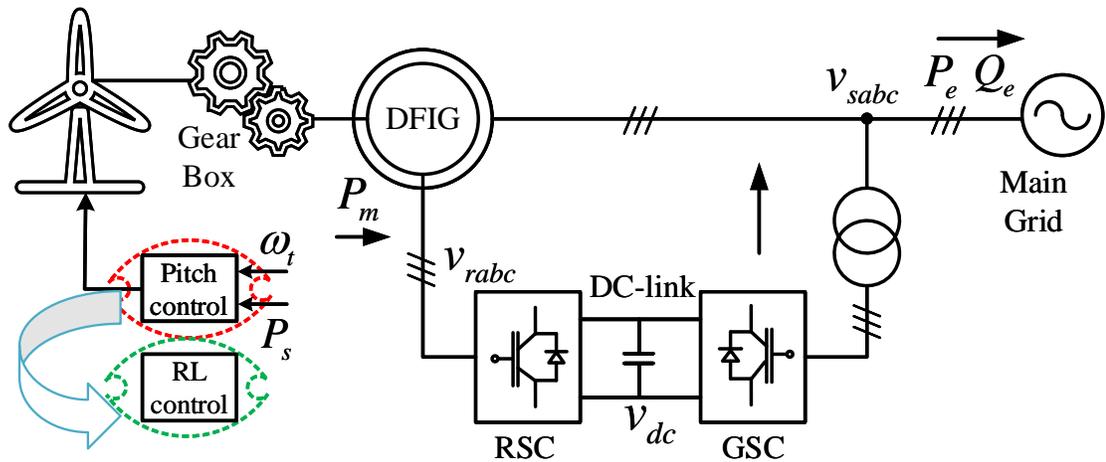


Figure 5.1: DFIG grid connection structure for frequency response [85]

5.2 Controller

The Maximum Power Point Tracking (MPPT) technique is used to optimize the power output of the wind turbine by adjusting the pitch angle and rotor speed. The pitch angle is the angle between the chord line of the wind turbine blade and the plane of rotation, and it determines the aerodynamic force on the blade. The MPPT control system adjusts the pitch angle to maintain the wind turbine at the optimal operating point, which maximizes the power output. The rotor speed is adjusted using the rotor-side converter to maintain the frequency of the generated voltage at the grid frequency.

However, deloaded operation is necessary for frequency response purposes, which requires a positive pitch angle during normal operation. Deloading means reducing the power output of the wind turbine below the rated power to ensure stable grid operation during transient events such as sudden changes in wind speed or grid faults. The positive pitch angle reduces the aerodynamic torque on the wind turbine blade, which reduces the power output. This control strategy is essential for maintaining the stability of the power sys-

tem and preventing voltage and frequency fluctuations that can cause damage to electrical equipment.

5.3 Pitch Angle based Deloading Controller

At MPPT, the coefficient of performance index is maximum, and the power output is at its highest. However, when the pitch angle is set to a positive value, the power output is reduced. In this condition, the DFIG can be deloaded, which allows additional power margin to be developed. The pitch angle can be slowly increased to reach the desired deloading value. At a super synchronous speed, the pitch angle cannot be controlled further, but it can be temporarily reduced to extract extra power for regulating the dip in frequency. This allows for an instantaneous deloading effect, and the output power of the DFIG can cross its rating but can be controlled not to go beyond its thermal limit [114].

5.4 Design of GSES-based Control Strategy

The complexity of the optimization problem for determining the optimal pitch angle value for dynamic frequency response poses a significant challenge due to the presence of several nonlinearities and interdependent effects. To address such situations, an RL-based algorithm can be deemed suitable. Hence, the GSES algorithm, characterized by the following specifics, is utilized:

5.4.1 State Space

In order to apply the GSES algorithm for determining the optimal pitch angle of DFIG to improve frequency response, relevant variables that represent the current status of the DFIG system need to be extracted. Our primary objective is to stabilize the system frequency, therefore, we select system frequency f as a component of the state space. Fur-

thermore, we incorporate the regulated pitch angle signal bt and the mechanical power P_m of the wind turbine, which determines the availability of wind power for frequency response, in the state space. Both electrical power P_e and reactive power Q_e are included as the system frequency is regulated by the provided wind power. The rotor rotational speed ω_r of DFIG indicates the power flow exchange between the stator and rotor of DFIG during operation. Finally, the terminal voltage v_g of DFIG is chosen to represent the system voltage fluctuations at the system connection point. As a result, the state space that provides the most useful information about the DFIG system can be represented as:

$$\mathcal{S} = \{f, bt, P_m, P_e, Q_e, \omega_r, v_g\} \quad (5.1)$$

5.4.2 Action Space

In order to achieve the dynamic regulation of the pitch angle reference value of the DFIG, shown in Figure 5.1, the possible actions can be formulated as:

$$\mathcal{A} = \{bt^*\} \quad (5.2)$$

5.4.3 Reward Function

To incentivize the desired behaviors of the agent and steer it towards sensible exploration of the environment, the reward function design is critical. To achieve the objective of maintaining a stable system frequency of 60 Hz, the reward function must penalize the agent for any undesirable outcomes resulting from its actions, while also rewarding desirable outcomes. In this context, the system frequency (f) is incorporated into the reward function design. Our preference for consistency is reflected in the decision to apply only penalties to the reward function. This approach is captured by the following equation:

$$R = \begin{cases} [\alpha_{1,f}(f - 60)]^3, & f \geq 60Hz \\ [\alpha_{2,f}(60 - f)]^3, & f < 60Hz \end{cases} \quad (5.3)$$

Here, $\alpha_{1,f}$ and $\alpha_{2,f}$ are negative penalty coefficients.

5.5 Implementation Platform and Parallel Computation

In this research, the modified IEEE39 bus system integrated with DFIG is simulated using the PSCAD/EMTDC software platform. To gather predefined states s_t at each time step t , an RL agent is integrated into the PSCAD/EMTDC as a user-defined component. The reward function presented in (Equation 5.3) is employed to determine the reward r_t for each time step t . To implement the proposed GSES algorithm, a neural network with parameters θ is employed to represent the policy π_θ , which generates the output action based on the states and trained policy π_θ , as described in (Equation 5.2). After one simulation episode, the episode reward is determined by calculating the discounted cumulative reward, as defined in (Equation 2.8). This cumulative reward value is then utilized to estimate the gradient of network parameters for updating the policy.

In addition to exploring the policy parameter space in multiple directions, namely $\epsilon_{j,+}$ and $\epsilon_{j,-}$ for $j \in 1, 2, \dots, N$, the GSES algorithm also balances the full search direction and guided subspace by calculating the distribution of perturbation and updating the policy. The m directions with the highest average episode reward are chosen as the guided subspace direction. These directions are opposite to each other, and their corresponding gradients are employed to update the policy parameters.

As the simulation of the system is performed in PSCAD/EMTDC software, effective handling of data communication is required to facilitate the exchange of simulation data, including states, actions, rewards, and parameters between PSCAD/EMTDC and GSES al-

gorithm. We implemented the GSES algorithm in Python, a robust programming language commonly used in AI, while building the neural network in FORTRAN for faster computation. Due to the absence of an API in PSCAD/EMTDC for real-time data exchange with Python, we have created our own API, as shown in Figure 5.3. The simulation data, including states, actions, rewards, and parameters, can be saved in a local database that can be accessed by both PSCAD/EMTDC and the Python program.

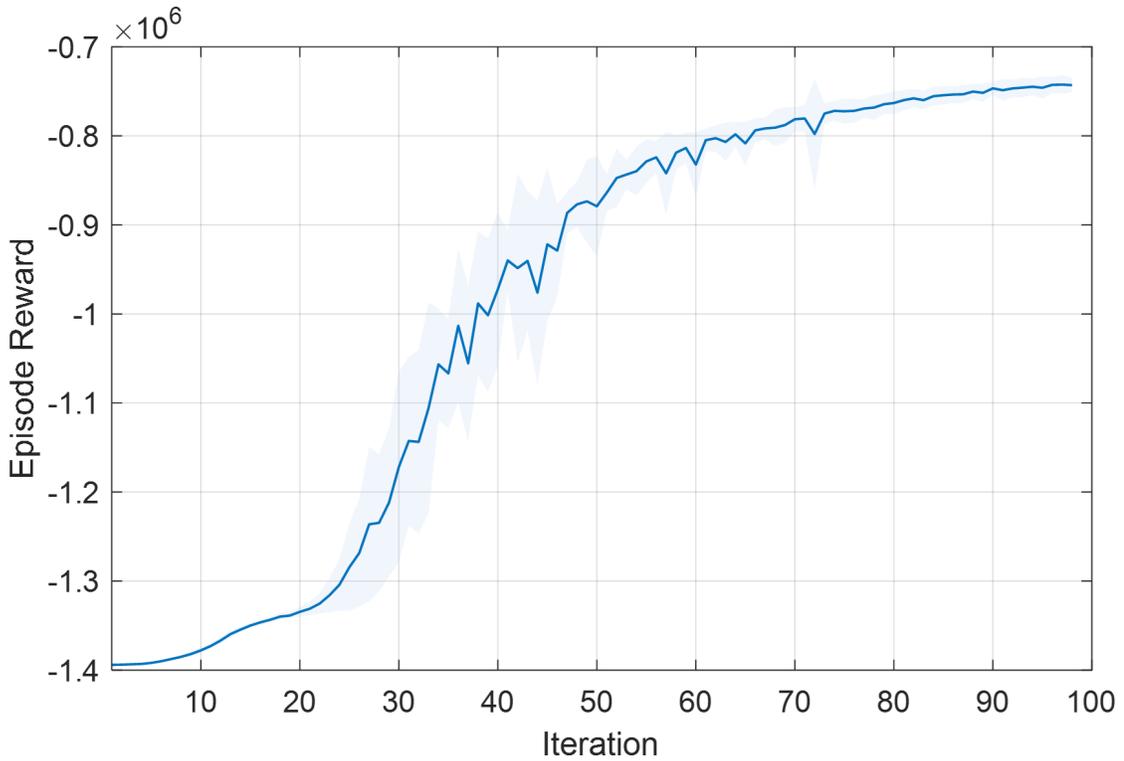


Figure 5.2: History of GSES training rewards

To expedite the training process, we leverage parallel computation techniques, allowing exploration of policy parameters in multiple independent and simultaneous directions. In particular, we employ the distributed AI framework, Ray [112], on an 8-core Intel Core i7-10700 CPU to simultaneously drive 8 PSCAD/EMTDC instances. At the end of each iteration, the Ray framework executes the GSES algorithm, which collects all the aggre-

5.6 Case Study and Discussion

A modified version of the IEEE 39-bus system is utilized in this study to simulate the dynamic frequency response of the power system. This system includes 10 generators with a total capacity of around 6 GW. To compare the effectiveness of our proposed method with conventional frequency response techniques, we substitute the synchronous generator located at Bus 36 with a DFIG wind farm of comparable size. Figure 5.4 illustrates the test system employed in this research. The DFIG wind farm has a rated voltage of 34.5 kV and a rated active power of 550 MW, and it is connected to the power grid through a 34.5/230 kV step-up transformer. The wind speed is set to 11 m/s, and the DFIG output power is regulated to approximately 400 MW in deloaded mode.

A three-hidden-layer multi-layer perceptron network with 128, 128, and 62 neurons for each layer, respectively, has been developed to serve as the policy neural network. The output layer of the neural network includes a neuron for the actions of bt^* . The continuous ranges for bt^* of action space have been chosen from (Equation 5.2), and they fall within the range of $[0, 20]$ degrees. The reward function's coefficients are established as $\alpha_{1,f} = -100$ and $\alpha_{2,f} = -200$ for case studies that the frequency needs to be boosted (e.g. decreased generation and increased load), $\alpha_{1,f} = -200$ and $\alpha_{2,f} = -100$ for case studies that the frequency needs to be suppressed (e.g. increased generation and decreased load). The reason is because that when the frequency is smaller than 60 Hz, we drive the RL agent to boost the frequency to 60 Hz by giving more reward (less penalty) if it is closer than 60 Hz. The GSES RL agent has been trained for various contingencies, intending to maintain the system frequency. After the agent is adequately trained, it is employed to regulate the frequency response of the DFIG. To evaluate the proposed GSES-based method's efficiency, we compared it to the original IEEE 39-Bus system, in which all 9 synchronous generators regulate the system frequency. To ensure stability, we select an occurrence time of $t = 42s$

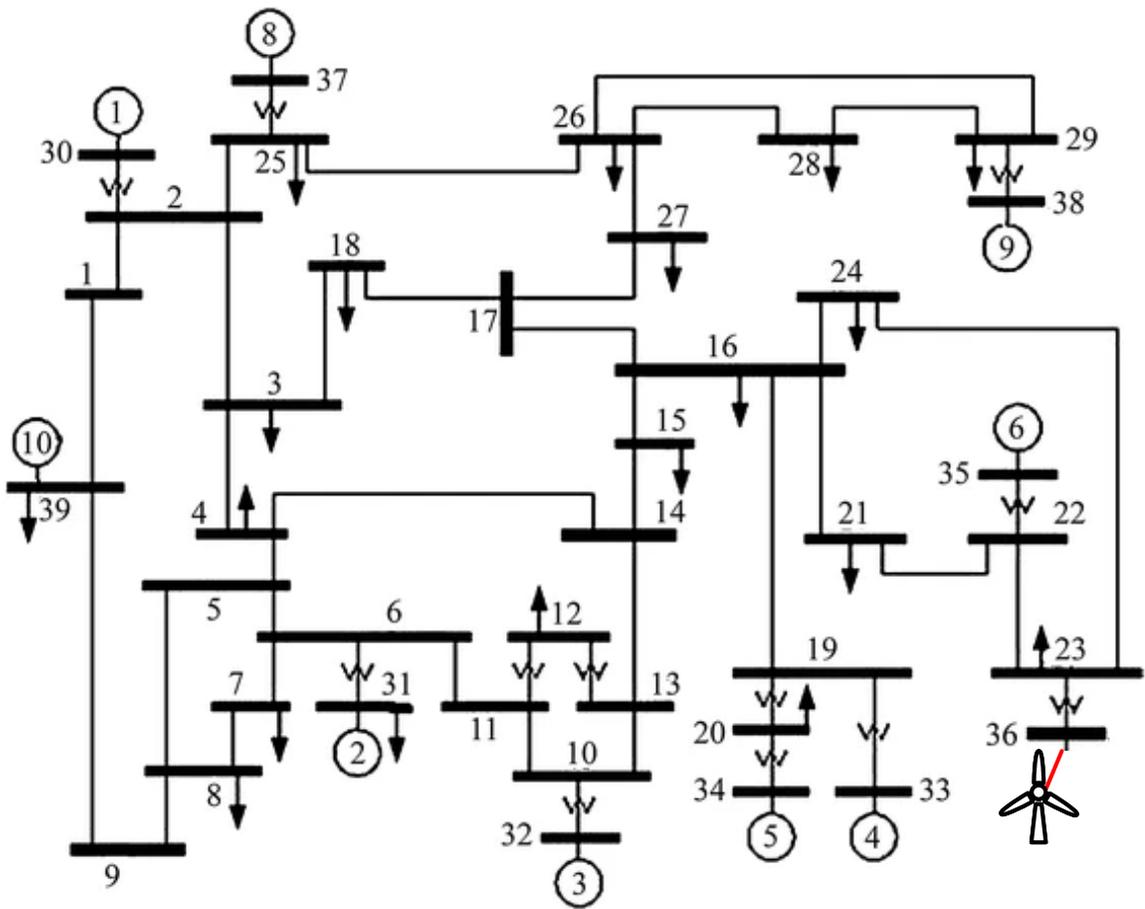


Figure 5.4: Modified IEEE 39-Bus test system for case studies

for all transients in all case studies, as the numerical simulation takes approximately 37s to stabilize.

5.6.1 Loss of Generator

The loss-of-generator emergency studied in this scenario involved an internal fault in the power plant, leading to the sudden cut-off of the generator on Bus 33 and a subsequent loss of approximately 620 MW. This resulted in a significant frequency drop at time $t = 42s$.

Figure 5.5 illustrates the frequency response outcomes for two approaches: the traditional generator-based approach, which does not involve wind farms, and the proposed GSES-based approach. Under the traditional approach, the generators gradually increased power output using conventional droop control methods, causing the system frequency to drop to a minimum of 59.47 Hz before bouncing back. This nadir point is below the 59.5 Hz threshold that typically triggers load-shedding relays to cut off loads, thereby leading to a loss of certain loads. In contrast, the GSES-based DFIG control approach quickly released the reserved power of the wind farm by changing the pitch angle to support the frequency response. The system frequency recovered much faster under this approach, and the lowest value (59.62 Hz) exceeded the 59.5 Hz threshold. Consequently, no additional system loads were cut by the load-shedding relays.

The performance of the frequency response under the two approaches was quantified using the reward function in (Equation 5.3). The calculated total reward over the simulation was $-3,372k$ without the GSES-controlled DFIG, compared to $-716k$ with the proposed GSES approach. The results confirmed the superior performance of the GSES-based frequency response approach over traditional methods during a loss of generator.

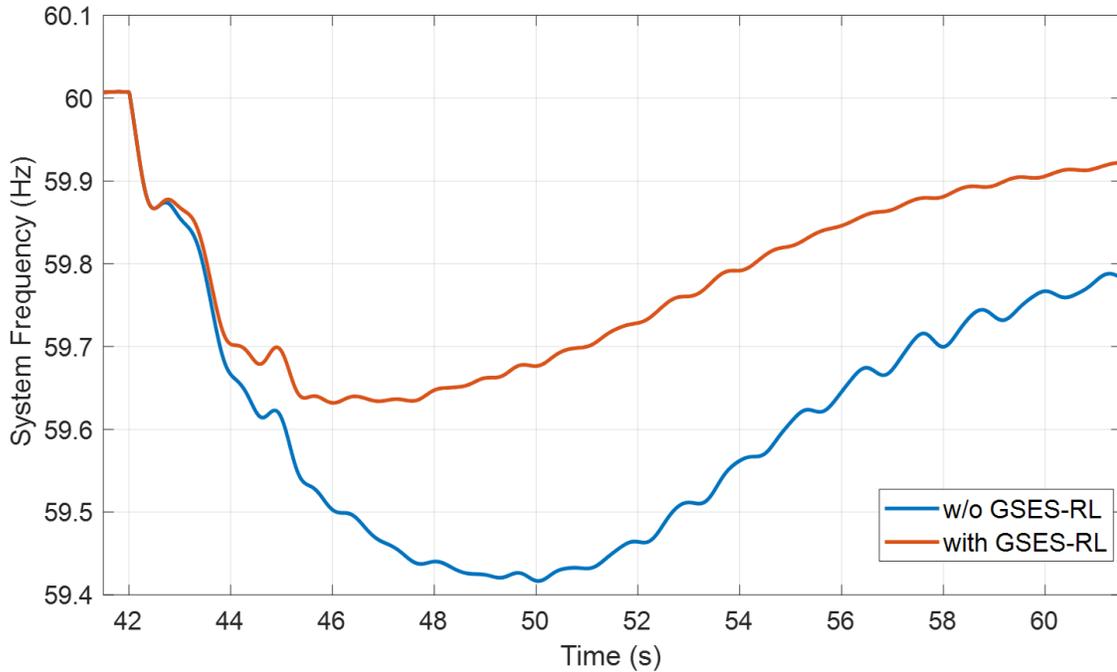


Figure 5.5: Result of frequency response for a loss of generator

5.6.2 Sudden increase of Load

In this study, we investigated the response of the power system to a sudden increase in load on Bus 16, caused by the connection of an unexpected load area. The sudden increase in load necessitated additional generation, which resulted in a significant drop in frequency at time $t = 42s$.

Figure 5.6 shows the frequency response results obtained using the traditional generator-based approach (without wind farms) and the proposed GSES-based approach. When the wind farm is not under GSES control, the generators gradually increase power output using traditional droop control methods, resulting in a system frequency drop to a minimum of 59.42 Hz before rebounding. This frequency dip falls below the critical threshold of 59.5 Hz for about 2.1s, triggering the operation of load-shedding relays and resulting in some load loss.

In contrast, when the GSES-based DFIG control approach is employed, the wind farm's reserved power is quickly released by changing the pitch angle to support the frequency response. The system frequency recovers much faster with the GSES-based control method compared to the traditional method, and the minimum frequency value (59.66 Hz) remains above the 59.5 Hz threshold. As a result, no additional system loads are lost due to load-shedding relays.

We evaluated the performance of the frequency response under the two methods using the reward function defined by (Equation 5.3). The total reward calculated over the simulation period is $-2,748k$ for the traditional approach without GSES control, while it is $-572k$ for the proposed GSES approach. These results validate the superior performance of the GSES-based frequency response approach during a sudden increase in load.

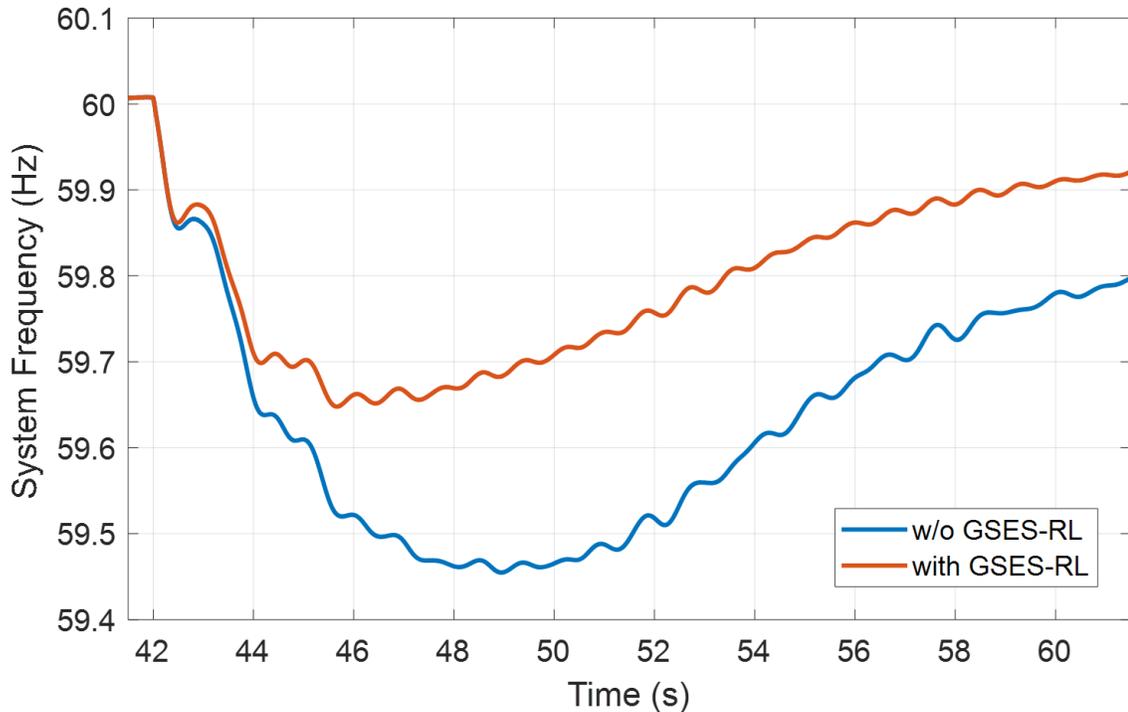


Figure 5.6: Result of frequency response for a sudden increase of load

5.6.3 Decrease of Load

In the last event, generator 4 experienced a decrease in load by approximately 500 MW at time $t = 42$ s. The system frequency response is presented in Figure 5.7, showing an increase in frequency to about 60.4 Hz without the GSES algorithm. However, with the GSES algorithm, the nadir point of the system frequency f was notably reduced by employing the extra power from DFIG, indicating the high effectiveness of the GSES algorithm in regulating the system frequency during such events.

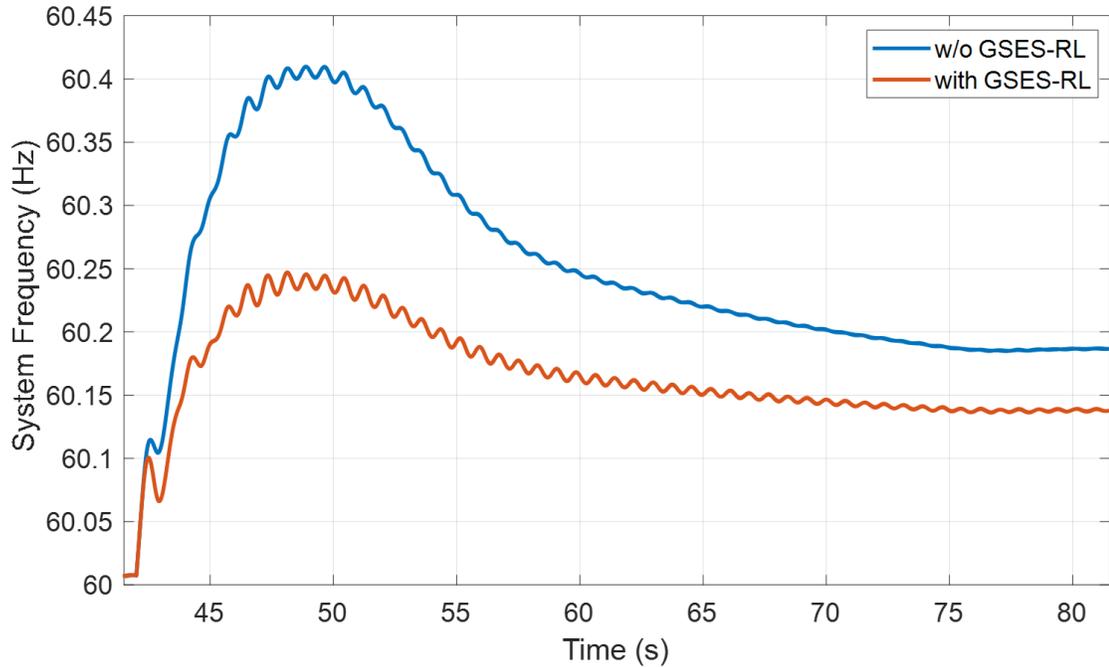


Figure 5.7: Result of frequency response for a sudden decrease of load

Similar to the previous events, we compare the episode reward R_f of the system frequency f with and without the GSES algorithm, as shown in Table 5.1. The results reveal an improvement in R_f with the implementation of the GSES algorithm.

Table 5.1: Reward Comparison with and without GSES under different scenarios

Scenarios	Without GSES	With GSES
Loss of Generator	-3372,639.1	-716,112.0
Increase of Load	-2747,914.4	-572,394.3
Decrease of Load	-2507,166.5	-74,545.2

CHAPTER 6

CONCLUSIONS AND FUTURE WORK

This dissertation proposed a new RL based solution to improve the performance for typical dynamic control problem in power systems including HVDC, DFIG and modified IEEE-39 bus based system during many different scenarios.

6.1 Conclusions

- **GSES-based HVDC Damping Control System:** We proposed a novel approach for damping inter-area oscillations using High Voltage Direct Current (HVDC) transmission. The approach used a Reinforcement Learning (RL)-based control algorithm called Guided Surrogate Gradient-based Evolution Strategy (GSES) to dynamically set the parameters of the controller. The GSES-based controller was efficiently trained using parallel computation techniques to learn to modulate the HVDC transferred power during various transient event-triggered inter-area oscillations. The effectiveness and superiority of the proposed GSES-based HVDC oscillation damping controller over conventional controllers were demonstrated in extensive studies.
- **GSES-based DFIG-FRT System:** The performance of the Doubly-fed Induction Generator (DFIG) system was improved using an advanced GSES algorithm. The algorithm was designed and implemented in PSCAD software to control the DFIG sys-

tem. A communication framework was established between PSCAD and Python to facilitate machine learning. The use of parallel computation improved the computational and training speed. The proposed GSES-based control was tested in a detailed DFIG system under different fault scenarios. Simulation results showed that the algorithm effectively reduced the rotor over-current and DC-link over-voltage.

- **GSES-based IEEE-39 Bus Frequency Regulation System:** We presented a novel approach for frequency response that utilizes the reserved power of deloaded wind farms to support the system frequency under emergencies. Our approach employed a robust GSES-based RL agent to intelligently control the wind farm output and learn the best policy for frequency response. A user interface was designed, and parallel computation techniques were used to efficiently train the GSES-based RL agent. To validate the efficacy of our proposed approach, we conducted experiments on a modified IEEE 39-Bus system and compared its performance with traditional frequency response methods under sudden load increase and generator loss scenarios. The proposed approach demonstrated superior performance over traditional frequency response methods.

Overall, the proposed RL-based solutions demonstrated promising results in improving the performance of power systems. The use of parallel computation techniques and the communication framework between different simulation environments and Python provided more efficient and practical solutions for real-world applications.

6.2 Future Work

Overall, future work on this research is expected to explore more complex systems and challenging scenarios and to investigate the scalability and robustness of the proposed

approaches. In addition, we plan to have the detailed future work for each specific system in the next subsections.

- GSES-based HVDC Damping Control System: Future work on this subject will focus on studying the impact of communication delays on the proposed approach and exploring the use of meta-learning to adapt to various transient events.
- GSES-based DFIG-FRT System: It is worth noting that the proposed GSES-based control can work in combination with traditional hardware solutions such as the crowbar protection. For this subject, further investigation is needed to explore the combined performance.
- GSES-based IEEE-39 Bus Frequency Regulation System: Future work on this subject is expected to introduce more wind turbines for frequency regulation support. Thus those wind turbines can not only enhance the capability of frequency regulation, but also can operate coordinately to increase the stability of the system.

The use of RL in power systems is an exciting and rapidly evolving field, and we look forward to seeing more innovative solutions in the future.

PUBLICATIONS

1. **Gao, W.** Fan, R., Qiao, W., Wang, S., & Gao, W. (2023). Fast Frequency Response using Reinforcement Learning-Controlled Wind Turbines. IEEE IAS Annual Meeting, submitted.
2. **Gao, W.**, Fan, R., Huang, R., Huang, Q., Gao, W., & Du, L. (2023). Augmented random search based inter-area oscillation damping using high voltage DC transmission. Electric Power Systems Research, 216, 109063.
3. **Gao, W.**, Fan, R., Huang, R., Huang, Q., Du, Y., Qiao, W., ... & Gao, D. W. (2022). Improving DFIG performance under fault scenarios through evolutionary reinforcement learning based control. IET Generation, Transmission & Distribution.
4. **Gao, W.** (2021, April). PV Array Fault Detection Based on Deep Neural Network. In 2021 IEEE Green Technologies Conference (GreenTech) (pp. 42-47). IEEE.
5. **Gao, W.** (2020, July). Microgrid control strategy based on battery energy storage system-virtual synchronous generator (BESS-VSG). In 2020 IEEE Kansas Power and Energy Conference (KPEC) (pp. 1-6). IEEE.
6. Huang, R., **Gao, W.**, Fan, R., & Huang, Q. (2022). Damping inter-area oscillation using reinforcement learning controlled TCSC. IET Generation, Transmission & Distribution.

7. Huang, R., **Gao, W.**, Fan, R., & Huang, Q. (2022). A Guided Evolutionary Strategy Based Static Var Compensator Control Approach for Inter-area Oscillation Damping. *IEEE Transactions on Industrial Informatics*.
8. Li, Y., **Gao, W.**, Huang, S., Wang, R., Yan, W., Gevorgian, V., & Gao, D. W. (2021). Data-driven optimal control strategy for virtual synchronous generator via deep reinforcement learning approach. *Journal of Modern Power Systems and Clean Energy*, 9(4), 919-929.

APPENDIX A

COMPARISON BETWEEN BRS AND GSES

A.1 Basic Random Search

The GSES algorithm is an extension of the Basic Random Search (BRS) algorithm. In the BRS, a neural network is also used to represent the mapping between the state s to the action a . Therefore, the RL policy π_θ is based on the neural network mapping, and $\theta \in \mathbb{R}^n$ is the parameter of the neural network [115]. The objective of BRS is to find the optimal policy π_θ^* that maximizes the expected total reward through updating the neural network parameter θ :

$$\text{Objective : } \max_{\theta} \mathbb{E} [R(\pi_\theta)] \tag{A.1}$$

where $R(\pi_\theta)$ is the total reward using the policy π_θ .

The BRS initializes the policy π_θ and perturbs the parameter θ by applying $+\epsilon\delta$ and $-\epsilon\delta$. The ϵ is a positive constant exploration noise that is small than 1, and δ is the random sample generated by Gaussian distribution. The perturbed policy pair $\pi_{\theta+\epsilon\delta}$ and $\pi_{\theta-\epsilon\delta}$ will be evaluated by their corresponding received total rewards $R(\pi_{\theta+\epsilon\delta})$ and $R(\pi_{\theta-\epsilon\delta})$. When the parallel computation technique is used, N perturbed policy pairs can be generated at the same time to calculate the $2N$ received total rewards. These rewards are used to determine the gradient $\Delta\theta$ of the parameter θ as follows (Equation A.2):

$$\Delta\theta = \frac{1}{N} \sum_{k=1}^N [R(\pi_{\theta+\epsilon\delta_k}) - R(\pi_{\theta-\epsilon\delta_k})] \delta_k \quad (\text{A.2})$$

The neural network parameter θ is then updated by

$$\theta \leftarrow \theta + \alpha\Delta\theta \quad (\text{A.3})$$

Unlike many other RL algorithms that require a back-propagation process for gradient update, the BRS determines the $\Delta\theta$ using a simultaneous perturbation stochastic approximation approach, and thus it achieves a much faster and more robust learning convergence. However, the BRS faces several challenges like the exploding standard deviation in the collected rewards, unbalanced system state magnitudes, and un-prioritized gradient update directions. The GSES algorithm keeps the benefits of the BRS and solves the above challenges.

A.2 Guided Surrogate Gradient-based Evolution Strategy

During the training of BRS, the standard deviation σ_N of the N pairs of total rewards $R(\pi_{\theta+\epsilon\delta})$ and $R(\pi_{\theta-\epsilon\delta})$ would gradually grow from a small value to a very large number. Therefore, in the gradient updating stage, it is very difficult to choose a suitable learning rate α that can properly handle the increasing σ_N when the training process is approaching the optimal points. The GSES addresses this problem through dividing the gradient $\Delta\theta$ by the magnitude of the standard deviation σ_N [91],

$$\Delta\theta = \frac{1}{N\sigma_N} \sum_{k=1}^N [R(\pi_{\theta+\epsilon\delta_k}) - R(\pi_{\theta-\epsilon\delta_k})] \delta_k \quad (\text{A.4})$$

The above method can effectively reduce the impact of the varying total reward. Another issue with BRS is that the system states could be in very different scales, which

could contribute to inequality of the inputs of the neural network. Taking the power system as an example, the system states are voltage magnitudes, phase angles, bus frequencies, and machine speeds. These states could be in different units and scales. For instance, the voltage magnitude could be around 1.0 per unit, while the frequency difference at the HVDC terminals can be around 0.01 Hz. The differences in the input scales would result in over-weighting the voltage magnitude than the frequency difference. The GSES solves this problem through implementing a state normalization to fit the neural network as

$$\pi'_{\theta+\epsilon\delta_k}(s) = \pi_{\theta+\epsilon\delta_k}(\text{diag}(\Sigma)^{-1/2}(s - \mu)) \quad (\text{A.5})$$

$$\pi'_{\theta-\epsilon\delta_k}(s) = \pi_{\theta-\epsilon\delta_k}(\text{diag}(\Sigma)^{-1/2}(s - \mu)) \quad (\text{A.6})$$

where Σ and μ are the variance and mean value of the input states s .

The last problem with BRS is that the gradient $\Delta\theta$ is estimated by using N pairs of total rewards of the perturbed policies $\pi_{\theta+\epsilon\delta}$ and $\pi_{\theta-\epsilon\delta}$. Because the perturbed direction $+\epsilon\delta$ and $-\epsilon\delta$ are randomly selected, there is a very high chance that some direction could decrease (instead of enhancing) the performance of the RL policy π_{θ} . The GSES evaluates all the rewards of each perturbed policy and ranks the rewards from high to low. Only the top M pairs are kept by the GSES to determine the gradient updating direction, while the other $N - M$ pairs are discarded.

In summary, the GSES enhances the performance of the BRS by replacing the previous gradient updating process in (Equation A.7) as:

$$\Delta\theta = \frac{1}{M\sigma_M} \sum_{k=1}^M [R(\pi'_{\theta+\epsilon\delta_k}) - R(\pi'_{\theta-\epsilon\delta_k})] \delta_k \quad (\text{A.7})$$

REFERENCES

- [1] Y. Zhang, X. Shi, H. Zhang, Y. Cao, and V. Terzija, "Review on deep learning applications in frequency analysis and control of modern power system," *International Journal of Electrical Power & Energy Systems*, vol. 136, p. 107 744, 2022.
- [2] R. He, H. Xie, J. Deng, T. Feng, L. L. Lai, and M. Shahidehpour, "Reliability modeling and assessment of cyber space in cyber-physical power systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 5, pp. 3763–3773, 2020.
- [3] U. Adhikari, T. Morris, and S. Pan, "Wams cyber-physical test bed for power system, cybersecurity study, and data mining," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2744–2753, 2016.
- [4] Z. Lin *et al.*, "Wams-based coherency detection for situational awareness in power systems with renewables," *IEEE Transactions on Power Systems*, vol. 33, no. 5, pp. 5410–5426, 2018.
- [5] Y. Zhang *et al.*, "Distributed electrical energy systems: Needs, concepts, approaches and vision," *Acta Automatica Sinica*, vol. 43, no. 9, Sep. 2017. [Online]. Available: <https://www.osti.gov/biblio/1413187>.
- [6] J. J. Zhang *et al.*, "Social energy: Mining energy from the society," *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 3, pp. 466–482, 2017.
- [7] J. Hao, K. Zhang, X. Dai, and J. J. Zhang, "Game-theoretic energy consumption planning at distribution level for future smart grid," in *2019 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI)*, 2019, pp. 203–206.
- [8] J. Hao, "Deep reinforcement learning for the optimization of building energy control and management," Ph.D. dissertation, University of Denver, 2020.
- [9] J. Hao, D. W. Gao, and J. J. Zhang, "Reinforcement learning for building energy optimization through controlling of central hvac system," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 320–328, 2020.
- [10] J. Hao, X. Dai, Y. Zhang, J. Zhang, and W. Gao, "Distribution locational real-time pricing based smart building control and management," in *2016 North American Power Symposium (NAPS)*, 2016, pp. 1–6.
- [11] J. Hao, Y. Gu, Y. Zhang, J. J. Zhang, and D. W. Gao, "Locational marginal pricing in the campus power system at the power distribution level," in *2016 IEEE Power and Energy Society General Meeting (PESGM)*, 2016, pp. 1–5.
- [12] V. Telukunta, J. Pradhan, A. Agrawal, M. Singh, and S. G. Srivani, "Protection challenges under bulk penetration of renewable energy resources in power systems:

- A review,” *CSEE journal of power and energy systems*, vol. 3, no. 4, pp. 365–379, 2017.
- [13] H. Karbouj, Z. H. Rather, D. Flynn, and H. W. Qazi, “Non-synchronous fast frequency reserves in renewable energy integrated power systems: A critical review,” *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 488–501, 2019.
- [14] M. Mahzarnia, M. P. Moghaddam, P. T. Baboli, and P. Siano, “A review of the measures to enhance power systems resilience,” *IEEE Systems Journal*, vol. 14, no. 3, pp. 4059–4070, 2020.
- [15] S. Liu *et al.*, “A data-driven approach for online dynamic security assessment with spatial-temporal dynamic visualization using random bits forest,” *International Journal of Electrical Power & Energy Systems*, vol. 124, p. 106316, 2021.
- [16] H. Bevrani, H. Golpira, A. R. Messina, N. Hatziargyriou, F. Milano, and T. Ise, “Power system frequency control: An updated review of current solutions and new challenges,” *Electric Power Systems Research*, vol. 194, p. 107114, 2021.
- [17] D. Fernández-Muñoz, J. I. Pérez-Diéz, I. Guisández, M. Chazarra, and Á. Fernández-Espina, “Fast frequency control ancillary services: An international review,” *Renewable and Sustainable Energy Reviews*, vol. 120, p. 109662, 2020.
- [18] A. Fernández-Guillamón, E. Gómez-Lázaro, E. Muljadi, and Á. Molina-Garciéa, “Power systems with high renewable energy sources: A review of inertia and frequency control strategies over time,” *Renewable and Sustainable Energy Reviews*, vol. 115, p. 109369, 2019.
- [19] D. J. Sobajic and Y.-H. Pao, “Artificial neural-net based dynamic security assessment for electric power systems,” *IEEE Transactions on Power Systems*, vol. 4, no. 1, pp. 220–228, 1989.
- [20] Y. Asadi, M. M. Farsangi, E. Bijami, A. M. Amani, and K. Y. Lee, “Data-driven adaptive control of wide-area non-linear systems with input and output saturation: A power system application,” *International Journal of Electrical Power & Energy Systems*, vol. 133, p. 107225, 2021.
- [21] E. Todorov, T. Erez, and Y. Tassa, “Mujoco: A physics engine for model-based control,” in *2012 IEEE/RSJ international conference on intelligent robots and systems*, IEEE, 2012, pp. 5026–5033.
- [22] V. Mnih *et al.*, “Asynchronous methods for deep reinforcement learning,” in *International conference on machine learning*, PMLR, 2016, pp. 1928–1937.
- [23] J. Schulman, P. Moritz, S. Levine, M. Jordan, and P. Abbeel, “High-dimensional continuous control using generalized advantage estimation,” *arXiv preprint arXiv:1506.02438*, 2015.

- [24] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International conference on machine learning*, PMLR, 2015, pp. 1889–1897.
- [25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [26] V. Mnih *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [27] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.
- [28] A. Messina *et al.*, "Use of power system stabilizers for damping inter-area oscillations in the south systems of the mexican electrical grid," *Electric power systems research*, vol. 76, no. 4, pp. 169–179, 2006.
- [29] K. Sebaa and M. Boudour, "Optimal locations and tuning of robust power system stabilizer using genetic algorithms," *Electric Power Systems Research*, vol. 79, no. 2, pp. 406–416, 2009.
- [30] N. Mithulananthan, C. A. Canizares, J. Reeve, and G. J. Rogers, "Comparison of pss, svc, and statcom controllers for damping power system oscillations," *IEEE transactions on power systems*, vol. 18, no. 2, pp. 786–792, 2003.
- [31] Y. Zhang and A. Bose, "Design of wide-area damping controllers for interarea oscillations," *IEEE transactions on power systems*, vol. 23, no. 3, pp. 1136–1143, 2008.
- [32] C. Gama, "Brazilian north-south interconnection control-application and operating experience with a tcsc," in *1999 IEEE Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No. 99CH36364)*, IEEE, vol. 2, 1999, pp. 1103–1108.
- [33] X. Li, "Nonlinear controller design of thyristor controlled series compensation for damping inter-area power oscillation," *Electric Power Systems Research*, vol. 76, no. 12, pp. 1040–1046, 2006.
- [34] M. Weiss, B. N. Abu-Jaradeh, A. Chakraborty, A. Jamehbozorg, F. Habibi-Ashrafi, and A. Salazar, "A wide-area svc controller design for inter-area oscillation damping in wecc based on a structured dynamic equivalent model," *Electric Power Systems Research*, vol. 133, pp. 1–11, 2016.
- [35] E. Al Murawwi, M. Mohsin, L. El Chaar, and L. Lamont, "Interconnection between two variant frequency systems using hvdc links," in *2011 2nd International Conference on Electric Power and Energy Conversion Systems (EPECS)*, IEEE, 2011, pp. 1–6.

- [36] A. Alassi, S. Bañales, O. Ellabban, G. Adam, and C. MacIver, “Hvdc transmission: Technology review, market trends and future outlook,” *Renewable and Sustainable Energy Reviews*, vol. 112, pp. 530–554, 2019.
- [37] M. A. Elizondo *et al.*, “Interarea oscillation damping control using high-voltage dc transmission: A survey,” *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6915–6923, 2018.
- [38] R. Fan, R. Huang, and L. Sun, “Nonlinear model predictive control of hvdc for inter-area oscillation damping,” *Electric Power Systems Research*, vol. 165, pp. 27–34, 2018.
- [39] V. C. Ganti, B. Singh, S. K. Aggarwal, and T. C. Kandpal, “Dfig-based wind power conversion with grid power leveling for reduced gusts,” *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, 2012.
- [40] I. D. Margaritis, S. A. Papathanassiou, N. D. Hatziargyriou, A. D. Hansen, and P. Sorensen, “Frequency control in autonomous power systems with high wind power penetration,” *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 189–199, 2012.
- [41] M. Yamamoto and O. Motoyoshi, “Active and reactive power control for doubly-fed wound rotor induction generator,” *IEEE Transactions on Power Electronics*, vol. 6, no. 4, pp. 624–629, 1991.
- [42] M. Edrah, K. L. Lo, and O. Anaya-Lara, “Impacts of high penetration of dfig wind turbines on rotor angle stability of power systems,” *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 759–766, 2015.
- [43] R. Cardenas, R. Peña, S. Alepuz, and G. Asher, “Overview of control systems for the operation of dfigs in wind energy applications,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 7, 2013.
- [44] M. Mohseni, S. Islam, and M. Masoum, “Fault ride-through capability enhancement of doubly-fed induction wind generators,” *IET Renewable Power Generation*, vol. 5, no. 5, pp. 368–376, 2011.
- [45] K. E. Okedu, S. M. Muyeen, R. Takahashi, and J. Tamura, “Wind farms fault ride through using dfig with new protection scheme,” *IEEE Transactions on Sustainable Energy*, vol. 3, no. 2, pp. 242–254, 2012.
- [46] G. Joos, “Wind turbine generator low voltage ride through requirements and solutions,” in *2008 IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century*, IEEE, 2008, pp. 1–7.
- [47] Y.-W. Shen, D.-P. Ke, W. Qiao, Y.-Z. Sun, D. S. Kirschen, and C. Wei, “Transient reconfiguration and coordinated control for power converters to enhance the lVRT of

- a dfig wind turbine with an energy storage device,” *IEEE Transactions on energy conversion*, vol. 30, no. 4, 2015.
- [48] S. Huang, Q. Wu, Y. Guo, and F. Rong, “Hierarchical active power control of dfig-based wind farm with distributed energy storage systems based on adm,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1528–1538, 2020.
- [49] J. López, E. Gubiéa, E. Olea, J. Ruiz, and L. Marroyo, “Ride through of wind turbines with doubly fed induction generator under symmetrical voltage dips,” *IEEE Transactions on Industrial Electronics*, vol. 56, no. 10, pp. 4246–4254, 2009.
- [50] M. V. Gururaj and N. P. Padhy, “Phil experimentation on fault ride through behavior of doubly fed induction generator-based wind system in the presence of fault current limiter,” *IEEE Transactions on Industry Applications*, vol. 56, no. 3, pp. 2988–3005, 2020.
- [51] M. M. Hossain and M. H. Ali, “Transient stability improvement of doubly fed induction generator based variable speed wind generator using dc resistive fault current limiter,” *IET Renewable Power Generation*, vol. 10, no. 2, pp. 150–157, 2016.
- [52] D. Xiang, L. Ran, P. J. Tavner, and S. Yang, “Control of a doubly fed induction generator in a wind turbine during grid fault ride-through,” *IEEE transactions on energy conversion*, vol. 21, no. 3, 2006.
- [53] R. Zhu, Z. Chen, X. Wu, and F. Deng, “Virtual damping flux-based lvr control for dfig-based wind turbine,” *IEEE Transactions on Energy Conversion*, vol. 30, no. 2, pp. 714–725, 2015.
- [54] J. Liang, W. Qiao, and R. G. Harley, “Feed-forward transient current control for low-voltage ride-through enhancement of dfig wind turbines,” *IEEE Transactions on Energy Conversion*, vol. 25, no. 3, pp. 836–843, 2010.
- [55] J. Driesen and K. Visscher, “Virtual synchronous generators,” in *2008 IEEE Power and Energy Society General Meeting-Conversion and Delivery of Electrical Energy in the 21st Century*, IEEE, 2008.
- [56] S. Wang and L. Shang, “Fault ride through strategy of virtual synchronous controlled dfig-based wind turbines under symmetrical grid faults,” *IEEE Transactions on Energy Conversion*, vol. 35, no. 3, pp. 1360–1371, 2020.
- [57] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.
- [58] N. Maheswaranathan, L. Metz, G. Tucker, D. Choi, and J. Sohl-Dickstein, “Guided evolutionary strategies: Augmenting random search with surrogate gradients,” in *International Conference on Machine Learning*, 2019, pp. 4264–4273.

- [59] J. Oh *et al.*, “Discovering reinforcement learning algorithms,” *arXiv preprint arXiv:2007.08794*, 2020.
- [60] H. Bevrani, A. Ghosh, and G. Ledwich, “Renewable energy sources and frequency regulation: Survey and new perspectives,” *IET Renewable Power Generation*, vol. 4, no. 5, pp. 438–457, 2010.
- [61] L. Yang, G. Y. Yang, Z. Xu, Z. Dong, K. P. Wong, and X. Ma, “Optimal controller design of a doubly-fed induction generator wind turbine system for small signal stability enhancement,” *IET generation, transmission & distribution*, vol. 4, no. 5, pp. 579–597, 2010.
- [62] G. P. Prajapat, N. Senroy, and I. Narayan Kar, “Stability enhancement of dfig-based wind turbine system through linear quadratic regulator,” *IET Generation, Transmission & Distribution*, vol. 12, no. 6, pp. 1331–1338, 2018.
- [63] N. Nguyen and J. Mitra, “An analysis of the effects and dependency of wind power penetration on system frequency regulation,” *IEEE Transactions on Sustainable Energy*, vol. 7, no. 1, pp. 354–363, 2015.
- [64] K. Vidyanandan and N. Senroy, “Primary frequency regulation by deloaded wind turbines using variable droop,” *IEEE transactions on Power Systems*, vol. 28, no. 2, pp. 837–846, 2012.
- [65] G. Ramtharan, N. Jenkins, and J. Ekanayake, “Frequency support from doubly fed induction generator wind turbines,” *IET Renewable Power Generation*, vol. 1, no. 1, pp. 3–9, 2007.
- [66] R. G. De Almeida and J. P. Lopes, “Participation of doubly fed induction wind generators in system frequency regulation,” *IEEE transactions on power systems*, vol. 22, no. 3, pp. 944–950, 2007.
- [67] A. D. Hansen, P. Sørensen, F. Iov, and F. Blaabjerg, “Centralised power control of wind farm with doubly fed induction generators,” *Renewable energy*, vol. 31, no. 7, pp. 935–951, 2006.
- [68] M. Kayikçi and J. V. Milanovic, “Dynamic contribution of dfig-based wind plants to system frequency disturbances,” *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 859–867, 2009.
- [69] M. El Mokadem, V. Courtecuisse, C. Saudemont, B. Robyns, and J. Deuse, “Experimental study of variable speed wind generator contribution to primary frequency control,” *Renewable Energy*, vol. 34, no. 3, pp. 833–844, 2009.
- [70] L.-R. Chang-Chien, W.-T. Lin, and Y.-C. Yin, “Enhancing frequency response control by dfigs in the high wind penetrated power systems,” *IEEE transactions on power systems*, vol. 26, no. 2, pp. 710–718, 2010.

- [71] N. A. Janssens, G. Lambin, and N. Bragard, "Active power control strategies of dfig wind turbines," in *2007 IEEE Lausanne Power Tech*, IEEE, 2007, pp. 516–521.
- [72] I. Erlich and M. Wilch, "Primary frequency control by wind turbines," in *IEEE PES general meeting*, IEEE, 2010, pp. 1–8.
- [73] U. Akram, M. Nadarajah, R. Shah, and F. Milano, "A review on rapid responsive energy storage technologies for frequency regulation in modern power systems," *Renewable and Sustainable Energy Reviews*, vol. 120, p. 109 626, 2020.
- [74] S. Baros and M. D. Ilić, "Distributed torque control of deloaded wind dfigs for wind farm power output regulation," *IEEE Transactions on Power Systems*, vol. 32, no. 6, pp. 4590–4599, 2017.
- [75] L. Zhou, A. Swain, and A. Ukil, "Reinforcement learning controllers for enhancement of low voltage ride through capability in hybrid power systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5023–5031, 2019.
- [76] 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, 2019.
- [77] D. Cao *et al.*, "Reinforcement learning and its applications in modern power and energy systems: A review," *Journal of modern power systems and clean energy*, vol. 8, no. 6, pp. 1029–1042, 2020.
- [78] W. Gao *et al.*, "Improving dfig performance under fault scenarios through evolutionary reinforcement learning based control," *IET Generation, Transmission & Distribution*, vol. 16, no. 19, pp. 3825–3836, 2022.
- [79] W. Gao, "Pv array fault detection based on deep neural network," in *2021 IEEE Green Technologies Conference (GreenTech)*, IEEE, 2021, pp. 42–47.
- [80] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [81] A. K. Ozcanli, F. Yaprakdal, and M. Baysal, "Deep learning methods and applications for electrical power systems: A comprehensive review," *International Journal of Energy Research*, vol. 44, no. 9, pp. 7136–7157, 2020.
- [82] M. Khodayar, G. Liu, J. Wang, and M. E. Khodayar, "Deep learning in power systems research: A review," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 2, pp. 209–220, 2020.
- [83] M. L. Puterman, "Markov decision processes," *Handbooks in operations research and management science*, vol. 2, pp. 331–434, 1990.
- [84] D. Silver *et al.*, "Mastering the game of go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484–489, 2016.

- [85] W. Gao *et al.*, “Improving dfig performance under fault scenarios through evolutionary reinforcement learning based control,” *IET Generation, Transmission & Distribution*, vol. 16, no. 19, pp. 3825–3836, 2022.
- [86] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3, pp. 279–292, 1992.
- [87] R. S. Sutton, D. McAllester, S. Singh, and Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation,” *Advances in neural information processing systems*, vol. 12, 1999.
- [88] V. Konda and J. Tsitsiklis, “Actor-critic algorithms,” *Advances in neural information processing systems*, vol. 12, 1999.
- [89] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, J. Pineau, *et al.*, “An introduction to deep reinforcement learning,” *Foundations and Trends® in Machine Learning*, vol. 11, no. 3-4, pp. 219–354, 2018.
- [90] F. Meier, A. Mujika, M. M. Gauy, and A. Steger, “Improving gradient estimation in evolutionary strategies with past descent directions,” *arXiv preprint arXiv:1910.05268*, 2019.
- [91] H. Mania, A. Guy, and B. Recht, “Simple random search provides a competitive approach to reinforcement learning,” *arXiv preprint arXiv:1803.07055*, 2018.
- [92] W. Gao, R. Fan, R. Huang, Q. Huang, W. Gao, and L. Du, “Augmented random search based inter-area oscillation damping using high voltage dc transmission,” *Electric Power Systems Research*, vol. 216, p. 109 063, 2023.
- [93] R. Huang, W. Gao, R. Fan, and Q. Huang, “Damping inter-area oscillation using reinforcement learning controlled tcsc,” *IET Generation, Transmission & Distribution*, vol. 16, no. 11, pp. 2265–2275, 2022.
- [94] R. Huang, W. Gao, R. Fan, and Q. Huang, “A guided evolutionary strategy based static var compensator control approach for inter-area oscillation damping,” *IEEE Transactions on Industrial Informatics*, 2022.
- [95] N. Flourentzou, V. G. Agelidis, and G. D. Demetriades, “Vsc-based hvdc power transmission systems: An overview,” *IEEE Transactions on power electronics*, vol. 24, no. 3, pp. 592–602, 2009.
- [96] R. Rudervall, J. Charpentier, R. Sharma, *et al.*, “High voltage direct current (hvdc) transmission systems technology review paper,” *Energy week*, vol. 2000, pp. 1–19, 2000.
- [97] P. Bresesti, W. L. Kling, R. L. Hendriks, and R. Vailati, “Hvdc connection of offshore wind farms to the transmission system,” *IEEE Transactions on energy conversion*, vol. 22, no. 1, pp. 37–43, 2007.

- [98] S. P. Azad, R. Iravani, and J. E. Tate, “Damping inter-area oscillations based on a model predictive control (mpc) hvdc supplementary controller,” *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3174–3183, 2013.
- [99] Y.-X. Ni and A. Fouad, “A simplified two-terminal hvdc model and its use in direct transient stability assessment,” *IEEE transactions on power systems*, vol. 2, no. 4, pp. 1006–1012, 1987.
- [100] R. Fan, J. Lian, K. Kalsi, and M. A. Elizondo, “Impact of cyber attacks on high voltage dc transmission damping control,” *Energies*, vol. 11, no. 5, p. 1046, 2018.
- [101] M. E. Aboul-Ela, A. Sallam, J. D. McCalley, and A. Fouad, “Damping controller design for power system oscillations using global signals,” *IEEE Transactions on Power Systems*, vol. 11, no. 2, pp. 767–773, 1996.
- [102] D. A. Schoenwald, B. J. Pierre, F. Wilches-Bernal, and D. J. Trudnowski, “Design and implementation of a wide-area damping controller using high voltage dc modulation and synchrophasor feedback,” *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 67–72, 2017.
- [103] R. Fan, S. Wang, R. Huang, J. Lian, and Z. Huang, “Wide-area measurement-based modal decoupling for power system oscillation damping,” *Electric Power Systems Research*, vol. 178, p. 106 022, 2020.
- [104] R. Fan, M. A. Elizondo, H. Kirkham, J. Lian, F. Wilches-Bernal, and D. Schoenwald, “Oscillation damping control using multiple high voltage dc transmission lines: Controllability exploration,” in *2018 IEEE/PES Transmission and Distribution Conference and Exposition (T&D)*, IEEE, 2018, pp. 1–9.
- [105] J. H. Chow and K. W. Cheung, “A toolbox for power system dynamics and control engineering education and research,” *IEEE transactions on Power Systems*, vol. 7, no. 4, pp. 1559–1564, 1992.
- [106] R. Nishihara, P. Moritz, *et al.*, “A distributed framework for emerging ai applications,” *University of California, Berkeley*, 2018.
- [107] D. Trudnowski, D. Kosterev, and J. Undrill, “Pdc damping control analysis for the western north american power system,” in *2013 IEEE Power & Energy Society General Meeting*, IEEE, 2013, pp. 1–5.
- [108] J. Hauer, “Application of prony analysis to the determination of modal content and equivalent models for measured power system response,” *IEEE Transactions on Power Systems*, vol. 6, no. 3, pp. 1062–1068, 1991.
- [109] W. Qiao, “Dynamic modeling and control of doubly fed induction generators driven by wind turbines,” in *2009 IEEE/PES Power Systems Conference and Exposition*, IEEE, 2009, pp. 1–8.

- [110] A. B. Attya and T. Hartkopf, “Wind turbine contribution in frequency drop mitigation modified operation and estimating released supportive energy,” *IET Generation, Transmission & Distribution*, vol. 8, no. 5, pp. 862–872, 2014.
- [111] W. Yan, L. Cheng, S. Yan, W. Gao, and D. W. Gao, “Enabling and evaluation of inertial control for pmsg-wtg using synchronverter with multiple virtual rotating masses in microgrid,” *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1078–1088, 2019.
- [112] P. Moritz *et al.*, “Ray: A distributed framework for emerging {ai} applications,” in *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*, 2018, pp. 561–577.
- [113] N. W. Miller, W. W. Price, and J. J. Sanchez-Gasca, “Dynamic modeling of ge 1.5 and 3.6 wind turbine-generators,” *GE-Power systems energy consulting*, no. 3.0, 2003.
- [114] S. Ghosh, S. Kamalasan, N. Senroy, and J. Enslin, “Doubly fed induction generator (dfig)-based wind farm control framework for primary frequency and inertial response application,” *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 1861–1871, 2015.
- [115] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization.,” *Journal of machine learning research*, vol. 13, no. 2, 2012.