DEEP REINFORCEMENT LEARNING-BASED NONLINEAR CONTROL FOR MAGNETIC LEVITATION SYSTEMS OF MAGLEV TRAINS

ZHU QI

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Civil and Environmental Engineering

Deep Reinforcement Learning-Based Nonlinear Control for Magnetic levitation Systems of Maglev Trains

ZHU Qi

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

May 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

ZHU Qi (Name of student)

ACKNOWLEDGEMENT

First of all, I would like to express my deepest gratitude to my supervisor, Prof. Yi-qing NI, and co-supervisor, Dr. Su-mei WANG, for their continuous support. Their enlightening ideas and endless guidance help me to find this area that fits my interests and abilities and learn the way to do research.

I would like to thank all my group mates as well for their accompany. Finally, I want to express my sincerest gratitude for the financial support provided by the Hong Kong Government and the Hong Kong Polytechnic University.

ABSTRACT

Magnetic levitation (maglev) train as a novel transportation can offer various advantages including non-contact operation, minimal vibration, low noise, no risk of derailment, flexible route selection, and cost-effective construction. However, the levitation for the electromagnetic suspension system (EMS) of the maglev train is accomplished based on the magnetic attraction force between electromagnets and guideway. When considered without feedback control, the maglev train system exhibits inherently unstable open-loop dynamics. Furthermore, the maglev train system possesses complex, nonlinear dynamic characteristics. External disturbances, such as wind and variations in passenger load, can also impact the system during operation. These problems can seriously affect the stability and reliability of the maglev train and may lead to partial levitation-point failure. Designing appropriate controllers for the maglev train system is a pivotal problem for maglev train.

In recent years, although the control of maglev system has obtained great achievement, there still exists some drawbacks: (1) The maglev system for controller design is commonly linearized, and lack in automatic adjustment of control strategies. (2) Uncertainty in maglev train system modelling. (3) Most existing control methods can not guarantee the safe boundary for the controller. (4) Existing researches neglected the effect of crosswinds when designing controllers. (5) Lack of methods for considering coupling effect between two levitation points at one side of the bogie.

In this dissertation, the EMS-type maglev control system of the maglev train is chosen as the research object. Based on the problems and drawbacks mentioned, further researches have been carried out to solve the levitation problems with a novel method named deep reinforcement learning (DRL) in this chapter. Main work for this chapter have been listed as follows:

1. Transfer learning-based DRL (TL–DRL) is proposed to develop an adaptive nonlinear levitation system controller that enables automatic adjustment of control strategies. First, levitation control based on DRL is mathematically modeled using Markov decision processes, and the nonlinear state space of a single electromagnet levitation control system is established as an agent–environment interaction with the developed deep reinforcement learning controller. Then a twin delayed deep deterministic policy gradient algorithm in an actor–critic framework is adopted to solve the Markov decision processes. To address the dispersion caused by nonlinear suspension control, a transfer learning-based two-stage training process is devised that first trains the twin delayed deep deterministic policy gradient networks on a linearized model and then transfers the networks to a nonlinear model. The effectiveness of the new controller is verified by comparing it with a conventional proportional–integral–derivative (PID) controller and an adaptive sliding mode controller. The robustness of the TL–DRL controller is examined in the presence of uncertainty, such as train load changes and disturbance forces in the suspension system.

2. Considering the coupling effect between two levitation units of the levitation bogie, a cooperative levitation controller based on the Hamilton-Jacobi-Bellman incorporated multi-agent reinforcement learning (HJB–MADRL) is proposed. The MADRL is adopted for the two-point levitation control considering the coupling effect between the two levitation points. To improve the training of the value network in the MADRL, the HJB function is used in control theory to evaluate the optimality of the value function. The proposed algorithm shows an improved performance compared to the original MADRL algorithm. The effectiveness of the proposed cooperative controller using the proposed algorithm is verified by comparing with a conventional PID controller and a model-guided controller. The robustness of the HJB–MADRL controller is examined in the presence of pitch motion, change in train load, disturbance force, and track irregularity.

3. To ensure the stability and safety of the air gap between the train and its guideway, a safe deep reinforcement learning (SDRL) controller for the maglev system considering the deformation of the flexible guideway is proposed. Notably, a reciprocal control barrier function (RCBF) is augmented in the reward function of the DRL to ensure safety and optimality of the controller. Additionally, a damping coefficient is incorporated into the designed RCBF to specify the trade-off between safety and optimality. The improved performance of the proposed SDRL is verified by comparing to original DRL algorithm. The superiority of the proposed controller is validated through a comparative analysis with a traditional PID controller and a

genetic algorithm tuned super twisting sliding mode controller (GA–ST–SMC) via simulations. Additionally, the robustness of the proposed controller is assessed under conditions of changing train loads, load fluctuations, external disturbances, and track irregularities. Furthermore, experiments have also been conducted to validate the control performance of the proposed RCBF–SDRL controller in comparison to the PID controller on a magnetic levitation system.

4. To investigate the impact of crosswinds on maglev trains, a numerical model is constructed in ANSYS Fluent Meshing, considering the complexities of the environment. Validation of this numerical model is conducted through a wind tunnel test. Subsequently, the principles of fluid mechanics similarity are employed to scale wind forces to a real-world maglev train scenario. To reduce the wind effect on the maglev train, a safe deep reinforcement learning (SDRL) controller is adopted to adjust the control signal for the maglev train–guideway coupling system. Notably, a reciprocal control barrier function (RCBF) is augmented in the reward function of the DRL to ensure safety and optimality of the controller. The superiority of the proposed controller in terms of efficiency and accuracy is validated through a comparative analysis with a traditional PID controller under varying crosswind speeds and train speeds.

Finally, the future work plans are presented.

Keywords: Maglev levitation system, nonlinear control, cooperative control, transfer learning, deep reinforcement learning

LIST OF PUBLICATIONS

Journal Articles

Zhu, Q., Wang, S. M., & Ni, Y. Q. (2024). Cooperative Control of Maglev Levitation System via Hamilton-Jacobi-Bellman Multi-Agent Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, 73(9), 12747–12759. (SCI, Q1, IF=6.1)

Zhu, Q., Wang, S. M., & Ni,Y. Q. (2024). A Review of Levitation Control Methods for Low- and Medium Speed Maglev Systems. *Buildings*,14(837). (SCI, Q2, IF=3.1)

Wang, S. M., **Zhu, Q.**, Ni, Y. Q., Junqi Xu, J. Q., & Chen, F. (2023). Dynamic performance of low-and medium-speed maglev train running on the turnout. *Mechatronics and Intelligent Transportation Systems*, 2(1), 32-41.

Li, H. W., Zhang, D., Lu, Y., Ni, Y. Q., Xu, Z. D., **Zhu, Q.**, & Wang, S. M. (2025). Self-Tuning Dual-Layer Sliding Mode Control of Electromagnetic Suspension System. *IEEE Transactions on Intelligent Transportation Systems*, 26(2), 2366–2380. (SCI, Q1, IF=7.9)

Zhu, Q., Wang, S. M., & Ni,Y. Q. Transfer Learning-Based Deep Reinforcement Learning for Adaptive Control of Maglev Trains. *Engineering Applications of Artificial Intelligence*. (Pending)

Zhu, Q., Wang, S. M., & Ni,Y. Q. Reciprocal Control Barrier Function Enhanced Safe Deep Reinforcement Learning Control for Maglev Train–Guideway Coupling System. *Mechanical Systems and Signal Processing*. (Pending)

Zhu, Q., Wang, S. M., & Ni,Y. Q. Enhanced Deep Reinforcement Learning Controller for Maglev Train-Guideway Coupling Systems in Crosswind Conditions. *Vehicle System Dynamics*. (Accepted)

Wang, S. M., **Zhu, Q.**, Zhang, M. H. & Ni,Y. Q. Switching logic-based saturated

control for maglev suspension systems based on disturbance observer. *Nonlinear Dynamics*. (Pending)

Conference papers

Zhu, Q., Wang, S. M., Ni, Y. Q. (2023). An adaptive MADRL approach for cooperative control of nonlinear maglev suspension system. *The 7th International Conference on Structural Dynamics*, July 2-5 2023, Delft, Netherlands.

Zhu, Q., Wang, S. M., Ni, Y. Q. (2023). An adaptive MADRL-HJB approach for cooperative control of nonlinear maglev suspension system. *The 11th National Academic Conference on Maglev Technology and Vibration Control*, August 4-7 2023, Changsha, China.

Zhu, Q., Wang, S. M., Ni, Y. Q. (2024). An adaptive MADRL-HJB approach for cooperative control of nonlinear maglev suspension system. *The World Transportation Convention 2024*, June 26-29 2023, Qingdao, China.

Awards

Outstanding Project Award in Reinforcement Learning Innovation and Creativity Competition. held by Jiangsu Association of Artificial Intelligence, Digital Brain Laboratory, & POLIXIR, (2022). (Participator: **Zhu Q.** and Lu Y.)

Patents

Chinese Patent: Maglev train control system based maglev control method, device and controller (Wang, S. M., Zhang, M. H., **Zhu Q.,** Ni, Y. Q.). China Invention Patent No. ZL202211013918.0. (published)

Chinese Patent: Deep reinforcement learning and disturbance observer based dynamic maglev levitation control method and system. (**Zhu Q.,** Wang, S. M., Ni, Y. Q.). Chinese Invention Patent Application No. 202210429956.8. (submitted)

TABLE OF CONTENTS

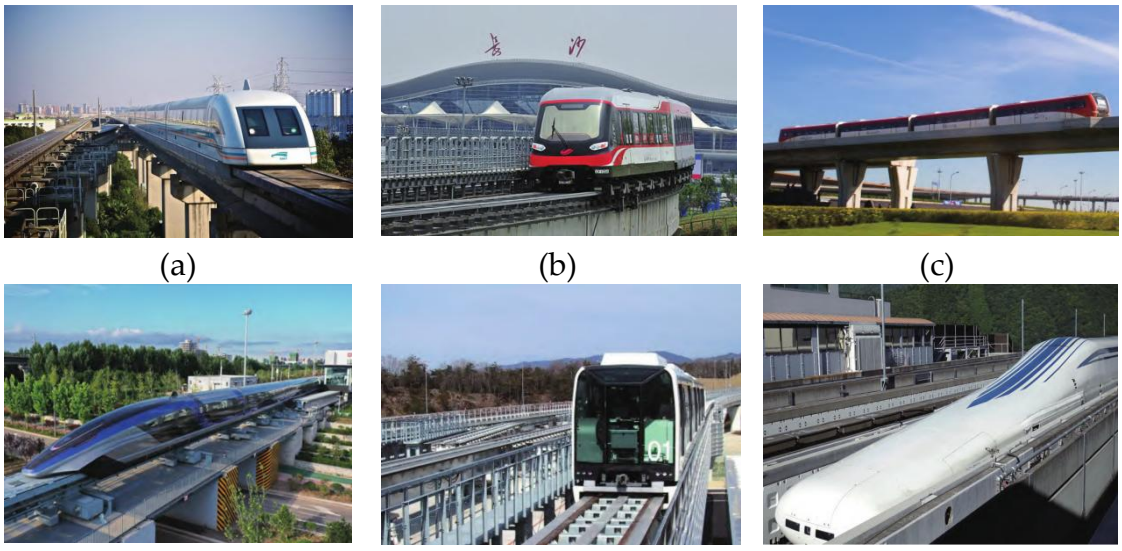
LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Research Background

In recent decades, railway transportation has experienced rapid global development due to its energy efficiency and increased transport capacity during operation (Wang et al., 2023). Nevertheless, conventional wheel-based rail transport systems face several technical challenges associated with noise, adhesion, rail-wheel wear, and vibration (Liu et al. 2022). Maglev train systems have emerged as innovative railway transport systems in response to the increasing demand for higher speeds and enhanced comfort in intercity travel. They offer various advantages including non-contact operation, minimal vibration, low noise, no risk of derailment, flexible route selection, and cost-effective construction. As a result, the maglev trains have garnered significant attention for their development and implementation in countries such as Japan, Korea, and China (Lee et al., 2006). Some maglev trains and lines are depicted in **Figure** 1–1.



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 1–1 Maglev trains and lines: (a) Shanghai high-speed maglev demonstration line; (b) Changsha maglev express; (c) Beijing S1 maglev demonstration line; (d) Qingdao high-speed maglev transportation system; (e) Japan medium and low speed maglev linimo line; (f) Japan high-speed L0 maglev train; (g) Korea Ecobee maglev line; (h) America MagTube

Based on the levitation mechanism in levitation, maglev train systems can be broadly categorized into electromagnetic levitation (EMS), electrodynamic levitation (EDS), and hybrid electromagnetic suspension (HEMS) types as in **Figure** 1–2 (Lee et al., 2006). The EDS-type system utilizes repulsive forces for levitation and is inherently stable due to the proportional relationship between the repulsive force and the gap. When the gap decreases, the repulsive force increases, which ensures a stable levitation state (Thornton, 1991). However, EDS-type requires sufficient speed to acquire sufficient induced currents for levitation. EDS systems can be divided into permanent magnet (PM) and superconducting magnet (SCM) types regarding the magnets. The structure of the PM-type system is very simple and does not require an electric power supply. However, it is limited to small-scale applications due to the

lack of high-power permanent magnets. In contrast, the SCM-type system has a more complex structure, and its operation may face challenges such as quenching and evaporation of liquid helium caused by heat generated from induced currents. Unlike the EDS-type system, the EMS-type system achieves levitation by utilizing the magnetic attraction force between the guideway and electromagnets. Such levitation is unstable due to the nature of the magnetic circuit, and precise control is necessary to maintain the uniform air gap. In the EMS-type maglev system, there are two levitation technologies: the integrated levitation and guidance type and the separated levitation and guidance type. The integrated type is more suitable for low-cost and low-speed operations, as it reduces the number of electromagnets and controllers required, with the guiding force generated automatically due to differences in reluctance. The separated type is utilized for high-speed operations, where levitation and guidance function independently without interference. However, this type requires a greater number of controllers. Nonetheless, the EMS-type maglev system has a significant advantage compared with the EDS-type in that it is easier and can provide an attractive force at zero or low speed (Taghirad et al., 2023). To reduce electric power consumption in EMS-type trains, the HEMS-type maglev system incorporates permanent magnets (PMs) alongside electromagnets. However, HEMS requires significantly larger variations in current amplitude compared to EMS, as the permeability of PMs is the same as that of air.

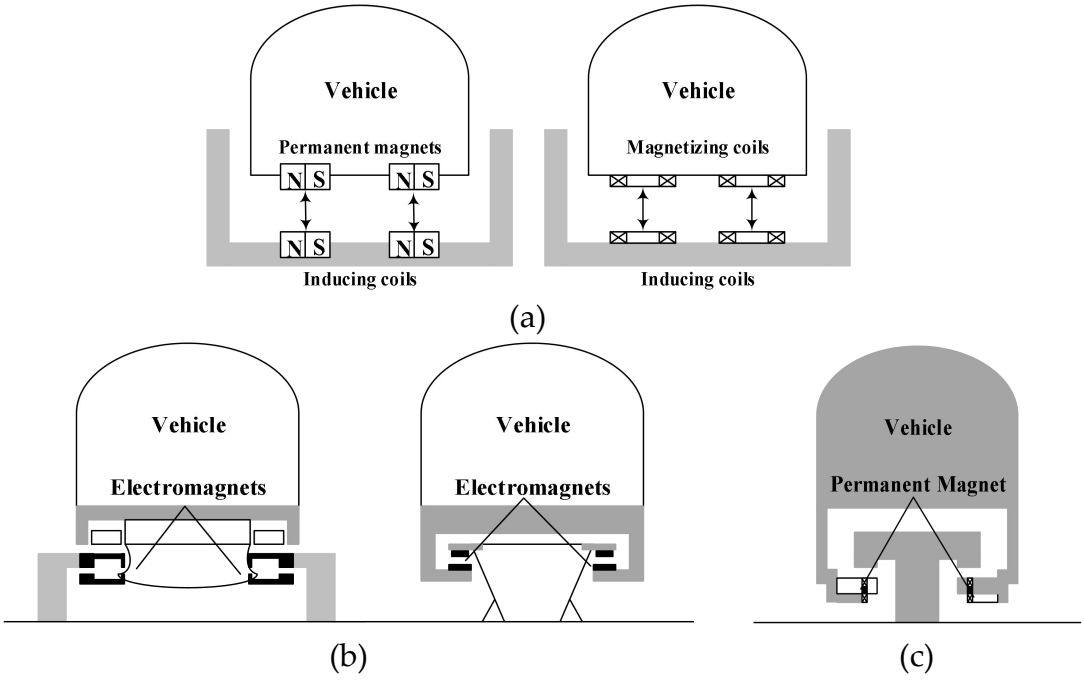


Figure 1 - 2 Levitation types: (a) Electrodynamic suspension type: permanent magnets and superconducting magnets; (b) Electromagnetic suspension type: integrated and separated types; (c) Hybrid electromagnetic suspension type

At present, commercial maglev train lines primarily utilize EMS-type systems, such as the Changsha Maglev line in China and the EcoBee line in South Korea. These systems are predominantly designed for low- to medium-speed operations. Given the substantial impact of the dynamic performance of the magnetic levitation system on the overall performance of maglev trains, the implementation of an active levitation control system is crucial to ensure the safety and stability of EMS-type maglev trains. The maglev levitation system models for control design in the literature can mainly be divided into two categories: dynamic equation-based modeling and data-driven-based modeling. In dynamic equation-based methods, numerical models

are built based on Newton's law and Kirchhoff's law. The parameters of the models are predefined using the mechanical properties of the maglev levitation system, and uncertainties associated with the system and environmental disturbances such as wind load can be updated using the measured data. Unlike dynamic equation-based modeling, data-driven-based modeling directly uses the history data for training.

The levitation control techniques can mainly be classified into three categories according to the control algorithms: classical linear control methods, classical nonlinear control methods, and artificial intelligent control methods. In general, a linearized model of an EMS-type system is established around the equilibrium point and controlled using linear control algorithms such as the proportional-integral-derivative (PID) (Sun et al., 2016; Yang et al., 2004) and linear quadratic regulator (LQR) (Long et al., 2007; Unni et al., 2016; Yang et al., 2011) theories. Nevertheless, it is worth noting that maglev levitation controllers based on linearized models are typically implemented within proximity to the equilibrium point. Due to the inherently strong nonlinearity of the magnetic levitation system, controllers based on linear simplified designs may become unstable or ineffective when the system is subjected to external disturbances. To enable the designed levitation controller to smoothly and efficiently handle external disturbances, advanced levitation controllers based on different nonlinear control theories have been proposed to improve the overall performance of the control system, such as robust control (Liu and Yao, 2016; Xu et al., 2015; Xu et al., 2018), adaptive control (Nguyen, 2018; Zhang and Li, 2018), and sliding mode control (SMC) (Kong et al., 2011; Li et al., 2022; Zhu et al., 2022). However, most of these control strategies degrade in performance when there are disturbances because they rely on a precise model and

detailed information about the system. Artificial intelligent control methods, including the fuzzy logic method (Chen et al., 2020; He et al., 2015), neural networks (NN) (Sun et al., 2019; Wai and Lee, 2008 & 2009; Wai et al., 2014; Wai et al., 2015), deep belief networks (DBN) (Sun et al., 2021), and deep reinforcement learning (DRL) (Wang et al., 2020), have been implemented to increase the robustness of the control system by automatically adjusting the parameters in the controller or directly acting as a controller.

In this study, advanced deep reinforcement learning (DRL) methods will be adopted to the EMS-type low- and medium-speed system control. The research objectives are listed in Section 1.2.

1.2 Research Gaps and Objectives

After conducting a literature review, the research gaps of EMS-type low- and medium-speed system control are summarized as below:

1) Limitation of linear controller: The controllers adopted in real maglev commercial lines are linear controllers like PID. When deviates from the equilibrium point, the linear controller would easily fail.

2) Requirement for precise modelling: The EMS-type maglev train exhibits highly nonlinear behavior due to the complex interactions between its magnetic field, quality of tracks, and driving conditions. Obtaining an accurate model of the system is challenging due to these nonlinear behaviors.

3) Robustness against external disturbances: The maglev system is subject to various external disturbances during operation, including load changes caused by passengers during operation, track irregularity, and disturbances caused by wind or other unpredictable disturbances in the operating environment. The control method needs to be robust enough to handle these disturbances and maintain stability of the levitation system.

4) Limitations of the control system structure: The structure of the control system in the EMS-type maglev levitation systems can have limitations on control performance. In the decentralized control structure, each levitation point on a decentralized maglev levitation bogie has its control loop but all levitation points

share the same control parameters. Designing control parameters for all levitation points to ensure robustness of all the levitation points under worst-case condition poses a significant challenge.

5) Train-track coupling effect: The construction cost of the rail-bridge system exceeds 60 – 70% of the total initial investment in a maglev train. Therefore, slimmer guideway and looser construction tolerances are urgently needed to lower overall construction costs. However, a lighter guideway increases flexibility, leading to stronger coupling between the maglev train's control loops and the flexible guideway.

To overcome the research gaps mentioned, advanced nonlinear control methods using deep reinforcement learning (DRL) algorithms are proposed in this research. The main objectives of this research are as follows:

1) Taking full advantage of the inherent nonlinearities of maglev levitation systems, enhancing the levitation performance, and simultaneously, achieving much less energy consumption.

2) Develop a multi-point levitation DRL control method with model uncertainty and control robustness guaranteed.

3) Develop a model-free maglev levitation system based safe DRL (SDRL) controller considering the train-track coupling effect in the control design with safety ensured.

4) Develop a model-free SDRL controller to achieve maglev levitation control with various external disturbances considered, especially the crosswinds.

1.3 Thesis Outline

This thesis mainly covered the nonlinear control of EMS-type low and medium speed maglev trains using DRL controllers. The outline is listed as follows:

Chapter 1 is an introduction, including the brief research background, main objectives, and report outline.

Chapter 2 presents the literature review, including overview of the EMS-type maglev train, a short history of maglev train development, and maglev levitation control methods, categorized as conventional linear control, conventional nonlinear control, and artificial intelligent control.

Chapter 3 proposes a transfer learning-based DRL (TL – DRL) algorithm to develop an adaptive nonlinear levitation system controller that enables automatic adjustment of control strategies. First, levitation control based on DRL is mathematically modeled using Markov decision processes, and the nonlinear state space of a single electromagnet levitation control system is established as an agent – environment interaction with the developed deep reinforcement learning controller. Then a twin delayed deep deterministic policy gradient algorithm in an actor – critic framework is adopted to solve the Markov decision processes. To address the dispersion caused by nonlinear suspension control, a transfer learning-based two-stage training process is devised that first trains the twin delayed deep deterministic policy

gradient networks on a linearized model and then transfers the networks to a nonlinear model. The effectiveness of the new controller is verified by comparing it with a conventional proportional – integral – derivative (PID) controller and an adaptive sliding mode controller. The robustness of the TL – DRL controller is examined in the presence of uncertainty, such as train load changes and disturbance forces in the suspension system.

Chapter 4 proposes a cooperative levitation controller based on the Hamilton-Jacobi-Bellman incorporated multi-agent reinforcement learning (HJB – MADRL) considering the coupling effect between two levitation units of the levitation bogie. The MADRL is adopted for the two-point levitation control considering the coupling effect between the two levitation points. To improve the training of the value network in the MADRL, the HJB function is used in control theory to evaluate the optimality of the value function. The proposed algorithm shows an improved performance compared to the original MADRL algorithm. The effectiveness of the proposed cooperative controller using the proposed algorithm is verified by comparing with a conventional PID controller and a model-guided controller. The robustness of the HJB – MADRL controller is examined in the presence of pitch motion, change in train load, disturbance force, and track irregularity.

Chapter 5 proposes a safe deep reinforcement learning (SDRL) controller for the maglev system considering the deformation of the flexible guideway, so as to

ensure the stability and safety of the air gap between the train and its guideway. Notably, a reciprocal control barrier function (RCBF) is augmented in the reward function of the DRL to ensure safety and optimality of the controller. Additionally, the designed RCBF includes a damping coefficient to balance safety and optimality. The improved performance of the proposed SDRL is verified by comparing to original DRL algorithm. The superiority of the proposed controller is validated through a comparative analysis with a traditional PID controller and a genetic algorithm tuned super twisting sliding mode controller (GA – ST – SMC) via simulations. Additionally, the robustness of the proposed controller is assessed under conditions of changing train loads, load fluctuations, external disturbances, and track irregularities. Furthermore, experiments have also been conducted to validate the control performance of the proposed RCBF – SDRL controller in comparison to the PID controller on a magnetic levitation system.

Chapter 6 investigates the impact of crosswinds on maglev trains. Firstly, a numerical model is constructed in ANSYS Fluent Meshing, considering the complexities of the environment. Validation of this numerical model is conducted through a wind tunnel test. Subsequently, the principles of fluid mechanics similarity are employed to scale wind forces to a real-world maglev train scenario. To reduce the wind effect on the maglev train, a safe deep reinforcement learning (SDRL) controller is adopted to adjust the control signal for the maglev train – guideway coupling system. Notably, a reciprocal control barrier function (RCBF) is augmented in the

reward function of the DRL to ensure safety and optimality of the controller. The superiority of the proposed controller in terms of efficiency and accuracy is validated through a comparative analysis with a traditional PID controller under varying crosswind speeds and train speeds..

Chapter 7 summarizes the key conclusions of this thesis and provide recommendations for future research.

CHAPTER 2 LITERATURE REVIEW

2.1 Overview of Maglev Train History and System

2.1.1 History background of the EMS-type maglev train

The development of maglev trains can be traced back to 1934 when Hermann Kemper of Germany patented the concept. In 1969, the German company Krauss-Maffei achieved a significant milestone by developing the first prototype model of a Maglev train in their laboratory, known as Transrapid (TR) 01. In the early 1970s, the German government initiated a high-speed maglev train transportation development plan, leading Krauss-Maffei to develop subsequent models, namely TR02 and TR04, based on the TR01 prototype. In 1991, the German government conducted the evaluation of the TR07 model and deemed it ready for practical application. Following this, the development of the TR08 and TR09 maglev trains took place. The TR08 model achieved a significant milestone as it became the first commercially operational maglev train in the world. It was successfully implemented in the joint Germany and China project, connecting Shanghai Pudong Airport to Longyang Road subway station.

In 1974, Japan Airlines (JAL) made a significant investment by purchasing the TR04 model from Krauss-Maffei. Building upon this technology, JAL developed a series of low-speed maglev trains and high-speed surface transport (HSST) systems

based on the TR04 model. The bogie structure utilized in the HSST-05 design became widely adopted in subsequent low-speed maglev train designs. Based on the HSST-05, two models were developed: HSST-100S and HSST-100L, capable of reaching speeds of up to 100 km/h. The HSST-100L model was implemented on the Linimo line in Nagoya, Japan, showcasing its practical application. The Japanese National Railways (JNR) have also made significant contributions to maglev train research. In 1972, they successfully developed the ML100 model, followed by the ML500, which set a world record speed of 517 km/h in 1979. The MLX01, developed in 1997, achieved a remarkable speed of 550 km/h. Furthermore, Japan's efforts in maglev train technology extended to the development of the low-temperature superconducting electric maglev vehicle known as L0. Tested on the Yamanashi test line, the L0 achieved a remarkable world record speed of 603 km/h, highlighting Japan's advancements in high-speed maglev train capabilities.

Korea has also made notable advancements in the development of low-speed maglev trains, intending to address traffic congestion issues in major cities. The improvement of the urban air mobility (UTM)-01 took place in 1999, leading to the subsequent development of UTM-02. In 2004, a study focusing on the commercial operation of UTM-02 based low-speed maglev trains commenced. The demonstration line for this study was chosen to be located at Incheon International Airport. However, it took until 2016 for the commercial line to be officially put into operation, serving as a solution to the city's transportation needs.

China initiated research on maglev trains in the 1980s, with various educational institutions and research institutes actively participating in the field. Notable contributors include many universities and institutes like the National University of Defense Technology, the Chinese Academy of Sciences, etc. Several organizations embarked on the development of a hybrid magnet maglev train, achieving success in 2012. A 425-meter-long test line was constructed in Chengdu, and a low-speed electromagnetic maglev train was developed. This train underwent testing on the line in 2006. Later, 1,500-meter-long low-speed maglev test line was established in Shanghai, resulting in the development of a low-speed maglev train. Additionally, the China Southern Railway Corporation Limited group (CSR) began studying low-speed maglev train technologies, establishing a test line at the Zhuzhou Railway Vehicle Factory.

2.1.2 Overview of EMS-type maglev train system

The EMS-type maglev train is suspended on the track via an electromagnetic force instead of the wheel-rail contact force in conventional rail transport systems. The electromagnet is adopted in the maglev train for both levitation and guidance functions, and the linear motor is utilized for the traction function. In general, EMS-type maglev trains consist of five main components: levitation and guidance, traction, bogie, braking, and rail-bridge systems. A detailed introduction to the five components is provided in the following sections.

1) Levitation and guidance system

Figure 2 - 1 shows the levitation and guidance system of an EMS-type maglev train. Its levitation and guidance functions are realized through electromagnets installed on the levitation bogie and an F-type rail. The electromagnets are energized to attract the F-type rail to achieve levitation of the vehicle, and the current adjustment in the electromagnets ensures that the train is stably suspended in a dynamic stability range (typically 8-12 mm).

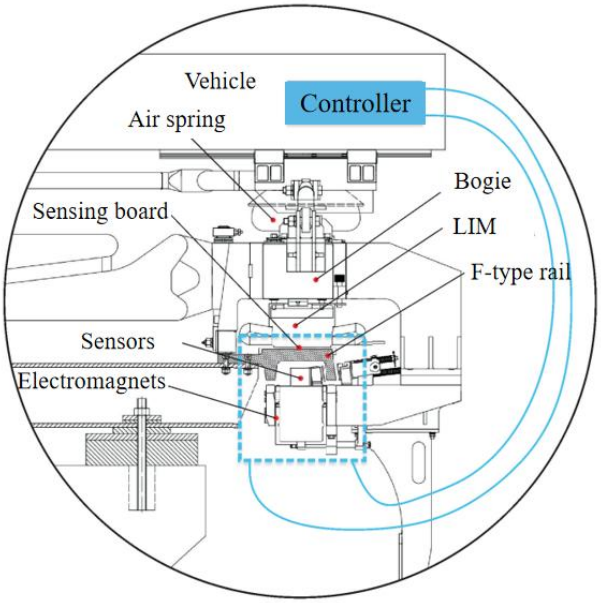


Figure 2 - 1 Cross section of levitation-guidance system of EMS-type maglev train

In the levitation control system, the electromagnetic force is inversely related to the air gap square, leading to instability in the maglev levitation system. Hence, feedback control is required to stabilize the air gap between the electromagnet and the

F-type rail. The absolute accelerations of the electromagnets and the relative air gap are monitored by the sensors and fed back to the suspension controller in real-time. The control algorithm then deals with the obtained sensor signals and adjusts the current accordingly. The levitation controller is composed of a control unit and a power unit. The control unit receives and filters the sensor signal, conducts the control algorithm, and generates a drive signal for the power unit. The power unit is utilized as a chopper to provide power to the electromagnets. The control structure of the levitation system is illustrated in **Figure 2–2**.

In the suspended state, if the levitation system causes a lateral offset relative to the track electromagnets, the electromagnetic suction generates lateral components due to the inverted track configuration, providing guidance for the train.

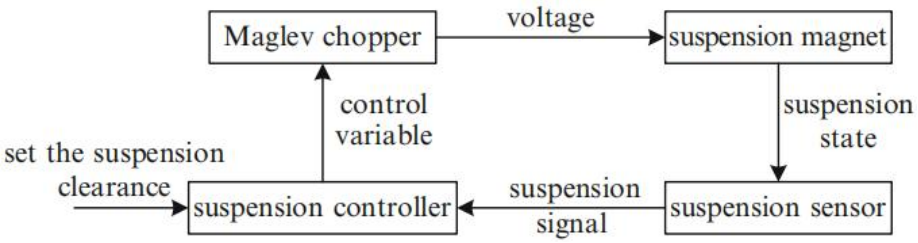


Figure 2–2 Control structure of suspension system

2) Traction system

The linear motor is used for the traction of the maglev trains. It can be divided into two types: long-stator linear synchronous motors (LSM) and short-stator linear induced motors (SSLIM). The long-stator LSM is better suited for high-speed maglev

trains, whereas the latter is typically used for low- and medium-speed maglev trains due to its relatively simple structure, lower manufacturing costs, and lower efficiency and power factor.

3) Bogie system

The bogie system integrates functions of levitation, guidance, and traction. The levitation electromagnets are mounted on a box beam made up of magnetic modules. For each module, there are four electromagnet pole-line packages and two electromagnets included, which are part of an independent levitation control system. Two sides of modules are connected by two anti-roller beams, which join the components to form a complete bogie. The four levitation points are mechanically decoupled by anti-roller beams to enhance levitation stability. A schematic of the bogie structure of the EMS-type maglev trains is shown in **Figure 2–3**.

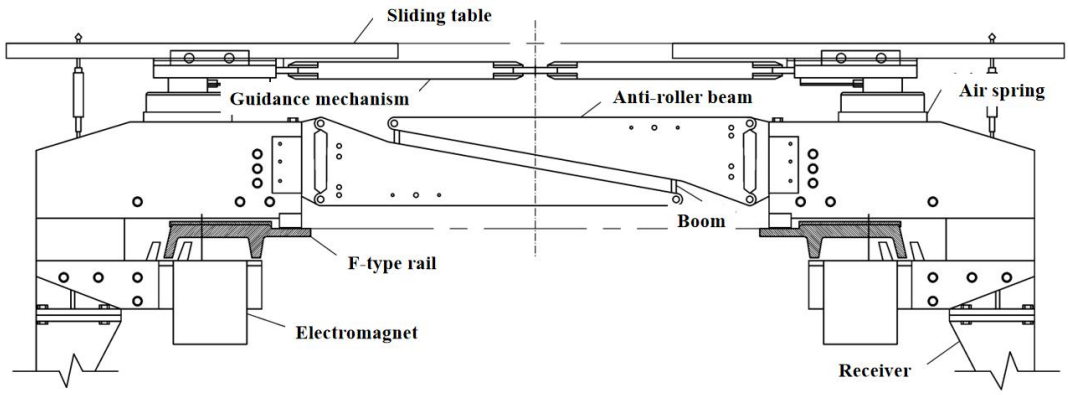


Figure 2–3 Bogie structure for EMS-type maglev trains

4) Braking system

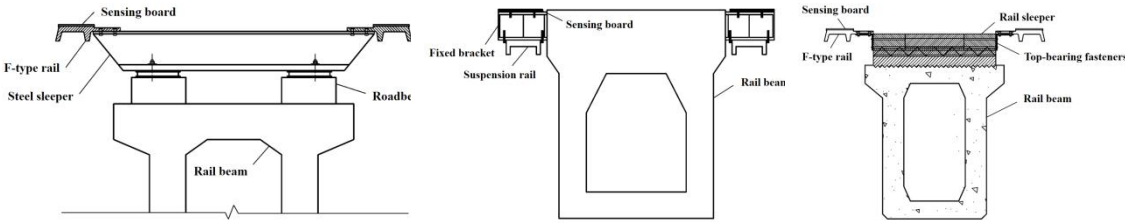
The braking system is a key component to ensure the safe operation of maglev trains. The braking systems of EMS-type maglev trains can be divided into common,

fast, emergency, and holding brake methods, which include electric braking, gas-to-liquid braking, and hydraulic braking. Gas-to-liquid braking and hydraulic braking are two common methods adopted in braking systems.

5) Guideway system

The difference in the supporting and guidance functions between the maglev train system and the conventional rail transport system results in a difference in the guideway design. The construction cost of the rail-bridge system exceeds 60-70 % of the total initial investment in a maglev train. Thus, a slimmer guideway is essential to lower construction costs of the maglev line. However, the coupling between the vehicle and the flexible guideway is enhanced due to the greater flexibility of the guideway. A compromise between stability and loosening tolerances can be achieved through an analysis of vehicle/guideway interaction.

In the design stage, issues such as guideway deflection, span length, guideway mass, natural frequency of the main beam, natural frequency of the track, rail joint, surface roughness, and long-stator profile must be investigated by conducting tests with experimental or numerical vehicle-guideway dynamic interaction analysis. For EMS-type maglev trains, the guideway structures can be divided into three types: steel sleeper-based rail structures, direct-connected rail structures without sleepers, and integral bed rail structures, as shown in **Figure 2–4**.



(a) (b) (c)

Figure 2–4 Structure sections of guideway of EMS-type maglev trains: (a) steel sleeper based rail structure, (b) direct-connected rail structure without sleeper, (c) integral bed rail structure

2.2 Maglev levitation System Control

With the availability of various control methods, they have been applied to control the electromagnetic levitation system in the maglev train (**Figure** 2–5). Control methods in recent years will be discussed in this section.

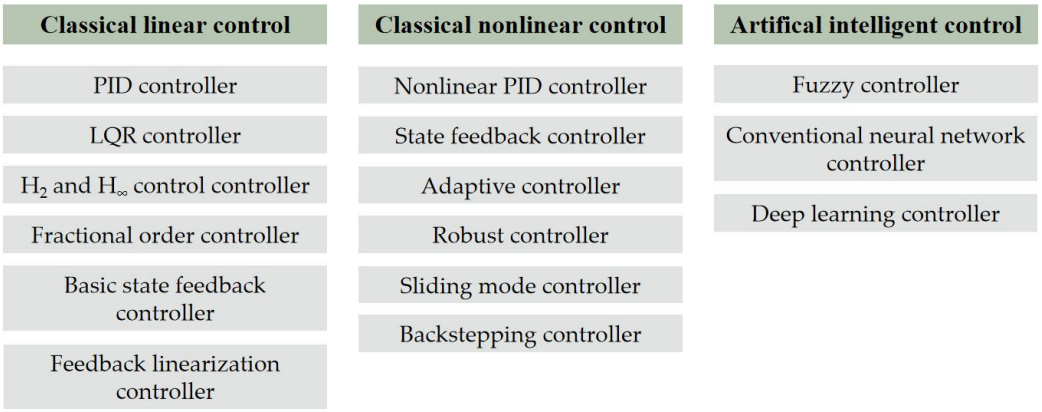


Figure 2–5 Schematic of the control methods

2.2.1 Conventional Linear Control

The conventional levitation controller for maglev trains uses linear control methods based on a linearized maglev train system (Sun et al., 2016; Yang et al., 2004). These methods are generally based on the single-point levitation model and assume that the levitation electromagnet and guideway are in a stable position (Yang et al., 2004). Based on the linearized system model, this section introduces typical control methods such as the proportional-integral-differential (PID) control (Duka et al., 2016; Morari et al., 1984; Sun et al., 2016; Yang et al., 2004), linear quadratic

regulator control (LQR) (Long et al., 2007; Unni et al., 2016; Yang et al., 2011), and state feedback control (Liu et al., 2022; He et al., 2010; Xia et al., 2020).

However, when the air gap is far from the equilibrium point, the stability of the control system based on these linear control methods significantly decreases. To solve this problem, a feedback linearization method, which is a first-order derivative transformation process for a nonlinear system at the desired inputs, was proposed (de Jesús Rubio, 2018). Advantage and disadvantage of different linear control algorithms are summarized as below:

1) The PID control algorithm is easy to apply, and the choice of the three coefficients is of great importance for designing the proper PID controller. However, there are two disadvantages in PID: 1) compared with other controllers such as the LQR controller, the PID controller exhibits larger peak overshoot and settling time, and 2) the disturbance cannot be ignored in real applications. Real-time changes in the three coefficients are required to adapt to different scenarios. Therefore, the fuzzy logic and IMC methods have been used in PID control to better tune the coefficients in the PID controller.

2) The LQR controller assumes that all state variables are accessible for feedback. However, in a maglev system, it is widely recognized that the velocity of the air gap cannot be directly measured, and the presence of disturbances significantly influences the system's behavior. Thus, the state estimation and disturbance observer-based control (DOBC) can be combined to improve the performance of the

LQR controller.

3) H_2 and H_∞ control algorithms use the H_2 and H_∞ norms, which are two popular measures in optimal control theory for control design, and are robust when there are disturbances.

4) The fractional order control algorithm is normally combined with other methods such as PID and sliding mode control by introducing additional fractional order parameters in the control design. With additional parameters, the traditional controllers show better performance in terms of robustness and disturbance rejection.

5) The basic state feedback control algorithm uses the state variables of the system for control design. The state variables of the controller must be carefully selected to improve the accuracy and robustness of the controller, and the state observer or model predictive method can be combined to obtain sufficient state variables for controller design.

2.2.2 Conventional Nonlinear Control

The maglev levitation system is a strong nonlinear system, and the aforementioned linear control methods based on the linearized maglev levitation model show poor robustness and low practicality. In recent years, many researchers have applied nonlinear control methods, including adaptive control, backstepping control, and sliding mode control (SMC), to solve maglev levitation control problems. Advantage and disadvantage of different nonlinear control algorithms are summarized

as below:

1) The PID control algorithm applied to nonlinear system control must be carefully designed. For example, it is necessary to change the conventional transfer function into an exponential function to increase the stiffness of the system when it is apart from the operating point or combine it with other methods such as tuned mass damper (TMD) to decrease the system vibration and particle-swarm-optimization (PSO) to optimize the coefficients of the PID.

2) The state feedback control algorithm can address the time delay of the system and real-time disturbances such as nonlinear, periodic bounded disturbances, and uncertainties. It can combine with an optimal method such as PSO to improve the controller parameters.

3) The adaptive control algorithm shows great potential in enhancing the performance of control systems, particularly when handling uncertainties caused by factors such as degradation and modeling inaccuracies. In addition to directly applying adaptive control algorithms to maglev levitation systems, model-assisted/reference adaptive control algorithms are implemented.

4) A robust control algorithm can address the uncertain part of the system and develop an effective design method that considers uncertainty information. The robust controller can also be combined with a disturbance observer to increase robustness to disturbances.

5) SMC is a type of nonlinear control algorithm widely employed in the field of

nonlinear control due to its straightforward physical implementation, rapid response, and robustness. Modifications can be made to the SMC for various purposes. For example, the designed global fast terminal integral sliding mode controller (GFTISMC) improves the global fast response speed of the maglev system and reduces the steady-state error. The nonsingular robust SMC reduces the upper bound of the uncertainty and interference of the SMC. The SMC can also combine with other control algorithms or optimization algorithms, such as the fuzzy logic control method and PSO method.

6) The backstepping control algorithm is typically adopted to decouple nonlinearities and eliminate uncertainties. It can be combined with SMC to reduce the chattering of the SMC and improve the dynamic response of the system.

2.2.3 Artificial Intelligent Control

Artificial intelligent control methods have been introduced in modern control engineering for its self-adaptation, self-learning, self-optimization characteristics. For systems containing uncertain parameters, artificial intelligent control methods like fuzzy control, neural network control, and DL control have proved their superiority versus conventional linear and nonlinear control methods (Wang et al., 2024). Specifically, artificial neural networks are highly interconnected networks that excel in representing complex nonlinear functions with exceptional accuracy through the process of training. The weighting parameters of a neural network can be adjusted to

approximate any desired nonlinear function with the necessary accuracy. This capability allows neural networks to effectively represent complex nonlinear system models or unknown systems (Liu, 2018). With the development of DL networks in recent decades, some DL methods, such as DBNs, convolutional neural networks (CNN), have also been used for controller design. Instead of using networks to represent the system structure, DRL methods can interact with the environment to obtain the optimal series of control signals. Thus, artificial intelligent control methods are capable of dealing with the complex nonlinear problem of maglev levitation systems, and are classified into fuzzy logic control, conventional neural network algorithms, and deep learning algorithms in this study.

1) The fuzzy logic control algorithm can use expert knowledge to obtain the output control signal, which has better control performance than the PID controller and can be directly applied to nonlinear systems. However, traditional fuzzy logic controllers are designed based on human operator experience and cannot linguistically determine the exact action for the output. The T–S fuzzy control method was proposed as a mathematical tool to overcome this problem. In addition, optimization methods such as PSO can be combined to better tune the parameters of the controller. It can also combine the control to ensure that the controlled output is less than a prescribed level and PID to optimally adjust the coefficients and restrain the coupled vibration of the vehicle and guideway.

2) The conventional neural network algorithm can address complex nonlinear

problems that have the characteristics of approximation of nonlinear functions, a simple structure, and fast convergence. This solution can effectively improve the control performance against large uncertainties in the system. In maglev levitation control, conventional neural network algorithms are commonly used as controllers to output the control item or optimize the parameters of other controllers such as PID and state feedback controllers.

3) The deep learning algorithm places emphasis on designing and evaluating training algorithms and model architectures for contemporary neural networks. These networks can better represent the intricate features of complex problems. Although few studies have investigated deep-learning-based maglev levitation control, the application of DBN and DRL has demonstrated the robustness and effectiveness of the deep-learning algorithm in the maglev levitation area.

2.3 Vehicle-guideway vibration suppression control design

Due to the small levitation air gap and strong nonlinear behavior of the maglev control system, the acceleration amplitudes of a series of maglev trains will be significantly amplified (Yau, 2009) and may cause failure in maglev control. In addition, the resonance phenomenon of vehicle-guideway vibration is frequently observed in commercial lines (Wang et al., 2023) and affects the ride comfort and safety of the maglev train. To suppress the vehicle-guideway vibration, many studies have been conducted to analyze the dynamic performance of the vehicle-guideway coupling system and design appropriate controllers to reduce the vibration fluctuation.

For the vibration analysis of vehicle-guideway coupling models, Lengyel et al. (2014) used a simply supported beam connected with a free beam on distributed springs and dampers to describe the guideway and coupled maglev vehicle, and they analyzed dynamic behaviors such as the deflection and acceleration of the guideway and the deformation of the vehicle with different speeds. Kim et al. (2015) developed an integrated model of a UTM vehicle that incorporated a 3-Dimensional (3D) full vehicle model, including the bodies, joints, and force elements derived from prototyping. A 2-dimensional (2D) flexible guideway and feedback controllers were also included to accurately represent all electromagnet levitation units. This integrated model offers a more realistic representation of the UTM vehicle system. The established model was investigated under the standstill and low-speed operation

scenarios. Min et al. (2017) proposed a 3D mathematical maglev system model of the UTM vehicle, compared its dynamic performance with a 2D model, and analyzed the effect of the lateral and vertical track irregularities. For the single electromagnet-guideway coupling model (Chi and Li, 2017; Zhang, 2022), full vertical dynamic coupling vibration model of an HSST type low-speed maglev train (Wang et al., 2018), and mathematical model of the low-speed maglev test car of Tongji University (Sun et al., 2023; Xu et al., 2019), the effect of the time-delay (Wang et al., 2018; Xu et al., 2017), levitation current perturbation , and control gain (Wang et al., 2018) on vibration behavior were studied. Hopf bifurcation theory was adopted by researchers to analyze the mechanism of a fundamental theory of vehicle-guideway coupling vibration (Sun et al., 2023; Zhang, 2022).

Numerous controllers have been proposed to reduce the vibration of the vehicle-guideway coupling system. Some researchers used extra controllers to suppress the vibration. Li et al. (2021) adopted a tuned mass damper (TMD) installed at the guideway to avoid the Hopf bifurcation phenomenon of the train-switch vibration. Wang et al. (2020) introduced a novel approach to address resonant phenomena encountered by a vehicle traversing guideway girders. They proposed the use of an additional feedback controller based on the condensed virtual dynamic absorber (C-VDA) scheme. This approach aims to mitigate the resonant effects and enhance the overall stability of the vehicle's motion on the guideway girders.

Instead of using extra controllers, Yau et al. (2009) proposed a neural-PI control

using back-propagation (BP) neural network to train the PI controller. The authors demonstrated that the proposed controller effectively regulated the acceleration amplitude of running maglev vehicles to ensure that it remained within an acceptable range. Yau et al. (2010a & 2010b) also proposed an LQR+PID controller to reduce the substantial amplification of the acceleration amplitude, which running maglev vehicles at higher speeds experience due to aerodynamic forces. Furthermore, they addressed the dynamic response of maglev trains that travel over levitation bridges subjected to horizontal earthquakes. Kong et al. (2011) proposed a Kalman-filter-based SMC to reduce the vibration response. They discovered that the fluctuation of the air gap and vertical acceleration of the cabin's center of gravity (CG) was significantly influenced by the vehicle speed and guideway irregularity. However, they observed only minimal effects on these factors due to variations in vehicle mass. Wang et al. (2014) developed a full-state feedback controller and used the PSO to optimize the control gains; then, they verified the performance of the proposed controller through simulation and test rig even under violent external disturbances. Zhou et al. (2014) found that the vibration amplitude depended on the voltage supplied by the power source. To address this issue, they proposed a vibration amplitude control method utilizing a PI controller. This method can control the vibration amplitude by adjusting the voltage of the power supply. Zhou et al. (2017) presented an effective novel approach for vibration suppression under various common track irregularities, including sinusoidal track profiles, random track

irregularities, and track steps. Their proposed method uses a pair of mirror FIR filters alongside an adaptation mechanism controller. Sun et al. (2019) developed a fuzzy adaptive tuning PID controller, for which the controller gains were adjusted according to the identified disturbance or changes in the system parameters.

In 2020, Sun et al. (2020) employed an enhanced version of the Apriori algorithm to extract and process data from a stored historical database. This process facilitated the establishment of a reliable database, based on which the researchers developed an improved fuzzy adaptive controller. The proposed controller was verified on a full-scale Internet of Things (IoT) maglev train system at Tongji University. In addition, sliding mode robust adaptive control (Chen et al., 2019), robust control (Li and Shen, 2020), feedback linearization control (Zhang et al., 2022), double loop PID considering control gain perturbation (Sun, et al., 2023), and GA tuned Super Twisting-SMC (ST-SMC) (Teklu and Abdissa, 2023), have been proposed to suppress the vibration of the coupling system.

2.4 Reinforcement learning and deep reinforcement learning

2.4.1 Reinforcement learning

Reinforcement learning (RL) (Sutton and Barto, 2018) is a subset of machine learning focused on improving a system's decision-making capabilities through experience obtained from interacting with the environment. Specifically, Markov Decision Processes (MDPs) represent a mathematically idealized version of the RL problem. This section presents key elements of RL, including return, value functions, and policies.

2.4.1.1 The agent-environment interface

An RL agent learns by periodically interacting with an environment. The agent serves as both the learner and decision-maker, while the entity it interacts with is referred to as the environment. At each time step t, the agent gets a state s_t $(s_t \in \mathcal{S})$ from the environment and chooses an action a_t $(a_t \in \mathcal{A})$ following a policy $\pi(a_t|s_t)$. After executing the the action a_t, the agent receives a scalar reward r_t from the environment, and the state of the environment transitions to s_{t+1} according to the state transition probability $\mathcal{P}(s_{t+1}|s_t, a_t)$. In an episodic problem, the interaction persists until a terminal state occurs, and the agent's goal is to maximize the expected return from each state.

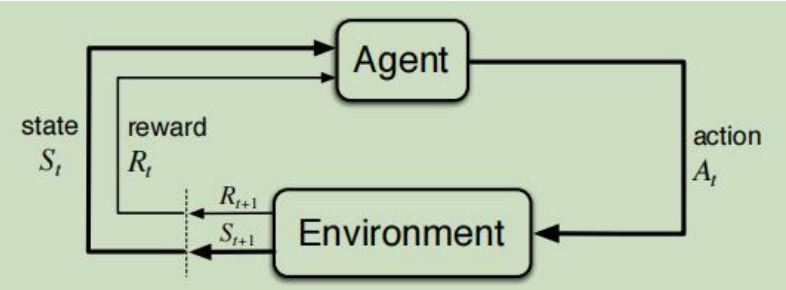


Figure 2 – 6 The agent-environment interaction process in a MDP

In a finite MDP, the sets of actions, states, and rewards contain finite elements. Under this scenario, discrete probability distributions can be adopted to calculate the reward R_t and state S_t. Specifically, for specific values of these random variables $s_{t+1} \in \mathcal{S}$ and $r_t \in \mathcal{R}$ at time step t, there is a probability associated with these values occurring, based on the prior state and action

$$p(s_{t+1}, r_t | s_t, a_t) = \Pr\{S_{t+1} = s_{t+1}, R_t = r_t | S_t = s_t, A_t = a_t\} \tag{2.1}$$

The probability function p indicates the environment dynamics, and defines a probability distribution for each option of s_t and a_t as

$$\sum_{s_{t+1} \in \mathcal{S}} \sum_{r_t \in \mathcal{R}} p(s_{t+1}, r_t | s_t, a_t) = 1 \tag{2.2}$$

The MDP framework is versatile and abstract, allowing it to be applied to a broad range of problems in different ways. For instance, the time steps can represent any sequence of decision-making and action stages, while the actions can range from low-level controls to high-level decisions, among other possibilities. In this thesis, we will use the MDP to construct the maglev system control system for control design.

2.4.1.2 Reward and return

In RL, the purpose of the agent is formalized in terms of rewards received from the environment. At each time step, the reward is a scalar, $R_t \in \mathbb{R}$. The objective can be informally described as maximizing the expected value of the cumulative scalar reward. Generally, we aim to maximize the expected return, G_t, defined as a particular function of the reward sequence. In a general case, the return can be expressed as the reward sum:

$$G_t = \sum_{i=1}^{T-t} R_{t+i} \tag{2.3}$$

where T is the final time step. This approach is logical in applications when a final time step is given, allowing the whole interaction with the environment naturally divided into sub-sequences known as episodes. Each episode concludes with terminal state, after which all states and actions are initialized.

To evaluate the present value of future rewards, an additional concept discounting is adopted. According to this approach, the expected discounted return can be revised into

$$G_t = \sum_{i=0}^{\infty} \gamma^i R_{t+i+1} = R_{t+1} + \gamma G_{t+1} \tag{2.4}$$

where $\gamma \in (0, 1]$ is a parameter named discounted rate.

2.4.1.3 Policies and value functions

Nearly all RL algorithms involve estimating value functions, which evaluate how

advantageous it is for the agent to be in a particular state. Since the rewards an agent can anticipate receiving in the future depend on the actions it chooses, value functions are defined in relation to specific strategies for action, known as policies. A policy is formally defined as a mapping from states to the probabilities of choosing each possible action. For example, if the agent follows a policy π at time t, then $\pi(a_t|s_t)$ is the probability that $A_t = a_t$ if $S_t = s_t$.

The value function of a state s_t under a policy π, denoted as $V_\pi(s_t)$, is the expected return when starting in s_t and following policy π thereafter. The $V_\pi(s_t)$ is named as the state-value function for policy π. For MDPs, it can be formally written as

$$V_\pi(s_t) = \mathbb{E}_\pi[G_t|S_t = s_t] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,|S_t = s_t\right] \tag{2.5}$$

Similarly, the action-value function for policy π, denoted as $Q_\pi(s_t, a_t)$, is the value of the expected return starting from s_t, taking the action a_t, and thereafter following policy π

$$Q_\pi(s_t, a_t) = \mathbb{E}_\pi[G_t|S_t = s_t, A_t = a_t] = \mathbb{E}_\pi\left[\sum_{k=0}^{\infty} \gamma^k R_{t+k+1} \,|S_t = s_t, A_t = a_t\right] \tag{2.6}$$

Solving a RL task means finding a policy that achieves a maximum return over the long run. An optimal policy exists when the policy is better than or equal to all other policies. The optimal policy is denoted by π_*, and the corresponding optimal state-value function is defined as

$$V_*(s_t) = \max_{\pi} V_\pi(s_t) \tag{2.7}$$

The optimal action-value function can also be obtained as

$$Q_*(s_t, a_t) = \mathbb{E}_\pi[R_{t+1} + \gamma V_*(s_{t+1}) | S_t = s_t, A_t = a_t] \tag{2.8}$$

2.4.1.4 Temporal difference learning

When the system model is accessible, system dynamics can be utilized to solve the MDP process. In the absence of a model, RL methods can be referred. These methods are also effective when the model is available. Furthermore, a RL environment can take the form of an MDP, a partially observed MDP, a game, and more.

Temporal difference learning (TD) is a key algorithm used for evaluating value functions Sutton (1988), and has been used in SARSA (Sutton and Barto, 2018), Q-learning (Watkins and Dayan, 1992), etc. TD error is adopted in TD algorithm to update the learned value function $V_\pi(s_t)$. The pseudo code for TD learning in tabular form is presented in Algorithm 1.

Algorithm 1 The pseudo code for TD learning

Input: the policy π to be evaluated

Initialize $V_\pi(s)$ arbitrarily for all states

For each episode **do**

Initialize state s

For each step *t*, if the state s_t is not terminal, **do**

$a_t = \pi(s_t)$

Execute action a_t, observe r_t and s_{t+1}

$V_\pi(s_t) = V_\pi(s_t) + \alpha[r_t + \gamma\, V_\pi(s_{t+1}) - V_\pi(s_t)]$

$s_t = s_{t+1}$

End

End

Output: State-value function $V_\pi(s)$

SARSA is an on-policy control method to find the optimal policy, the pseudo code for SARSA in tabular form is presented in Algorithm 2.

Algorithm 2 The pseudo code for SARSA

Initialize action-value function $Q_\pi(s, a)$ arbitrarily for all states

For each episode **do**

Initialize state s

For each step *t*, if the state s_t is not terminal, **do**

$a_t = \max_a Q_\pi(s_t, a)$

Execute action a_t, observe r_t and s_{t+1}

$a_{t+1} = \max_a Q_\pi(s_{t+1}, a)$

$Q_\pi(s_t, a_t) = Q_\pi(s_t, a_t) + \alpha[r_t + \gamma\, Q_\pi(s_{t+1}, a_{t+1}) - Q_\pi(s_t, a_t)]$

$s_t = s_{t+1},\ a_t = a_{t+1}$

End

End

Q-learning is an off-policy method used to determine the optimal policy, the pseudo code for Q-learning in tabular form is presented in Algorithm 3.

Algorithm 3 The pseudo code for Q-learning

Initialize action-value function $Q_\pi(s, a)$ arbitrarily for all states

For each episode **do**

Initialize state s

For each step *t*, if the state s_t is not terminal, **do**

$a_t = \max_a Q_\pi(s_t, a)$

Execute action a_t, observe r_t and s_{t+1}

$Q_\pi(s_t, a_t) = Q_\pi(s_t, a_t) + \alpha\left[r_t + \gamma \max_a Q_\pi(s_{t+1}, a) - Q_\pi(s_t, a_t)\right]$

$s_t = s_{t+1}$

End

End

2.4.1.5 Value-based learning and policy-based learning

Former algorithms are all in tabular form, value-based learning is a way to generalize the algorithms when the state and action spaces are larger or continuous. The pseudo code for TD learning in continuous form is presented in the Algorithm 4. $\hat{v}(s, \boldsymbol{w})$ is an approximation of value function, where $\boldsymbol{w}$ is the weight vector, and $\nabla\hat{v}(s, \boldsymbol{w})$ is the gradient of the approximate value function.

Algorithm 4 The pseudo code for TD learning

Input: the policy π to be evaluated

Initialize weight vector $\boldsymbol{w}$ arbitrarily

For each episode **do**

Initialize state s

For each step *t*, if the state s_t is not terminal, **do**

$a_t = \pi(s_t)$

Execute action a_t, observe r_t and s_{t+1}

$\boldsymbol{w} = \boldsymbol{w} + \alpha[r_t + \gamma\, \hat{v}(s_{t+1}, \boldsymbol{w}) - \hat{v}(s_t, \boldsymbol{w})]\nabla\hat{v}(s_t, \boldsymbol{w})$

$s_t = s_{t+1}$

End

End

Output: value function $\hat{v}(s, \boldsymbol{w})$

Unlike value-based methods aforementioned, policy-based methods optimize the policy $\pi(a_t|s_t, \boldsymbol{\theta})$ directly, and utilize the gradient ascent on $\mathbb{E}[R_t]$ to update the parameters $\boldsymbol{\theta}$. REINFORCE (Williams, 1992) is a policy gradient method, updating $\boldsymbol{\theta}$ in the direction of $\nabla_{\boldsymbol{\theta}} \log \pi(a_t|s_t, \boldsymbol{\theta}) R_t$. Usually a baseline $b_t(s_t)$ is deleted in the return to decrease the variance of gradient estimate as $\nabla_{\boldsymbol{\theta}} \log \pi(a_t|s_t, \boldsymbol{\theta})(R_t - b_t(s_t))$. The pseudo code for REINFORCE algorithm in the episodic case is presented in Algorithm 5.

Algorithm 5 The pseudo code for REINFORCE algorithm with baseline

Input: policy $\pi(a|s, \boldsymbol{\theta})$, $\hat{v}(s, \boldsymbol{w})$

Initialize state-value parameter $\boldsymbol{w}$ and policy weights $\boldsymbol{\theta}$ arbitrarily

For true **do**

Generate an episode transition $s_0, a_0, r_1, \cdots, s_{T-1}, a_{T-1}, r_T$, following the policy $\pi(a|s, \boldsymbol{\theta})$

For each step $t \in [0, T)$ **do**

Calculate return G_t

Obtain TD error as $\delta = G_t - \hat{v}(s, \boldsymbol{w})$

$\boldsymbol{w} = \boldsymbol{w} + \alpha \delta \nabla_{\boldsymbol{w}} \hat{v}(s_t, \boldsymbol{w})$

$\boldsymbol{\theta} = \boldsymbol{\theta} + \beta\gamma^t \delta \nabla_{\boldsymbol{\theta}} \log \pi(a_t|s_t, \boldsymbol{\theta})$

End

End

Output: policy $\pi(a|s, \boldsymbol{\theta})$

In actor-critic algorithms, the critic updates the parameters of the action-value function, while the actor updates the policy parameters. The pseudo code for one-step actor-critic algorithm in the episodic case is presented in Algorithm 6.

Algorithm 6 The pseudo code for actor-critic algorithm

Input: policy $\pi(a|s, \boldsymbol{\theta})$, $\hat{v}(s, \boldsymbol{w})$

Initialize state-value parameters $\boldsymbol{w}$ and policy weights $\boldsymbol{\theta}$ arbitrarily

For true **do**

Initialize the first state of the episode

I=1

For each step t, if the state s_t is not terminal, **do**

$a_t = \pi(s_t)$

Execute action a_t, observes s_{t+1} and r_t

Obtain TD error as $\delta = r_t + \gamma\hat{v}(s_{t+1}, \boldsymbol{w}) - \hat{v}(s_t, \boldsymbol{w})$

$\boldsymbol{w} = \boldsymbol{w} + \alpha\delta\nabla_{\boldsymbol{w}}\hat{v}(s_t, \boldsymbol{w})$

$\boldsymbol{\theta} = \boldsymbol{\theta} + \beta I \delta \nabla_{\boldsymbol{\theta}} \log \pi(a_t | s_t, \boldsymbol{\theta})$

$I = \gamma I$

$s_t = s_{t+1}$

End

End

Output: policy $\pi(a|s, \boldsymbol{\theta})$

2.4.2 Deep reinforcement learning

In recent years, deep learning (DL) has successfully surpassed traditional machine learning methods in efficiency and accuracy across various applications. Novel RL researches are mainly on the basis of the function approximation of value function $\hat{v}(s, \boldsymbol{w})$, policy function $\pi(a|s, \boldsymbol{\theta})$, and model by deep neural networks (DNN), referred to as deep reinforcement learning (DRL). In machine learning, generalization refers to an algorithm's ability to perform well across a variety of new inputs and applications. DRL is a powerful model-free learning technique that also possesses the property of generality. Consequently, DRL can learn directly from experience samples in both offline and online modes, even without complete knowledge of system dynamics. This makes DRL an efficient method for finding an approximately optimal policy for stochastic nonlinear systems with continuous state and action spaces.

Currently, DRL algorithms are being utilized in robotics to learn optimal control policies directly from visual inputs for various real-world challenges (Polydoros and Nalpantidis, 2017). In this section, some important DRL methods, such as Deep Q Learning (DQN) (Mnih et al., 2013; Mnih et al., 2015), Double DQN (Hasselt et al., 2016), Deep Deterministic Policy Gradient (DDPG) (Lillicrap et al., 2016), etc. will be discussed.

2.4.2.1 Value-based learning

In value-based learning algorithms, an initial random value function is selected for each state, and then updated using learning data. This process is repeated until the optimal value function is determined.

1) Deep Q Learning

The Q-Learning method creates a state-action tabular for the agent to estimate the optimal policy. For continuous problem, TD learning is utilized in Q-Learning. DQN comes up to be a DRL method that uses a DNN to approximate the Q value functions. In DQN, states are input into the DNN as images, and it outputs the estimated Q-values for all permissible actions, enabling easy selection of the optimal action. In detail, the basic idea of Q learning with gradient descent optimization methods are utilized. A replay buffer was also used to store the agent's former experience samples at each time step. Typically, a mini-batch of experience samples is randomly drawn from this memory and used to train RL agents. The pseudo code for

the DQN is presented in Algorithm 7.

Algorithm 7 The pseudo code for DQN

Input: $\hat{q}(s, a;\ \boldsymbol{w})$

Initialize experience replay buffer $\mathcal{B}$ and state-value weights $\boldsymbol{w}$ arbitrarily

For true **do**

Initialize the first state of the episode

For each step t, if the state s_t is not terminal, **do**

Select a random action $a_t \in \mathcal{A}$ in probability p

In probability $(1-p)$, select $a_t = \underset{a}{\operatorname{argmax}}\, \widehat{Q}(s_t, a; \boldsymbol{w})$

Execute action a_t, observes s_{t+1} and r_t

$s_t = s_{t+1}$

Store transition (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$

Sample random mini batch of transitions from the replay buffer $\mathcal{B}$

Obtain TD target as: $q_i = r_i + \gamma \underset{a}{\operatorname{argmax}}\, \widehat{Q}(s_{t+1}, a; \boldsymbol{w})$

Conduct a gradient descent step on $\left(q_i - \widehat{Q}(s_t, a_t; \boldsymbol{w})\right)^2$

End

End

Later in 2015, to avoid the problem of the moving target, a separate DNN was utilized for obtaining the target q_i values as $q_i = r_i + \gamma \underset{a}{\operatorname{argmax}} \widehat{Q}(s_{t+1}, a; \boldsymbol{w}^-)$. The weight parameter $\boldsymbol{w}^-$ was updated using $\boldsymbol{w}$ only after several time steps.

2) Double DQN

Q-learning can encounter issues with the overestimation of Q-values under certain conditions, and DQN faces the same problem. The fundamental concept of Double Q-learning (Hasselt et al., 2016), initially introduced for tabular settings, was integrated with large-scale function approximation to address the overestimation issue in DQN. By separating the max operation in target Q-value estimation into action evaluation and action selection, Double DQN seeks to minimize overestimation. In this approach, the weights of the primary network are used to evaluate the greedy policy, while the weights of the target network are employed to estimate the target value. The pseudo code for the Double DQN is presented in Algorithm 8.

Algorithm 8 The pseudo code for Double DQN

Input: $\hat{q}(s, a;\ \boldsymbol{w})$, target $\hat{q}(s, a;\ \boldsymbol{w}^-)$

Initialize experience replay buffer $\mathcal{B}$ and state-value weights $\boldsymbol{w}$ and $\boldsymbol{w}^-$ arbitrarily

For all episodes **do**

Initialize the first state of the episode

For each step t, if the state s_t is not terminal, **do**

Select $a_t = \underset{a}{\operatorname{argmax}}\, \widehat{Q}(s_t, a; \boldsymbol{w})$

Execute action a_t, observes s_{t+1} and r_t

$s_t = s_{t+1}$

Store (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$

Sample random mini batch of transitions from the replay buffer $\mathcal{B}$

Obtain TD target as: $q_i = r_i + \gamma \underset{a}{\operatorname{argmax}}\, \widehat{Q}(s_{t+1}, a; \boldsymbol{w}^-)$

Perform a gradient descent step on $\left(q_i - \widehat{Q}(s_t, a_t; \boldsymbol{w})\right)^2$

Update target network: $\boldsymbol{w}^- \leftarrow \boldsymbol{w}$

End

End

Besides, Dueling DQN (Wang et al., 2016), Prioritized replay DQN (Schaul et al., 2016), and a single learning algorithm in combination of several DQN enhancements (Hessel et al., 2018) has been presented to improve the overall performance.

2.4.2.2 Policy-based learning

In value-based learning, value functions are utilized to estimate the optimal policy. While in policy-based learning, the optimal behaviour policy is directly estimated without estimating the value functions. Some important policy-based learning algorithms are discussed in this section.

1) Policy gradient methods

In policy gradient (PG) methods, the policy is represented by a parameterized function $\hat{\pi}(a_t|s_t, \boldsymbol{\theta})$, which represents the probability of selecting an action at time t for a specific state s_t, using the policy parameter $\boldsymbol{\theta}$. A Performance measure $J(\boldsymbol{\theta})$ is employed to estimate the optimal policy parameters $\boldsymbol{\theta}$. In PG methods, the policy function $\hat{\pi}(a_t|s_t, \boldsymbol{\theta})$ change smoothly with adjustments to policy parameter $\boldsymbol{\theta}$, unlike in greedy approaches where the policy might change drastically with small changes in estimated values. The policy parameters is updated as: $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \boldsymbol{\alpha}\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}_t)$, where $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}_t)$ is the gradient of the performance measure.

According to the classical REINFORCE method, the performance gradient can be represented as:

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}_t) = \mathbb{E}_{\pi}\left[U_t \frac{\nabla_{\boldsymbol{\theta}}\hat{\pi}(a_t|s_t, \boldsymbol{\theta})}{\hat{\pi}(a_t|s_t, \boldsymbol{\theta})}\right] \tag{2.9}$$

The stochastic gradient ascent algorithm for the policy update using REINFORCE algorithm is as $\boldsymbol{\theta}_{t+1} = \boldsymbol{\theta}_t + \alpha U_t \frac{\nabla_{\boldsymbol{\theta}}\hat{\pi}(a_t|s_t, \boldsymbol{\theta})}{\hat{\pi}(a_t|s_t, \boldsymbol{\theta})}$. However, it is found out that the policy-based learning using REINFORCE learn slowly due to high variance (Baxter and Barlett, 2001). The policy updating theorem can be combined wit arbitrary baseline to reduce the variance.

2) Actor-Critic method

In actor-critic methods, both value functions and policy approximations are made, and the basic actor-critic architecture is as in **Figure 2 – 7**. Using value function

approximation, the policy function can be estimated and updated, while the critic network can be updated using TD methods. The pseudo code for the basic one-step actor-critic for optimal policy estimation is as in Algorithm 9. In this part, several landmark actor-critic algorithms are presented.

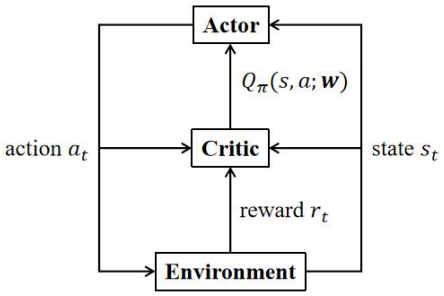


Figure 2 – 7 The actor-critic architecture

Algorithm 9 The pseudo code for actor-critic for optimal policy estimation

Input: Policy function $\hat{\pi}(a|s,\boldsymbol{\theta})$ and state value function $\hat{v}(s;\boldsymbol{w})$, learning rates as α_w and α_θ

Initialize weights $\boldsymbol{w}$ and $\boldsymbol{\theta}$ arbitrarily

For all episodes **do**

Initialize the first state of the episode

For each step t, if the state s_t is not terminal, **do**

Select $a_t = \hat{\pi}(\cdot|s_t,\boldsymbol{\theta})$

Execute action a_t, observes s_{t+1} and r_t

Obtain TD error as: $\delta_t = r_t + \gamma\hat{v}(s_{t+1};\, \boldsymbol{w}) - \hat{v}(s_t;\, \boldsymbol{w})$

Update weights: $\boldsymbol{w} = \boldsymbol{w} + \alpha_w \delta_t \nabla_w \hat{v}(s_t;\, \boldsymbol{w})$, $\boldsymbol{\theta} = \boldsymbol{\theta} + \alpha_{\boldsymbol{\theta}} \delta_t \nabla_{\boldsymbol{\theta}} \ln \hat{\pi}(a_t|s_t, \boldsymbol{\theta})$

$s_t = s_{t+1}$

End

End

A. Deep deterministic policy gradient

DQN has been successful in addressing problems with continuous state and discrete action spaces. However, many real-world control tasks involve continuous action spaces. The DDPG algorithm (Lillicrap et al., 2016) adapts the core concept of DQN to handle continuous action spaces with high-dimensional state spaces. DDPG is a model-free actor-critic approach based on a DPG method. Additionally, the batch normalization technique (Ioffe and Szegedy, 2015), is also used in the DDPG along with the basic idea of the DQN. The pseudo code for the DDPG is given in Algorithm 10.

Algorithm 10 The pseudo code for the DDPG

Initialize:

weights $\boldsymbol{w}$ and $\boldsymbol{\theta}$ for policy function $\hat{\pi}(a|s, \boldsymbol{\theta})$ and action value function

$\hat{q}(s, a;\ \boldsymbol{w})$, and weights $\boldsymbol{w}^-$ and $\boldsymbol{\theta}^-$ for target policy function $\hat{\pi}(a|s, \boldsymbol{\theta}^-)$ and target action value function $\hat{q}(s, a;\ \boldsymbol{w}^-)$

and experience replay buffer $\mathcal{B}$

For all episodes **do**

Initialize the first state of the episode

For each step t, if the state s_t is not terminal, **do**

Select $a_t = \hat{\pi}(\cdot|s_t, \boldsymbol{\theta}) + \mu_t$, where μ_t is exploration noise

Execute action a_t, observes s_{t+1} and r_t

Store (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$

Sample random mini batch of M transitions from the replay buffer $\mathcal{B}$

Obtain TD target as: $y_i = r_i + \gamma\hat{q}(s_{t+1}^i, \hat{\pi}(\cdot|s_{t+1}^i, \boldsymbol{\theta}^-);\ \boldsymbol{w}^-)$

Update critic by minimizing the loss:

$$L = \frac{1}{M}\sum_i (y_i - \hat{q}(s_t^i, a_t^i;\ \boldsymbol{w}))^2$$

Update the actor policy by utilizing the policy gradient:

$$\nabla_{\boldsymbol{\theta}} J = \frac{1}{M}\sum_i \nabla_{\boldsymbol{w}}\hat{q}(s_t^i, a_t^i;\ \boldsymbol{w})\nabla_{\boldsymbol{\theta}}\hat{\pi}(a_t^i|s_t^i, \boldsymbol{\theta})$$

Update target network:

$$\boldsymbol{w}^- \leftarrow \tau\boldsymbol{w} + (1-\tau)\boldsymbol{w}^-$$

$$\boldsymbol{\theta}^{-} \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^{-}$$

$s_t = s_{t+1}$

End

End

B. Twin delayed DDPG (TD3)

To address the accumulation of error and overestimation bias in TD methods within the actor-critic framework, a TD3 algorithm (Fujimoto et al., 2018) was proposed. To reduce variance, TD3 incorporates a regularization strategy that includes a delayed policy update. The pseudo code for the TD3 is given in Algorithm 11.

Algorithm 11 The pseudo code for the TD3

Initialize:

weights $\boldsymbol{w}_1$, $\boldsymbol{w}_2$, and $\boldsymbol{\theta}$ for policy function $\hat{\pi}(a|s, \boldsymbol{\theta})$ and two action value functions $\hat{q}(s, a; \boldsymbol{w}_1)$ and $\hat{q}(s, a; \boldsymbol{w}_2)$, and weights $\boldsymbol{w}_1^{-}$, $\boldsymbol{w}_2^{-}$, and $\boldsymbol{\theta}^{-}$ for target policy function $\hat{\pi}(a|s, \boldsymbol{\theta}^{-})$ and target action value functions $\hat{q}(s, a; \boldsymbol{w}_1^{-})$ and $\hat{q}(s, a; \boldsymbol{w}_2^{-})$

and experience replay buffer $\mathcal{B}$

For all episodes **do**

Initialize the first state of the episode

For each step t, if the state s_t is not terminal, **do**

Select $a_t = \hat{\pi}(\cdot|s_t, \boldsymbol{\theta}) + \mu_t$, where μ_t is exploration noise

Execute action a_t, observes s_{t+1} and r_t

Store (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$

Sample random mini batch of M transitions from the replay buffer $\mathcal{B}$

Obtain TD target as:

$$y_i = r_i + \gamma \min_{j=1,2} \hat{q}(s_{t+1}^i, \hat{\pi}(\cdot|s_{t+1}^i, \boldsymbol{\theta}^-);\ \boldsymbol{w}_j^-)$$

Update critic by minimizing the loss:

$$L = \frac{1}{M}\sum_i (y_i - \hat{q}(s_t^i, a_t^i;\ \boldsymbol{w}))^2$$

Update the actor policy by utilizing the policy gradient:

$$\nabla_{\boldsymbol{\theta}} J = \frac{1}{M}\sum_i \nabla_{\boldsymbol{w}} \hat{q}(s_t^i, a_t^i;\ \boldsymbol{w}) \nabla_{\boldsymbol{\theta}} \hat{\pi}(a_t^i|s_t^i, \boldsymbol{\theta})$$

Update target network:

$$\boldsymbol{w}^- \leftarrow \tau \boldsymbol{w} + (1-\tau)\boldsymbol{w}^-$$

$$\boldsymbol{\theta}^- \leftarrow \tau \boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^-$$

$s_t = s_{t+1}$

End

End

2.4.3 Applications in control

Reinforcement learning (RL) is capable to solve optimal control problems framed as Markov decision processes (MDPs). At each time step, the controller receives a state signal as feedback from the system and responds by taking an action. The objective of the RL optimal control is to maximize the cumulative reward. The challenge is one of sequential decision-making aimed at optimizing long-term performance. The two main advantages to utilize RL for control relies in generality and model-free. The RL algorithms can handle nonlinear and stochastic dynamics and non-quadratic reward functions. Moreover, it does not require system dynamics. Therefore, RL is an exceptionally valuable tool for finding optimal controllers for nonlinear stochastic systems, particularly when the dynamics are unknown or subject to significant uncertainty.

Nowadays, deep RL has been proved to be efficient in continuous tasks like Cart-pole Balancing, Double Inverted Pendulum Balancing, and some locomotion tasks. These applications show that DRL algorithms are effective methods for training deep neural network policies. Most of the control related researches are based on Open AI gyms or the robotics field. The application in field control like structure control and maglev system control is rear. In this work, more advanced algorithms in DRL will be adopted to cope with the nonlinear maglev system control.

CHAPTER 3 TRANSFER LEARNING-BASED DEEP REINFORCEMENT LEARNING FOR ADAPTIVE CONTROL OF MAGLEV TRAINS

As a key component of a maglev train, the magnetic levitation control system ensures that the air gap between the train and its guideway is stable. Although current levitation controllers meet basic engineering requirements, problems related to their performance often arise during long-term operations. These problems can seriously affect the stability and reliability of a maglev system and may lead to partial levitation-point failure. Hence, this part uses transfer learning-based deep reinforcement learning to develop an adaptive nonlinear levitation system controller that enables automatic adjustment of control strategies. First, levitation control based on deep reinforcement learning is mathematically modeled using Markov decision processes, and the nonlinear state space of a single electromagnet levitation control system is established as an agent–environment interaction with the developed deep reinforcement learning controller. Then a twin delayed deep deterministic policy gradient algorithm in an actor–critic framework is adopted to solve the Markov decision processes. To address the dispersion caused by nonlinear levitation control, a transfer learning based two-stage training process is devised that first trains the twin delayed deep deterministic policy gradient networks on a linearized model and then transfers the networks to a nonlinear model. The effectiveness of the new controller is verified by comparing it with a conventional PID controller and an adaptive sliding

mode controller. The robustness of the transfer learning–deep reinforcement learning controller is examined in the presence of uncertainty, such as train load changes and disturbance forces in the levitation system.

3.1 Introduction

Magnetic levitation (Maglev) technology has been implemented in numerous intercity and regional lines across the US, Germany, Japan, South Korea, and China, due to its high speed, low energy consumption and life cycle costs and reduced noise and vibration (Lee et al., 2006, Ni et al., 2021). Maglev techniques are categorized into two types of system based on their mechanism in levitation: EDS and EMS types. The EDS system uses repulsive forces for levitation. Thus, the electromagnetic force of the EDS system is partially stable because the repulsive force increases as the air gap between the vehicle and track decreases. However, the EDS system requires sufficient speed to acquire sufficient induced currents for levitation. In contrast, EMS system can provide an attractive force at zero or low speed to levitate the train. As the electromagnetic forces generated by a constant current are inversely proportional to the square of the levitation air gap (Li et al., 2013), all electromagnetic levitation systems are inherently unstable. Therefore, an active controller is necessary to stabilize a levitation system around the desired levitation air gap. Since the dynamic performance of a magnetic levitation system greatly affects the performance of a maglev train, a levitation control system that maintains a stable air gap is essential.

Maglev control problems (Li et al., 2013; Gottzein et al., 1977; Sinha and Pechev, 1999; Wai and Lee, 2005; Yau, 2009; Huang et al., 2000) have been investigated over the last decade and great progress has been achieved. In general, the linearized model

of an EMS system is established around its equilibrium point through linear control techniques (Ghosh et al., 2014; Hypiusova and Osusky, 2010; Ahmad et al., 2014; MacLeod and Goodall, 1996; Zhang et al., 2019), such as PID and LQR approaches. The equilibrium point is the real root of the system, about which the system would exhibit certain stability properties (Nguyen, 2018). In the maglev system, the equilibrium point is refer to the maglev train suspends on the track with a rating air gap (such as 10 mm) (Lee et al., 2006). However, levitation controllers based on a linearized model are only workable near an equilibrium point where the nonlinear system is linearized. Thus, as a magnetic levitation system is typically nonlinear, controllers based on a linear simplified design may become unstable or even fail when a system is exposed to external disturbance. Hence, various nonlinear levitation controllers (Liu et al., 2009; Kaloust et al., 2004) have been designed to improve the stability of levitation controllers. To ensure that a nonlinear levitation controller has the ability to efficiently cope with external disturbances, advanced levitation controllers based on different theories (Yang et al., 2004; Sinha and Pechev, 1999; Huang et al., 2000; Bidikli, 2000; Yaseen et al., 2022; Adil et al., 2020; Yang et al., 2004; Su et al., 2004) such as robust control methods and adaptive control methods, have been developed to improve the overall performance of control systems. Yang et al. (Yang et al., 2004) designed a robust nonlinear controller with improved position-tracking performance in presence of uncertain model parameters. Sinha and Pechev (Sinha and Pechev, 1999) developed a model-reference adaptive controller for

maglev systems using the stable maximum descent. Huang et al. (Huang et al., 2000) designed an adaptive backstepping controller to improve system stability in the presence of model uncertainty. Bidikli (2000) designed a controller with the capability to adaptively compensate for all parametric uncertainties during the control process. Yaseen et al. (2022) combined the adaptive control and sliding mode control (SMC) approaches to devise three nonlinear levitation controllers, i.e., adaptive terminal SMC, adaptive backstepping SMC and adaptive integral backstepping SMC, to ensure the air gap stayed within a desired range. Adil et al. (2020) developed a supervising and integral backstepping SMC that maintains the air gap at the desired value. In addition, fuzzy control approaches have also been used in maglev levitation control systems. Yang et al. (2004) devised a composite maglev levitation control model that combines the PID algorithm with the fuzzy control method. Su et al. (2004) developed a Takagi–Sugeno (T–S) model-based fuzzy levitation controller for maglev systems by designing a controller using a nonparallel-distributed compensation scheme. Compared with the linear PID controller used in current maglev levitation systems, the above-mentioned levitation controllers show improved stability performance. However, the majority of these control strategies exhibit degraded performance in the presence of disturbances because their performance relies on a precise model and detailed system information that is difficult to obtain in practice. In addition, as each levitation point on a decentralized maglev levitation bogie has its own control loop but all levitation points share the same identical control parameters,

designing control parameters in which all levitation points are highly robust under the worst condition is a significant challenge.

To solve some of the difficult problems in the traditional control field, intelligent control methods (Wai and Lee, 2008; Lin et al., 1998; Shiakolas et al., 2004; Phuah et al., 2005; Fatemimoghadam et al., 2020), such as neural network (NN), convolutional neural network (CNN) and deep belief network (DBN) methods, have been applied to complex nonlinear systems. Intelligent control methods have a degree of robustness to some disturbances and uncertain factors, which can improve overall control performance. However, they exhibit some problems, such as time-consuming during training step, easy falling into the local minimum, and a requirement for auxiliary control and prior information. Recently, reinforcement learning (RL) algorithms (Bellemare et al., 2013; Mnih et al., 2015; Koutnik et al., 2013; Lillicrap et al., 2016; Levine et al., 2016; Mahadevan and Connell, 1992; Watkins and Dayan, 1992) have provided an automated framework for decision-making and control, as they can automatically learn a control policy. They have been used in a wide range of applications, such as in video games (Bellemare et al., 2013; Mnih et al., 2015), autonomous vehicles (Koutnik et al., 2013; Lillicrap et al., 2016) and robotics (Levine et al., 2016; Mahadevan and Connell, 1992). RL was first applied to the optimal control problem by Watkins and Dayan (Watkins and Dayan, 1992), who used Q-learning, which integrates the Bellman equation and a Markov decision process (MDP) into a temporal-difference (TD) learning process. However, this method is

unable to solve high-dimensional problems with drastically increased calculations. The Deep Q Network (DQN) algorithm (Van Hasselt et al., 2016) achieves significant progress by combining advances in deep learning for sensory processing (Krizhevsky et al., 2012) with RL. Moreover, based on DQN, Deep RL (DRL) is capable of decision making and control and can automatically learn a control policy, which makes it a promising approach for solving real-world problems (Luo et al., 2017; Wan et al., 2018; Tan et al., 2019) such as optimal control (Luo et al., 2017), pedestrian regulation (Wan et al., 2018) and traffic grid signal control (Tan et al., 2019) problems.

Researchers have also used the DRL algorithm to solve the linearized maglev levitation control problem (Zhao et al., 2021), and these DRL controllers have less overshoot and more robust than conventional PID and LQR controllers. In DRL, the agent adopts “trial and error” mechanism to explore possible operations based on the current state. However, it is found that numerous invalid explorations and sparse rewards would easily lead to divergence and failure to obtain positive samples (Zheng et al., 2023). We also found that these problems occur when we directly implement the DRL algorithm to solve the nonlinear maglev suspension control problem. An example of the divergence phenomenon is given in **Figure 3–1** (a). The average return of the DRL is supposed to be close to zero during the training process, but it decreases to -9×10^6. In that case, we also found out that the airgap increases instead of approaching to the equilibrium point **Figure 3–1** (b). The recently

developed transfer learning (TL) method offers a potential solution to this problem, as it uses pre-built models developed for specific of datasets (source domain) as a starting point for building new models using a different dataset (target domain) with different attributes (Gupta et al., 2022). The development of deep learning methods has allowed the integration of TL into deep learning algorithms such as CNN (Cirillo et al., 2019), deep support vector (Li et al., 2015), autoencoders (Qi et al., 2017) and long-short-term-memory (LSTM) (Sina et al., 2014), and has improved performance of these methods. Therefore, in this chapter, we combine the TL and DRL methods to solve the nonlinear levitation control problem of maglev systems.

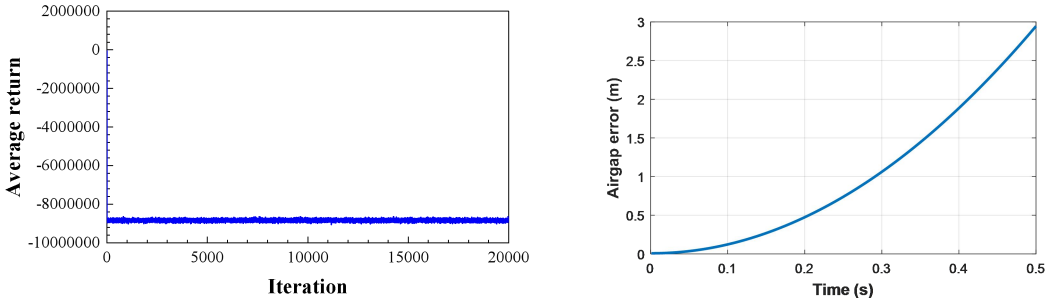


Figure 3–1 Divergence phenomenon when the DRL algorithm is directly used to solve the nonlinear maglev suspension control problem: (a) Average return; (b) Arigap error

This research makes three main contributions. 1) A TL–DRL levitation controller is designed to solve the nonlinear levitation control problem. 2) The performance of the TL–DRL levitation controller is verified by comparing it with the conventional PID controller and the adaptive sliding mode controller (ASMC) (Xu et al., 2018). 3) The robustness of the TL–DRL levitation controller is analyzed by considering

various sources of uncertainty, such as changes in train load during a field test period, load quality changes caused by passengers during operation, track irregularity and disturbances caused by wind or other unpredictable disturbances in a system's operating environment.

The reminder of the section is organized as follows. Section 3.2 presents a simplified nonlinear state space of a single-point levitation control system. The MDP problem of the system is briefly described in Section 3.3. Section 3.4 introduces the DRL-based solution of the MDP problem. The design of a TL–DRL controller is detailed in Section 3.5. The main results, including the performance of the TL–DRL controller, ASMC and PID controller and the robustness of the TL–DRL controller, are provided in Section 3.6. The experiments on a magnetic levitation system and a test platform are presented in Section 3.7. Finally, the conclusions are presented in Section 3.8.

3.2 Nonlinear dynamic modeling of maglev control systems

3.2.1 Overview of maglev control systems

Figure 3–2 shows the configuration of a typical EMS-type maglev system, which consists of maglev vehicles and elevated guideway. Each maglev vehicle has multiple levitation bogies and uses air springs to bear the weight of the vehicle body. Each levitation bogie is equipped with four levitation modules that are the basic levitation functional units of a maglev train. The levitation function is realized via the levitation controller, which controls the levitation air gap between the levitation electromagnet and the F-type rail in each levitation module. The four levitation modules are decoupled by anti-rolling beams to eliminate any coupling influence among the four electromagnets. The mechanical decoupling strategy allows each electromagnet to be controlled independently, and thus the levitation performance of a maglev system relies on the control performance of a single electromagnet. In addition, the mass and stiffness of the guideway on current maglev lines are designed to be very large. For instance, the mass of the guideway in the Changsha maglev line reaches 7,000 kg/m. Hence, the F-type rail on the guideway is generally regarded as a rigid beam. Thus, the levitation control problem for a maglev system can be simplified as a problem consisting of a single electromagnet levitation control system comprising a single electromagnet, a rigid F-type rail and a levitation controller. As shown in **Figure 3–2**, the single-point levitation control model including a vehicle

body, levitation bogie, an electromagnet and a F-type rail. In proper control mode, the levitation electromagnets would generate an adequate attractive force to levitate the vehicle. The principal of the attractive force to adjust the air gap and keep it stable around a target value, e.g., 8 or 10 mm.

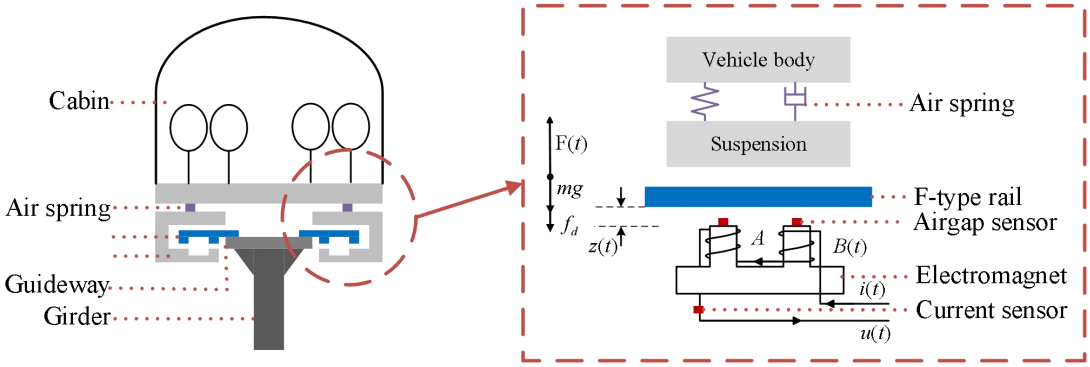


Figure 3–2 Cross-section of an EMS-type maglev system and a schematic of a single EMS module

3.2.2 Nonlinear levitation control model

The physical model of the control problem is a single electromagnet levitation control system. As illustrated in **Figure 3–2**, the levitation of a single electromagnet levitation control system is realized by adjusting the air gap between an inverted F-type rail and an electromagnet. The levitation system of the maglev train performs complex coupling with other elements of the maglev train, thus, it is necessary to make reasonable simplifications before modelling the systems. The assumptions in this chapter are as follows.

1) The impact of the secondary levitation on the dynamic performance of the levitation electromagnet is disregarded, and the secondary electromagnet is considered as a particle that supports the entire mass of the suspended unit.

2) The F-type rail on the guideway is regarded as a rigid beam.

3) The vehicle body and levitation bogie are assumed rigid contact with the electromagnet.

Thus, the single-point control problem can be treated as an electromagnet suspending above the F-type rail. After the electromagnet is energized, the alternating current in the electromagnet stabilizes the of the vehicle with a certain air gap. Based on the Biot–Savart theory and Maxwell equation, when the voltage u is applied to the end of the electromagnet, the electromagnetic force between the electromagnet and rail can be written as

$$f\big(i(t), z(t)\big) = \frac{\mu_0 A_m N_m^2}{4}\left[\frac{i(t)}{z(t)}\right]^2 \tag{3.1}$$

where i is the current, and z is the airgap between the electromagnet and the F-type rail, respectively; A_m is the pole face area; N_m is the number of turns of the magnet coils; and μ_0 is the vacuum permeability. The control voltage equation can be written as

$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = \frac{i(t)}{z(t)}\frac{\mathrm{d}z(t)}{\mathrm{d}t} + \frac{2z(t)}{\mu_0 N_m^2 A_m}(u(t) - i(t)R) \tag{3.2}$$

where $u(t)$ is the excitation voltage, and R is the electromagnet internal resistance.

Using Newton's second law, the governing equation of the single-magnet

suspension system is therefore

$$m\frac{\mathrm{d}^2 z(t)}{\mathrm{d}t^2} = -\frac{\mu_0 A_m N_m^2}{4}\left[\frac{i(t)}{z(t)}\right]^2 + mg + f_d \tag{3.3}$$

where f_d denotes an external disturbance to the suspension system. The dynamic equations of the single-magnet suspension system can be described by combining (3.2) and (3.3). Selecting $\mathbf{x} = [x_1(t), x_2(t), x_3(t)] = [z(t), \dot{z}(t), i(t)]$, the nonlinear state space expression of the single electromagnet suspension system is given by (Sun et al., 2016):

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\dfrac{\mu_0 A_m N_m^2}{4m}\left(\dfrac{x_3(t)}{x_1(t)}\right)^2 + g + \dfrac{1}{m} f_d \\ \dot{x}_3(t) = \dfrac{x_2(t) x_3(t)}{x_1(t)} + \dfrac{2x_1(t)}{\mu_0 A_m N_m^2}(u(t) - x_3(t)R) \\ y(t) = x_1(t) \end{cases} \tag{3.4}$$

Obviously, the suspension system is strongly nonlinear.

As only the current loop is considered in an actual system due to the use of choppers, (4) can be decomposed by deleting the third equation and translating the $x_3(t)$ in the second equation into $i(t)$. The modified nonlinear state space equation of the system equations can be rewritten as:

$$\begin{cases} \dot{\boldsymbol{x}} = f(\boldsymbol{x}) + g(\boldsymbol{x})u \\ y = h(x) \end{cases} \tag{3.5}$$

where

$$f(x) = \begin{bmatrix} x_2(t) \\ g + \dfrac{1}{m} f_d \end{bmatrix}, \quad g(x) = \begin{bmatrix} 0 \\ -\dfrac{\mu_0 A N^2}{4m}\left(\dfrac{1}{x_1(t)}\right)^2 \end{bmatrix}$$

$$h(x) = x_1(t), \;\; u = i^2(t)$$

The state space function is solved using Newton′s method (Kelley, 2003). Then, the established nonlinear single electromagnet suspension system can be used as the environment that interacts with the DRL algorithm in the design of a TL–DRL suspension controller.

3.3 MDP for a levitation control system

3.3.1 Nonlinear levitation control model

Levitation control based on DRL can be mathematically idealized using MDPs, which are a classic formalization of sequential decision making. MDPs learn from interactions between an agent and its environment, as specified in **Figure 3–3**. An MDP has four components: a state space $\mathcal{S}$; an action space $\mathcal{A}$; an immediate/instantaneous reward $\mathcal{R}$; and transition dynamics p that maps a state-action pair at time t into a distribution of states at time $t + 1$.

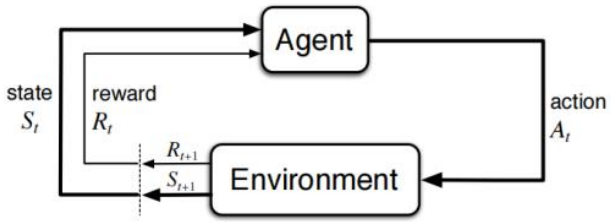


Figure 3–3 Agent–environment interaction in an MDP (Sutton and Barto, 1998)

For each time step t, in a given environmental state $s_t \in \mathcal{S}$, the agent selects action $a_t \in \mathcal{A}$ according to its policy $\pi(\cdot|s_t)$ and then observes a new state, s_{t+1}, and a numerical reward, $r_{t+1} \in \mathcal{R}$. The probability of the reward r_{t+1} and state s_{t+1} can be expressed by the transition dynamics $p(s_{t+1}, r_{t+1}|s_t, a_t)$. The goal of the agent is to find an optimal policy, π^*, by maximizing the total amount of the received reward (return) $U(t) = \sum_{k=0}^{\infty} \gamma^k \cdot R_{t+k}, 0 \leq \gamma \leq 1$,where γ is the discount rate

determining the present value of future rewards. The expected cumulative reward given the state s_t and a_t is determined by the action-value function, defined as

$$Q_\pi(s_t, a_t) = E[U_t|s_t, a_t] \quad s_t \in \mathcal{S}, a_t \in \mathcal{A} \tag{3.6}$$

$Q_\pi(s_t, a_t)$ can be used to evaluate the rewards an agent receives if they pick action a_t under state s_t and policy π . The optimal policy π^* can be found by solving a corresponding optimal action–value function, which is defined as the following maximum action–value function:

$$Q_*(s_t, a_t) = \max_\pi Q_\pi(s_t, a_t) \quad s_t \in \mathcal{S}, a_t \in \mathcal{A} \tag{3.7}$$

Given state s_t, the expected cumulative reward following policy π is defined as the following state–value function:

$$V_\pi(s_t) = E_{A_t \sim \pi(\cdot|s_t)}[Q_\pi(s_t, A_t)] \quad s_t \in \mathcal{S} \tag{3.8}$$

where A_t is a random variable and can be eliminated by calculating the expectation a_t. $V_\pi(s_t)$ can be used to evaluate the rewards for an agent under state s_t given policy π. The optimal policy π^* can then be identified by solving the maximum state value function $V_{\pi^*}(s_t)$.

3.3.2 MDP for a suspension system

The levitation control problem can be described by modeling the four components of an MDP, as **Figure 3–4** shows that the agent–environment interaction in a maglev system takes the form of an MDP. The state of an established MDP should fully capture the system behaviors and be adequate for calculating new states.

As the air gap error represents the difference between the real air gap $x_1(t)$ and objective air gap x_{eq}, and the velocity of the air gap $x_2(t)$ denotes the change in the air gap, the air gap error and the velocity of the air gap are defined as the state, i.e., $s_t = [x_1(t) - x_{eq}, x_2(t)]$. The control current $i(t)$, which is the input in the suspension control system, is regarded as the action a_t.

As the objective of the maglev control problem is to ensure that a stable air gap is established between the electromagnet and the F-type rail and the minimum energy consumption is consumed to do so, the levitation air gap error should be maintained at zero and the input current representing the energy consumption should be as small as possible. Hence, the reward r_t is defined as

$$r_t = -(x_1(t) - x_{eq})^2 - {a_t}^2 \tag{3.9}$$

Then, the levitation control problem is described by solving the MDP problem as an interaction between the DRL (agent) and the established nonlinear single-magnetic levitation system (the environment).

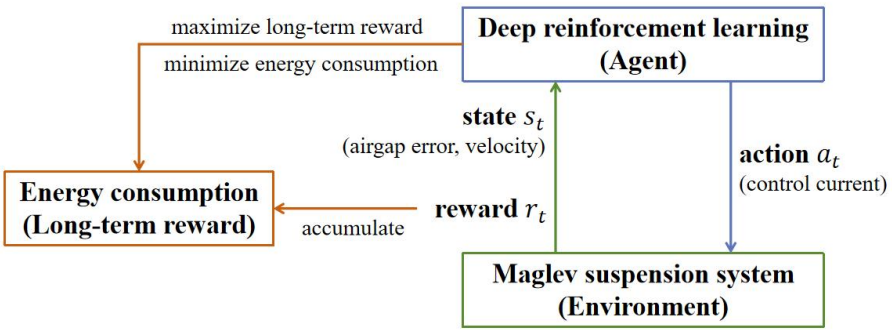


Figure 3–4 Agent–environment interaction in a maglev system control problem

3.4 DRL-based solution of an MDP

3.4.1 Approaches to solving MDPs

The objective of the RL is to find the optimal policy π^* in an MDP by maximizing the expected return from all states. One of two approaches is commonly used to find the optimal policy of the MDPs: the value-based learning approach and the policy-based learning approach (Li, 2017). The value-based learning approach finds the actions by maximizing the cumulative reward in the future for each step, which can be estimated using an optimal action–value function. The agent's decision-making process to find the actions can be expressed as

$$a_t = \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, Q_*(s_t, a) \tag{3.10}$$

In DRL, deep neural networks are trained to approximate the optimal action–value function, namely the value network. The value network can then be trained by using the TD method (Sutton, 1988). The difference between the estimated optimal action–value function value at time *t* and *t*+1 is used to update the network

$$J(\boldsymbol{w}) = \frac{1}{2}\left[Q_*(s_t, a_t; \boldsymbol{w}) - (r_t + \gamma \cdot \max_{a \in A} Q_*(s_{t+1}, a; \boldsymbol{w})\right]^2 \tag{3.11}$$

The object of the value network can be simplified to minimizing the target $J(\boldsymbol{w})$ by updating the parameter $\boldsymbol{w}$ using stochastic gradient descent (SGD) (Arefin and Asadujjaman, 2016). The Adam optimizer (Su et al., 2018) is adopted in this chapter for parameter training.

Policy-based learning calculates the probability of all of the actions given the

optimal policy and then chooses the action based on the calculated probability. For policy-based learning, the target function is the expectation of the state-value function, defined as

$$J(\boldsymbol{\theta}) = E_S[V_\pi(S)] \tag{3.12}$$

where $\boldsymbol{\theta}$ is parameter in policy network, S denotes random variable and can be eliminated by calculating the expectation s_t.

To approximate the optimal policy, a policy network based on deep neural networks is trained by maximizing the target function, and the policy gradient is defined as:

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = E_S\left[E_{A\sim\pi(\cdot|S;\boldsymbol{\theta})}\left[\frac{\partial \ln \pi(A|S;\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \cdot Q_\pi(S,A) + \frac{\partial Q_\pi(S,A)}{\partial \boldsymbol{\theta}}\right]\right] \tag{3.13}$$

$Q_\pi(S,A)$ is normally estimated using the REINFORCE method (Williams, 1992) with the real return or actor–critic methods (Li, 2017) used to obtain a value network for approximation.

The schematic of the actor–critic framework is shown in **Figure 3–5**. The actor (policy) identifies a state from the environment and chooses an action to perform. The actor (policy function) learns by using feedback from the critic (value function). Thus, the actor−critic method trades off the variance reduction of policy gradients with bias introduced by the value function methods (Konda and Tsitsiklis, 2003; Schulman et al., 2016). Thus, the gradient for the actor (policy) network can be expressed as

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \triangleq \boldsymbol{g}(s,a;\boldsymbol{\theta}) = Q_\pi(s,a;\boldsymbol{w}) \cdot \nabla_{\boldsymbol{\theta}} \ln \pi(a|s;\boldsymbol{\theta}) \tag{3.14}$$

where $Q_\pi(s, a; \boldsymbol{w})$ is estimated using the critic (value) network, $g(s, a; \boldsymbol{\theta})$ denotes the gradient. Parameters $\boldsymbol{\theta}$ for the actor network are updated as

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \beta \cdot \boldsymbol{g}(s, a; \boldsymbol{\theta}) \tag{3.15}$$

where β is the learning rate of the actor network.

The gradient for the critic network is calculated using the TD method given in (3.11), and the parameters $\boldsymbol{w}$ are updated as

$$\boldsymbol{w} \leftarrow \boldsymbol{w} - \alpha \cdot \nabla_{\boldsymbol{w}} J(\boldsymbol{w}) \tag{3.16}$$

where α is the learning rates of the critic networks.

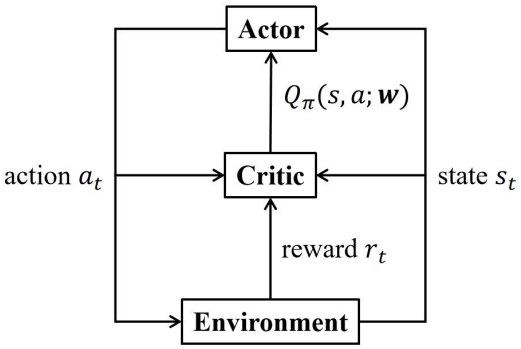


Figure 3–5 Schematic of the actor–critic framework

3.4.2 Twin delayed deep deterministic policy gradient

Generally, the on-policy gradient formulation is used in actor–critic algorithms to update the actor (Peters and Schaal, 2008). Though the on-policy has advantage in improving stability, it would also result in poor sample complexity (Schulman et al., 2017). To increase sample efficiency, algorithms (Donoghue et al., 2017; Gu et al.,

2017) that incorporate off-policy samples and use higher-order variance reduction techniques have been devised. Silver et al. (2014) devised an off-policy actor–critic algorithm named deterministic policy gradients (DPGs) for high-dimensional continuous control problems that performs significantly better than a stochastic policy gradient equivalent. DDPG (Lillicrap et al., 2016), which is an extension of DPG to DRL using a deep variant of the DPG (Silver et al., 2014) algorithm, uses a Q-function estimator to enable off-policy learning and a deterministic actor to maximize this Q-function. To reduce the overestimation bias and accumulating error caused by high variance estimates in the DDPG, Fujimoto et al. (2017) modified the DDPG to form the twin delayed deep deterministic policy gradient algorithm (TD3), which increases the algorithm's stability and performance while considering function approximation error. TD3 is thus an actor–critic algorithm that considers the interplay between function approximation error in both policy and value updates.

In a value-based network, there is overestimation bias caused by bootstrapping, which introduces bias spreading. The overestimation of the TD target is caused by the maximization of the value function $Q_*(s_{t+1}, a; \boldsymbol{w})$. The updating of the Q network is based on the trained Q network, as stated in (3.11). When the $Q_*(s_{t+1}, a; \boldsymbol{w})$ in (3.11) overestimates or underestimates the real Q value, the bias spreads to $Q_*(s_t, a_t; \boldsymbol{w})$ and results in either overestimation or underestimation of the $Q_*(s_t, a_t; \boldsymbol{w})$. In addition, if the greedy target adopted to update the estimated value is sensitive to error ϵ, then the real maximum would be over estimated (Thrun and Schwartz, 1993) (i.e.,

$E_\epsilon\left[\max_{a\in A} Q_*(s_{t+1}, a; \boldsymbol{w})\right] \geq \max_{a\in A} Q_{real}(s_{t+1}, a))$.

In Double Q-learning, two independent value estimates are kept to apart the greedy update from the value function. Then, unbiased estimate of the actions can be attained by selecting the separate estimated values. Hence, in TD3, the Double Q-learning method is adopted to reduce overestimation bias (Fujimoto et al., 2017). The Double Q-learning formulation can be introduced in the actor–critic framework, in the form of an actor a_t and critics $(Q(s, a; \boldsymbol{w}_1), Q(s, a; \boldsymbol{w}_2))$, where the actors are optimized by the critics respectively as:

$$\begin{cases} \eta_1 = r_t + \gamma \cdot Q(s_{t+1}, a_t; \boldsymbol{w}_1) \\ \eta_2 = r_t + \gamma \cdot Q(s_{t+1}, a_t; \boldsymbol{w}_2) \end{cases} \tag{3.17}$$

where η_1 and η_2 are TD target estimated by two critic networks, respectively.

In practice, since the opposite value estimates in target updating and the same replay buffer, the critics are not entirely independent. As a result, for some state *s*, we have $Q(s, a; \boldsymbol{w}_2) > Q\ (s, a; \boldsymbol{w}_1)$, as $Q\ (s, a; \boldsymbol{w}_1)$ commonly overestimates the real value. To solve this problem, TD3 adopts the biased estimate value, $Q\ (s, a; \boldsymbol{w}_1)$, to upper bound the less biased $Q(s, a; \boldsymbol{w}_2)$, so as to create the so called clipped double Q-learning algorithm. Thus, the lowest of the two estimates, is taken as the target update of the proposed algorithm:

$$\eta = r_t + \gamma \cdot \min_{i=1,2} Q(s_{t+1}, a; \boldsymbol{w}_i) \tag{3.18}$$

where η is the TD target estimated using the minimum value of two critic networks.

High variance estimates not only introduce overestimation bias, but also causes a noisy gradient for the actor update. In TD update, a value function is estimated based

on state s_{t+1}, thus errors would accumulate during training. Then, large overestimation bias, and even sub-optimal policy updates would occur. To minimize error in individual updates,get networks are introduced into the TD3, as they are a well-known tool for achieving stability in deep reinforcement learning by providing a stable objective in the learning procedure, thus allowing a good performance in convergence.

To reduce the error accumulated in actor, the actor network should be updated at a lower frequency than the critic, so as to reduce the error introduced. Hence, the TD3 delays actor updates to make sure the value error is small enough. To keep the TD error at a small value, the target networks are updated slowly in TD3, as follows:

$$\boldsymbol{\theta}^{-} \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^{-} \tag{3.19}$$

where $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^{-}$ are parameters in the actor network and target actor network, respectively; τ denotes update parameter.

As the TD3 adopts a deterministic policy that can overfit to narrow peaks in the value estimate, a learning target that employs a deterministic policy is highly susceptible to inaccuracies that are induced by function approximation error, thereby increasing the variance of the target. To solve this problem, target policy smoothing is introduced into TD3 as follows:

$$\eta = r_t + E_{\epsilon}[Q(s_{t+1}, a+\epsilon; \boldsymbol{w})] \tag{3.20}$$

where $\epsilon \sim clip(\mathcal{N}(0,\sigma), -c, c)$ is the random noise added to the estimated actions.

This expectation can be approximated over several actions and performing average operations over mini-batches. Then the target update comes to be

$$\eta = r_t + E_\epsilon[Q(s_{t+1}, a + \epsilon; \boldsymbol{w})] \tag{3.21}$$

The detailed TD3 algorithm in actor–critic form is as given in **Table 3-1**.

Table 3-1 TD3 algorithm in actor-critic form

Pseudo code of TD3 Algorithm
Initialize critic networks $Q(s, a; \boldsymbol{w}_1), Q(s, a; \boldsymbol{w}_2)$, and actor network $\mu(s; \boldsymbol{\theta})$ with random parameters $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{\theta}$.
Initialize target critic networks $Q(s, a; \boldsymbol{w}_1^-), Q(s, a; \boldsymbol{w}_2^-)$, and actor network $\mu(s; \boldsymbol{\theta}^-)$ with parameters $\boldsymbol{w}_1^- \leftarrow \boldsymbol{w}_1, \boldsymbol{w}_2^- \leftarrow \boldsymbol{w}_2, \boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$.
Initialize Replay buffer $\mathcal{B}$.
For iteration =1, 2, ..., T
Select action with exploration noise $a_t = \mu(s_t; \boldsymbol{\theta}) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma)$ and obtain reward $\mathrm{r_t}$ and next state s_{t+1}, then store the transition (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$.
Sample mini-batches of N transitions from the replay buffer $\mathcal{B}$
$a_{j+1}^- = \mu(s_{j+1}; \boldsymbol{\theta}^-) + \epsilon, \epsilon \sim clip(\mathcal{N}(0, \sigma), -c, c)$

$$Q_{1,j+1}^{-} = Q(s_{j+1}, a_{j+1}^{-}; \boldsymbol{w}_1^{-}), Q_{2,j+1}^{-} = Q(s_{j+1}, a_{j+1}^{-}; \boldsymbol{w}_2^{-})$$

$$Q_{1,j} = Q(s_j, a_j; \boldsymbol{w}_1), Q_{2,j} = Q(s_j, a_j; \boldsymbol{w}_2)$$

TD target: $\eta_j = r_j + \gamma \cdot min\{Q_{1,j+1}^{-}, Q_{2,j+1}^{-}\}$

TD error: $\delta_{1,j} = Q_{1,j} - \eta_j, \delta_{2,j} = Q_{2,j} - \eta_j$

Update of critic networks:

$$\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_1 - \alpha \cdot \delta_{1,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_1)$$

$$\boldsymbol{w}_2 \leftarrow \boldsymbol{w}_2 - \alpha \cdot \delta_{2,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_2)$$

If iteration mod k, **then**

Update actor network:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \beta \cdot \nabla_{\boldsymbol{\theta}} \mu(s_j; \boldsymbol{\theta}) \cdot \nabla_a Q(s_j, a_j; \boldsymbol{w}_1)$$

Update target networks:

$$\boldsymbol{\theta}^{-} \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^{-}$$

$$\boldsymbol{w}_1^{-} \leftarrow \tau\boldsymbol{w}_1 + (1-\tau)\boldsymbol{w}_1^{-}$$

$$\boldsymbol{w}_2^{-} \leftarrow \tau\boldsymbol{w}_2 + (1-\tau)\boldsymbol{w}_2^{-}$$

End if

End for

3.5 Levitation controller design

3.5.1 TD3 Controller Learning

Here, a DRL controller is designed using the TD3 method with the objective of maintaining a stable 8 mm air gap between the electromagnet and the F-type rail with the minimum energy consumption. The environment of the designed DRL controller is the nonlinear single electromagnet suspension system established in Section 3.2. In the nonlinear model in this chapter, the initial air gap is 16 mm and the target air gap x_{eq} is 8 mm (Zhao et al., 2021). The value of the parameters in the system model are given in **Table 3-2**.

Table 3-2 Parameter values of the maglev levitation system

Physical quantity	Value	Physical quantity	Value
Mass m / kg	700	Vacuum permeability μ_0 / (Hm^{-1})	$4\pi \cdot 10^{-7}$
Number of Turns of coil N_m	450	Stable current i_{eq}/A	17.0
Area of coil A_m/m^2	0.024	Stable air gap x_{eq} /m	0.008
Coil resistance R/Ω	1.2	-	-

If the nonlinear function is directly adopted as the environment for TD3 training, the results easily disperse. A transfer learning method based on two-stage training is devised to solve this nonlinear suspension control problem. First, a TD3 network is

trained on a linearized model, and then transfer the trained TD3 network for nonlinear model training. If the nonlinear model of the maglev system is linearized at the stable air gap (x_{eq}), then the linearized system can be written as

$$\begin{cases} \dot{x}_1(t) = x_2(t) \\ \dot{x}_2(t) = -\dfrac{\mu_0 A N^2}{4m} \cdot \dfrac{{i_{eq}}^2}{x_{eq}^2} \cdot i(t) + g + \dfrac{1}{m} f_d \end{cases} \tag{3.22}$$

Transfer learning allows users to apply the knowledge and skills gained in previous domains to a novel domain (Deng et al., 2020). The known domain is named the source domain $D_S(x_S)$ and obeys the distribution of $P(x_S)$, while the novel domain is named the target domain $D_T(x_T)$ and has a different distribution $P(x_T)$. x_S and x_T denote the source domain sample and target domain sample, respectively. When the source and target data have the same distribution, the pre-trained network based on source domain can be fine-tuned to the target domain. In this chapter, the source and target data have similar features, as they are both obtained from the same original system. Therefore, the TD3 algorithm pretrained on the source domain (linear system) can be directly fine-tuned on the target domain (nonlinear system). A schematic of the two-stage training strategy used in the transfer learning devised in this chapter is given in **Figure 3–6**.

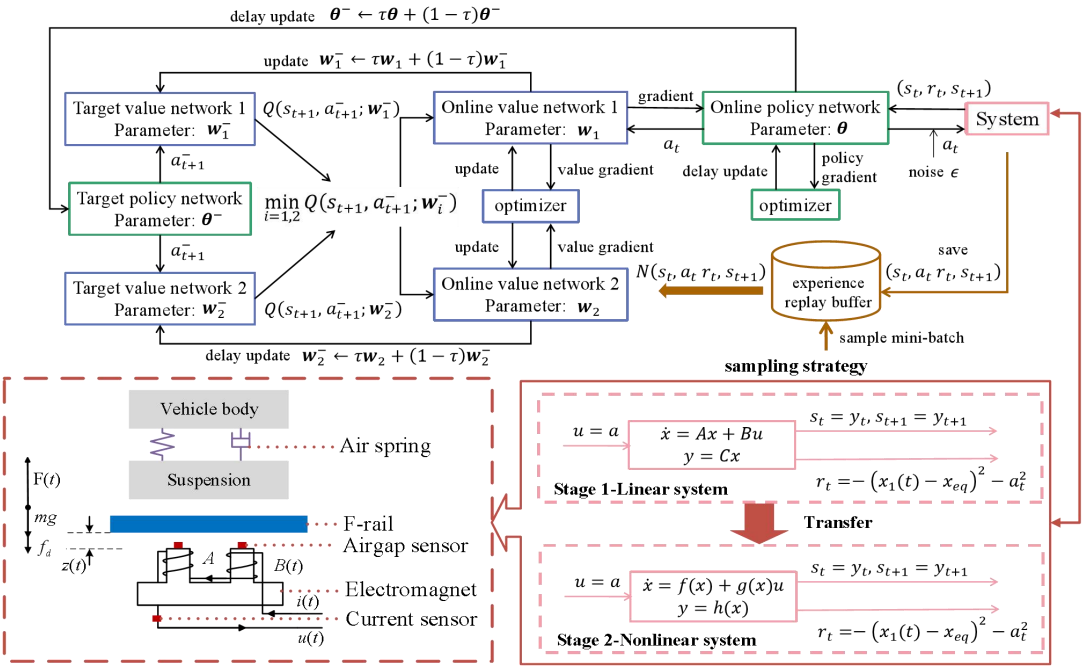


Figure 3–6 Schematic of the two-stage learning strategy

3.5.2 Hyperparameters of TD3 Controller

Both the critic networks and actor networks of the TD3 are designed with two hidden layers. The input for the critic networks are the state and action, and the output is the estimated results of the action-value function. The input for the actor networks is the state, and the output is the action with one dimension. The output of the hidden layers is processed using a rectified linear unit activation function (ReLU), and a tanh unit follows the output of the actor. The iterations of the training for the first and second stages are set as 20,000 and 10,000, respectively. The time step in each iteration is 500 ($\Delta t = 0.001s$). The discounted factor is set to 0.99. The networks are trained every fifth step with a mini batch of 512 transitions sampled from a replay

buffer $\mathcal{B}$ with a size of 1×10^6. The learning rates are as 1×10^{-3} for both the actor and critic networks, and the update parameter τ is 0.005. The policy noise is implemented by adding $\epsilon\sim clip(\mathcal{N}(0, 0.5), -0.2, 0.2)$ to the actions. The critic networks are updated every two iterations. The TD3 algorithm is implemented in in Python 3.6 with PyTorch 1.5.1.

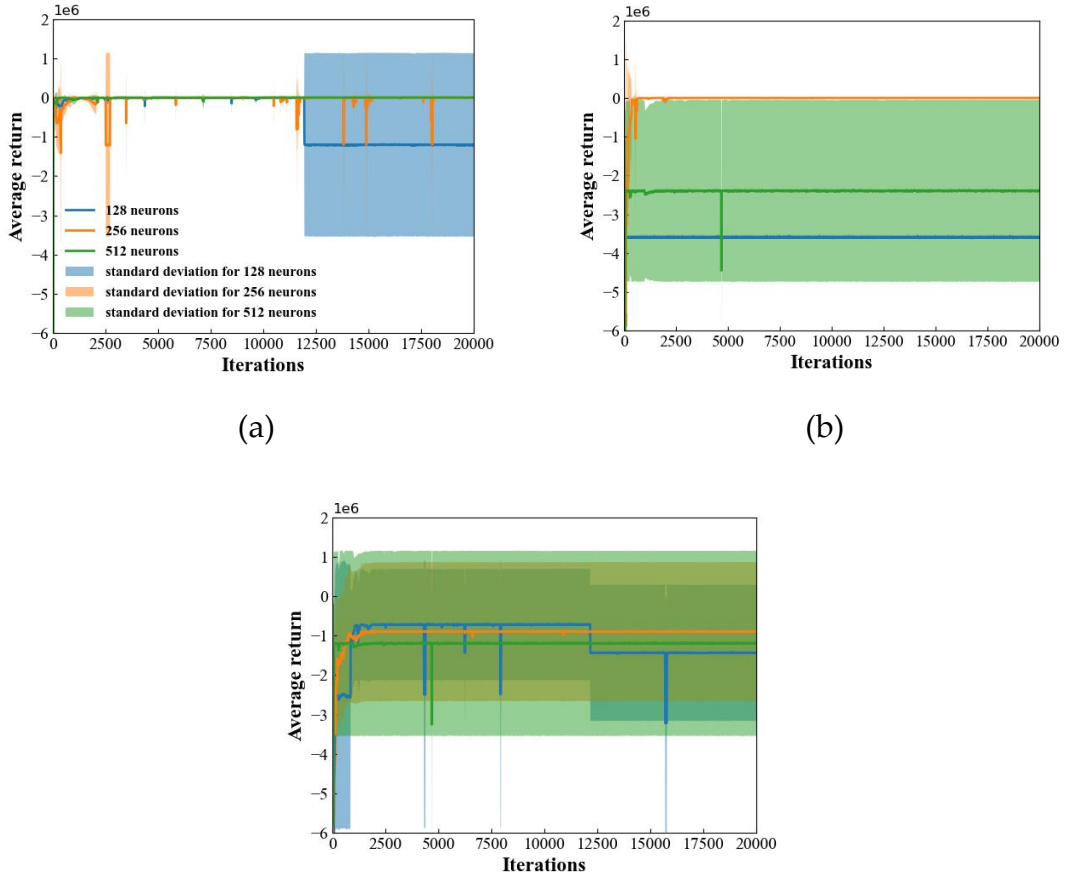


Figure 3–7 Learning curves of different hidden layers and neurons: (a) one hidden layer; (b) two hidden layers; (c) three hidden layers

To obtain the optimal parameters in the hidden layers, nine networks are designed for training using three different numbers of hidden layers (one hidden layer, two hidden layers and three hidden layers) and three different numbers of hidden neurons (128 neurons, 256 neurons and 512 neurons). For each network, five random seeds of the the network initialization and environment are considered in trials. The learning curves of the nine networks based on linearized model (first stage) are shown in **Figure 3–7**. The y-axle in **Figure 3–7** give the average return over the previous 10 evaluations. To show the convergence performance of different networks, the mean average return curve and the standard deviation over five trials for each network are plotted in **Figure 3–7**. As the legends in **Figure 3–7** (b) and (c) are similar to **Figure 3–7** (a), only the legend in **Figure 3–7** (a) is given. It can be observed that for networks with one hidden layer (**Figure 3–7** (a)), the mean value of the average return for the network with 128 neurons decreases and the standard deviation increases (the maximum standard deviation is 2.1×10^6) after 12, 000 iterations, which indicates the unstable convergence performance of the network. In addition, the average return curve of the network with 256 neurons fluctuates much more than that of the network with 512 neurons. For the networks with two hidden layers (**Figure 3–7** (b)), the mean average return values of the networks with 128 neurons and 512 neurons are much smaller than those of the networks with 256 neurons, and the networks with 512 neurons have larger fluctuations in convergence performance with a maximum standard deviation of 2.1×10^6. **Figure 3–7** (c) shows the average

return curves of the networks with three hidden layers. As can be seen, the average return of the networks with 512 neurons are much smaller than those of the other two curves but the standard deviation of the networks with 512 neurons is the largest (with a maximum standard deviation value of 2.6×10^6). In addition, the average return values of the network with 128 neurons becomes small after 12,000 iterations. It can be concluded from the average return curves that the optimal number of neurons when there is one hidden layer is 512 neurons, but the optimal number of neurons when there are the two or three hidden layers is 256 neurons.

To further investigate the optimal structure for a DRL controller, the control performance of the three networks with three different layers is examined in **Figure 3–8**. It can be observed that the convergence times of the networks with two or three hidden layers are shorter than the convergence time of the network one hidden layer. Additionally, the network with two hidden layers exhibits less fluctuation than the network with three hidden layers. Hence, two hidden layers with 256 neurons are adopted in the critic networks and actor networks of the TD3 in this chapter. The detailed structure for the final optimal TL–DRL controller based on the TD3 method is shown in **Figure 3–9**.

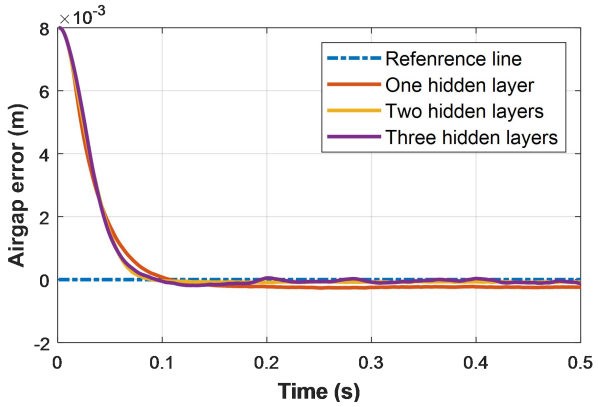


Figure 3–8 Control performance of three kinds of network architecture

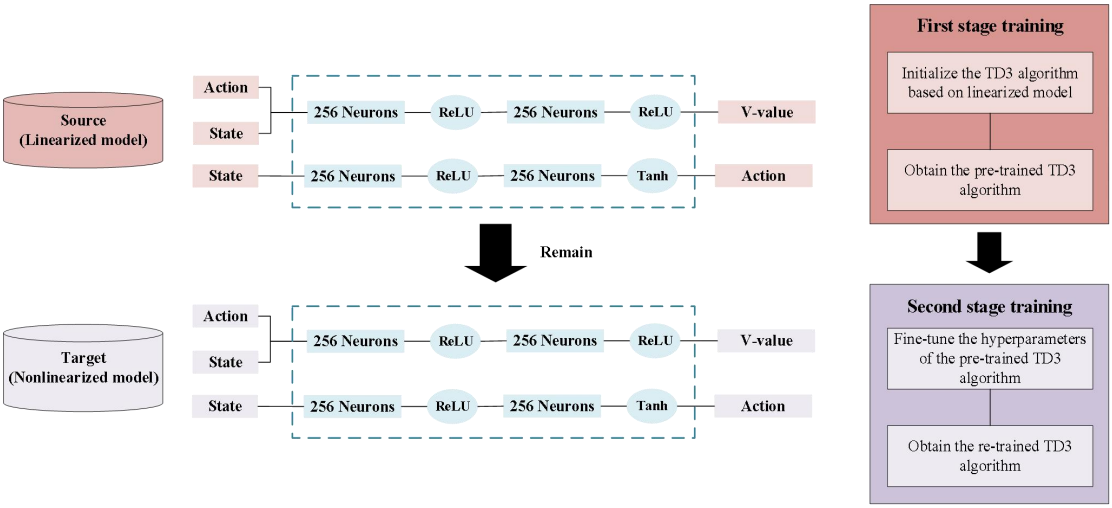


Figure 3–9 Structure of the transfer learning based TD3 controller

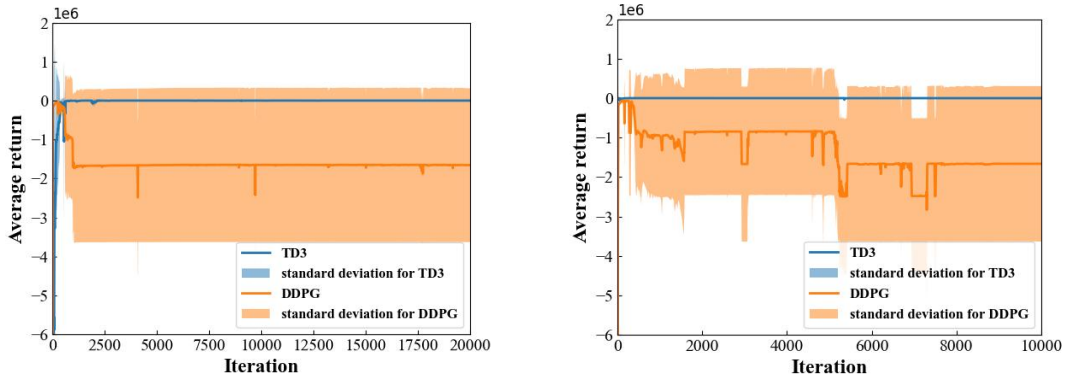
3.6 Simulation results

3.6.1 Effectiveness of the established TD3 controller

3.6.1.1 Network performance comparison with DDPG algorithm

The TD3 algorithm (Fujimoto et al., 2017) significantly improves the performance of the DDPG algorithm by adopting a clipped variant of Double Q-learning, which reduces overestimation bias. To verify the effectiveness of the TL–DRL controller based on the TD3 algorithm, a DDPG controller is designed for comparison. In the DDPG controller, the actor network has one hidden layer with 400 neurons and the critic network has one hidden layer with 100 neurons. The outputs of the hidden layers are processed using a ReLU, and a tanh follows the the actor output. The inputs to the critic network are the action and state. Both network parameters are updated using an Adam optimizer with learning rate of 1×10^{-3}.

The convergence performance at the first and second stages of training with the TD3 is compared with this performance with training with the DDPG. The learning curves, including the mean average return values and standard deviations at both training stages, are given in **Figure 3–10**. The mean average return curve of the DDPG declines after 200 iterations in both the first and second stages, and the standard deviation of the DDPG is much larger than that of the TD3 in both stages. These results indicate that the performance of the TD3 is more stable than that of the DDPG in both stages.



(a) (b)

Figure 3–10 Learning curves of TD3 and DDPG: (a) Stage 1: Linear model based training; (b) Stage 2: Nonlinear model based training

3.6.1.2 Controller performance comparison with PID and ASMC

After two-stage training, the best neural network parameters of the TD3 during training is set as the parameter of the TL–TD3 controller. The control performance of the trained TL–TD3 is then compared with a PID controller and ASMC, as the PID is commonly used in real maglev systems and the ASMC is adopted to verify the adaptive control performance of the proposed TL–DRL controller. In this chapter, the PID controller is designed using the linearized model. The transfer function of the PID controller is

$$G_p(s) = K_p + K_d s + K_i \frac{1}{s} \tag{3.23}$$

where K_p, K_d and K_i are proportional-control gain, differential-control gain, and integral-control gain, respectively. We set $K_p = 2{,}200$, $K_d = 8{,}000$ and $K_i = 6$ in this section based on trial and error. The control signal u (excitation voltage) of the adaptive sliding mode control law proposed by Xu et al. (2018) is presented as:

$$u = \frac{\mu_0 mN}{2B(t)}\left(\frac{4Rx_1B(t)^2}{\mu_0^2 mN^2} + \lambda_2\left(g - \frac{A_m B(t)^2}{\mu_0 m}\right) + \lambda_1 e_2 + \hat{w}sign(S_{smc})\right) \quad (3.24)$$

where $B(t)$ is the air gap flux density, the meaning of other physical quantities are the same as the maglev suspension model in this paper, λ_1 and λ_2 are positive constants, $\hat{w}$ is the adjustable estimated value of w, sign($\cdot$) is the signum function, and S_{smc} denotes the dynamic sliding surface defined by

$$S_{smc} = \lambda_1 e_1 + \lambda_2 e_2 + g - \frac{A_m B(t)^2}{\mu_0 m} \quad (3.25)$$

where $e_1 = x_1 - x_{eq}$, and $e_2 = x_2 - \dot{x}_{eq} = x_2$. The adaptive laws of $\hat{w}$ is proposed as $\dot{\hat{w}} = \frac{1}{\delta}|S_{smc}|$, and the $\delta \in \mathbb{R}^+$ denotes the adaptive gain. According to Xu et al. (2018), the coefficients of the ASMC are $\lambda_1 = 1200$, $\lambda_2 = 200$, and $\delta = 0.015$ for the simulation.

A comparison of the airgap errors by using the PID controller, ASMC, and TL–DRL controller based on TD3 are shown in **Figure 3–11**. As can be seen, the air gap achieves the desired level using the designed PID controller, ASMC, and TL–DRL controller. The convergent time to control the suspension airgap within the desired air gap is approximately 1.5 s for the PID, 1 s for the ASMC, and is only 0.1 s for the TL–DRL. The PID controller, being a linear control strategy, may not adapt well to the inherently nonlinear dynamics of the maglev system. In contrast, both ASMC and TL–DRL are better suited to handle nonlinearities, resulting in faster and more robust convergence. In addition, the observed air gap fluctuation is obviously observed larger for the PID controller than for the TL–DRL controller based on the TD3 algorithm, while the overshoot of the ASMC is observed larger than the

TL–DRL controller. The fluctuation and overshoot of the PID and ASMC is due to the coefficients chosen and characteristics of control methods. These results indicate that the accuracy and effectiveness of the TL–DRL controller designed in this chapter are much better than those of the PID controller and the ASMC.

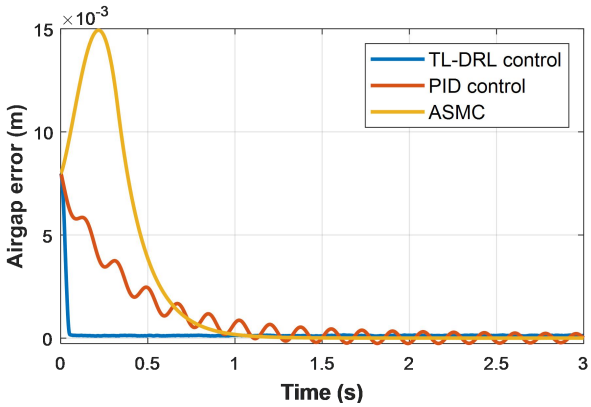


Figure 3–11 Control performance of ASMC, PID and TL–DRL controllers

3.6.2 Robustness of TL-DRL Controller

When a maglev train is in operation, the performance of its suspension controller tends to be affected by changes in train load, track irregularity and disturbances caused by wind force or other unpredictable disturbances in the operating environment. Hence, to verify the robustness of the designed TL–DRL controller, its performance under these three conditions, i.e., fluctuation of train load, track irregularity and disturbance force, is analyzed.

3.6.2.1 Effect of different train load

In field test, four train load operating conditions are considered in evaluations of suspension controller: AW0 (the weight of an empty vehicle), AW1 (the weight of an empty vehicle + the weight of seated passengers), AW2 (the weight of an empty vehicle + the weight of seated passengers + the weight of standees with five people/m2), and AW3 (the weight of an empty vehicle + the weight of seated passengers + the weight of standees with eight people/m2) (DBJ50T-347-2020). To simulate the different train load conditions, this chapter sets the masses for AW1, AW2 and AW3 as 115%, 130% and 145% of AW0, respectively. The airgap errors obtained using the TL–DRL, ASMC and PID controllers under the four train loads are given in **Figure 3–12**. **Figure 3–12** (a) and (c) show that the air gap error of the TL–DRL controller slightly increases as the train load increases, while the air gap error of the ASMC slightly decreases. After checking the control signal and acceleration of the maglev system using ASMC with different train load, it is found that the initial acceleration value decreases when the train load increases. The oscillation of the air gap error of the PID controller in **Figure 3–12** (b) increases with train load. As the coefficients of the PID is obtained with AW0 train load, the controller may not work well when system changes. These results indicate that the TL–DRL controller and ASMC is more robust to changes in train load than the PID controller. However, the overshoot of the ASMC can not be ignored.

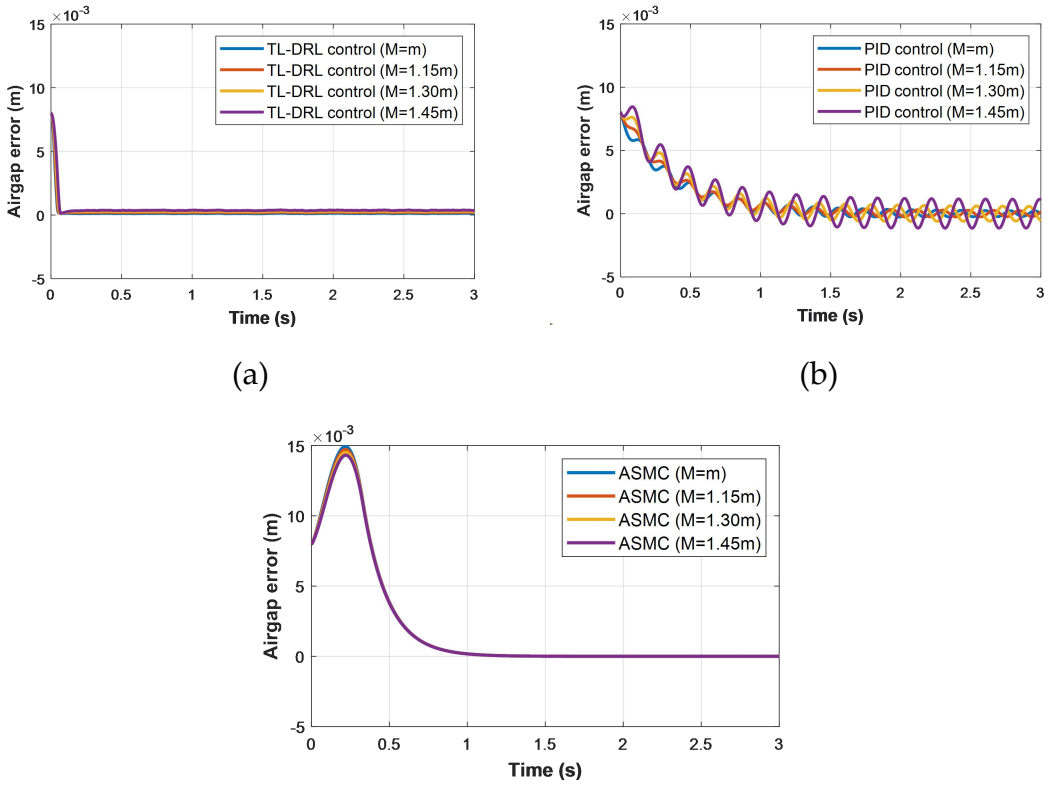
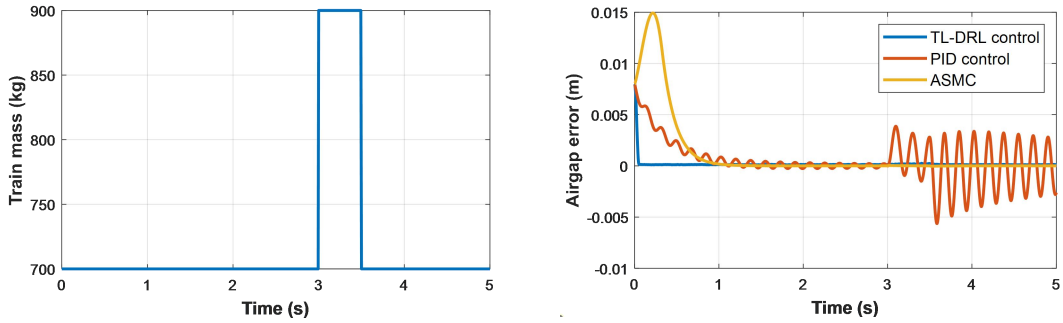


Figure 3–12 Comparison of performance under different load conditions: (a) TL–DRL controller; (b) PID controller; and (c) ASMC controller

3.6.2.2 Fluctuation of train load

During operation, changes in the number of passengers may cause sudden changes in the train load. To ensure the comfort and safety of a maglev train, a robust levitation system is required to adapt to disturbances caused by these sudden changes. To further verify the robustness of the TL–DRL controller based on TD3, this study simulates sudden changes in the train mass from 700 kg to 900 kg at 3 s that last for

0.5 s (see **Figure 3–13** (a)). The air gap error of the suspension systems with TL–DRL, ASMC and PID controllers, respectively, are shown in **Figure 3–13** (b). As can be seen, when the train mass increases at 3 s, there is only a slight oscillation in the air gap error in the system using the TL–DRL controller and ASMC. However, there is a huge oscillation in the air gap error of the system using PID controller and this oscillation continues for more than 5 s. This confirms that the adaptive controllers as TL–DRL controller and ASMC are much more stable than the PID controller to fluctuations in train load.



(a) (b)

Figure 3–13 Comparison of performance under changes in load: (a) mass change curve; (b) controller performance

3.6.2.3 Fluctuation of track irregularity

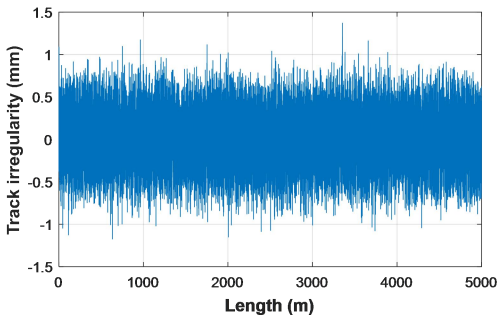
Track irregularity is the main source of excitation in maglev control systems. It can cause a substantial instability in a levitation controller, as it leads directly to fluctuations in the levitation air gap. To evaluate the effects of the random nature and

characteristics of rail irregularity on the designed TL–DRL controller, the following power spectrum density function is adopted to simulate the vertical profile of variations in the guideway geometry (Yang et al., 2004):

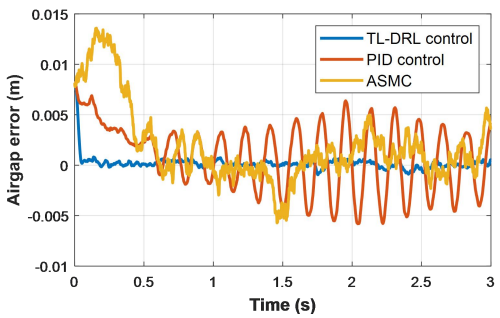
$$S(\Omega) = \frac{A_v \Omega_c^2}{(\Omega^2+\Omega_r^2)(\Omega^2+\Omega_c^2)} \tag{3.26}$$

where Ω is the spatial frequency, A_v is set as 1.5×10^{-7} m and Ω_c is 0.825 rad/m. The vertical profile of the track irregularity calculated using (3.26) is given in **Figure 3–14** (a).

The air gap errors associated with the TL–DRL, ASMC and PID controllers are shown in **Figure 3–14** (b), which shows the oscillation in the air gap error for the three controllers. The maximum air gap error for TL–DRL controller is 1.3 mm, which is less than the allowed maximum gap error of 2 mm. The maximum air gap error for the PID controller is 6 mm, and 5 mm for the ASMC. Thus, the TL–DRL controller can maintain an air gap around 8 mm under track irregularity, but the ASMC and PID controller are unable to do so.



(a)



(b)

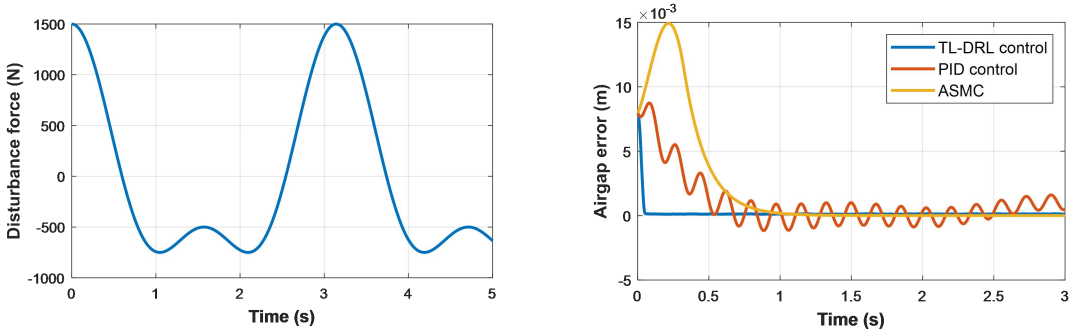
Figure 3–14 Comparison of the performance under track irregularity: (a) vertical profile of the track irregularity; and (b) controller performance

3.6.2.4 Fluctuation of disturbance force

To analyze the effect of disturbance forces on the controllers' performance, multi-amplitude and multi-period sine curves (Chen et al., 2020) are used to simulate disturbance forces, as follows:

$$f_d = 1000sin(2t + \frac{\pi}{2}) + 500sin(4t + \frac{\pi}{2}) \quad (3.27)$$

The disturbance forces and the control performance of the two controllers are shown in **Figure 3–15**. **Figure 3–15**(b) shows that the PID controller is greatly affected by the disturbance force, whereas the TL–DRL controller and ASMC are near-negligibly affected.



(a) (b)

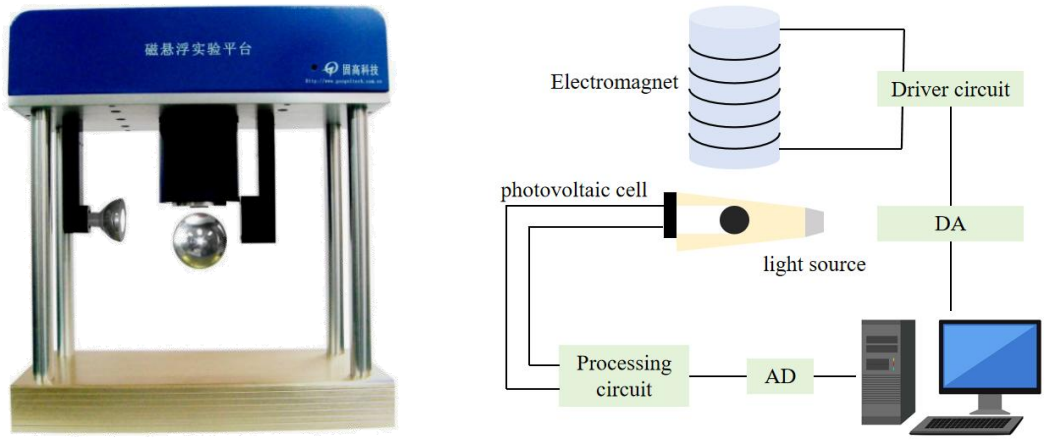
Figure 3–15 Comparison of the performance under disturbance forces: (a) disturbance force added; and (b) controller performance

3.7 Experiment results

To further verify the effectiveness of the proposed TL–DRL controller, experiments on a GML1001 magnetic levitation system is conducted.

3.7.1 Experiment setting

A GML1001 magnetic levitation system is adopted in the experiment. The GML1001 system is mainly composed of three parts as a main body, an electronic control box, and a data acquisition card. The main body of the magnetic levitation system and the diagram of the system structure are shown in **Figure 3–16**.



(a) (b)

Figure 3–16 The magnetic levitation system: (a) the main body, (b) the diagram of the system structure

During operation, the system sends analog voltage to the electromagnet using the data acquisition card installed in the PC through Matlab. Then the electromagnetic force can be generated, and the displacement of the ball is detected by the

photovoltaic cell. The close loop control is thus constructed to generate adequate electromagnetic force to levitate the ball to the reference place. The whole control system takes the current in the electromagnet as control input so as to control the electromagnetic force. The physical parameters of the system is as in **Table 3-3**.

Table 3-3 Parameter values of the GML1001 maglev levitation system

Physical quantity	Value	Physical quantity	Value
Mass	100 g	Reference air gap	35 mm
Core diameter	22 mm	Wire diameter	0.8 mm
Reluctance	13.8 Ω	Radius of the ball	22 mm
Number of Turns of coil	2450	K^*	6.146e-4 $\mathrm{Nm^2/A^2}$

3.7.2 Experiment results and discussion

In this part, some experimental results of the GML1001 magnetic levitation system are provided to verify the practical control performance of the designed TL–DRL controller. The proposed controller is firstly compared with the PID controller. The coefficient parameters of the PID control are set as $k_p = 4.5, k_i = 0.01, k_d = 50$. In **Figure 3–17**, the control performance of the PID and TL–DRL controllers are displayed. Notably, the TL–DRL method achieves the target air gap in just 0.6 s, a significant improvement over the PID controller, which requires 2.5s to reach the same target air gap. Furthermore, the TL–DRL controller exhibits a smaller

overshoot compared to the PID controller in this scenario.

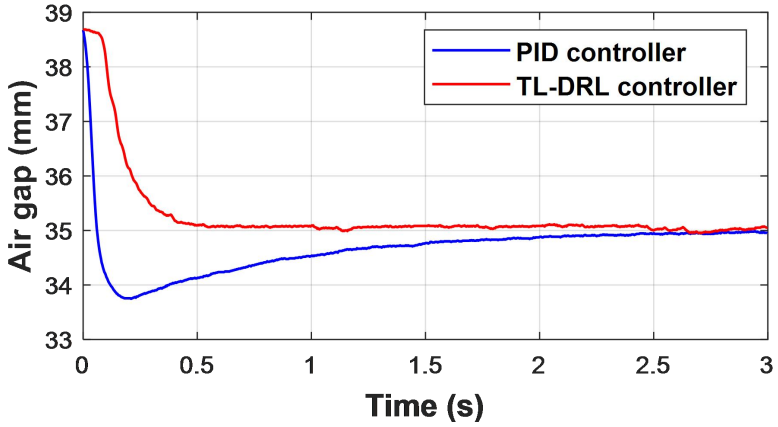


Figure 3–17 Control curves of PID and TL–DRL controllers

In order to assess the robustness of the TL–DRL controller, track irregularity is considered in the experiment. In this experimental setup, a random value between -0.1 mm and 0.1 mm is introduced to the measured air gap to simulate the track irregularity. The control curves of the PID and TL–DRL controllers are as in **Figure 3–18.** It can be observed that the TL–DRL control method achieves the target air gap more swiftly compared to the PID controller. Additionally, the TL–DRL control demonstrates a smaller overshoot than the PID controller. These outcomes underscore the superior efficiency and robustness of the TL–DRL control method over PID control.

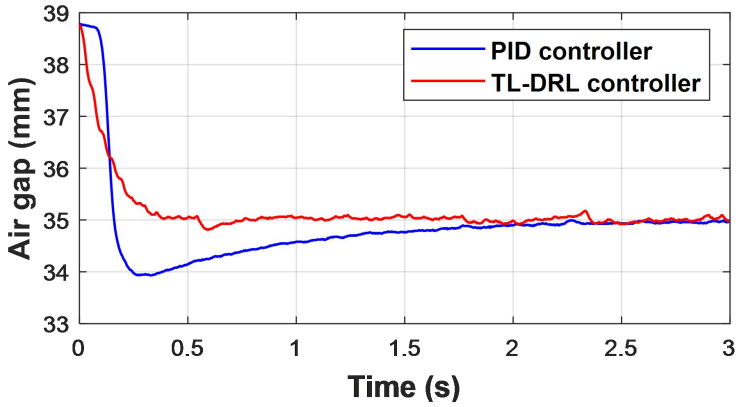


Figure 3–18 Control curves of PID and TL–DRL controllers under track irregularity

3.8 Conclusion

A TL–DRL controller based on a TD3 algorithm is devised for solving a nonlinear control problem for maglev levitation control. With an appropriate reward and state design, the TD3 algorithm can provide the optimal control solution for a maglev system. In addition, the two-stage training method based on TL can effectively solve the nonlinear control problem. The effectiveness of the TL–DRL controller is verified by comparing its performance with that of a DRL controller based on the DDPG algorithm and that of a traditional PID controller and an advanced adaptive controller named ASMC controller. In addition, the robustness of the TL–DRL controller is assessed by evaluating its response to disturbances caused by changes in train load, rail irregularity and disturbance forces. The experiment has also been conducted to verify the performance of the TL–DRL controller. The main results are as follows.

1)The convergence performance of the TL–DRL controller based on a TD3 algorithm is more stable than that of the DDPG algorithm in both training stages.

2)The convergent time of the TL–DRL controller is 0.1 s, whereas that of the ASMC is ~1s, and the conventional PID controller is ~ 1.5s; the maximum overshoot of the TL‐DRL controller is also smaller than that of the PID controller and ASMC controller.

3)The TL–DRL controller based on a TD3 algorithm is more robust than the

traditional PID controller and ASMC when the system is disturbed by changes in train load, track irregularity or disturbance forces through simulation.

4) From the experiment result, the TL–DRL controller has better control performance and robustness to the track irregularity compared with the PID controller.

Future work studies could consider the following issues to improve the TL–DRL method: data processing speed and data storage; multi-point suspension cooperative modeling and control using multi-agent RL; and experimental verification of the DRL method.

CHAPTER 4 COOPERATIVE CONTROL VIA HAMILTON-JACOBI-BELLMAN MULTI-AGENT DEEP REINFORCEMENT LEARNING

Currently, the levitation controller is designed based on a single-point model as presented in Chapter 3 in which the coupling effect between two suspension units at one side of the bogie in the maglev train is ignored. However, it is found that the coupling effect can cause unstable levitation problems during long-term operations. Hence, to solve the coupling problem of the two suspension units, a cooperative levitation controller based on the Hamilton-Jacobi-Bellman incorporated multi-agent reinforcement learning (HJB–MADRL) is proposed in this chapter. The MADRL is adopted for the two-point levitation control considering the coupling effect between the two levitation points. To improve the training of the value network in the MADRL, the HJB function is used in control theory to evaluate the optimality of the value function. The proposed algorithm shows an improved performance compared to the original MADRL algorithm. The effectiveness of the proposed cooperative controller using the proposed algorithm is verified by comparing with a conventional PID controller and a model-guided controller. The robustness of the HJB–MADRL controller is examined in the presence of pitch motion, change in train load, disturbance force, and track irregularity. Additionally, experiment on a full-scale maglev bogie is carried out to validate the performance of the HJB–MADRL

controller.

4.1 Introduction

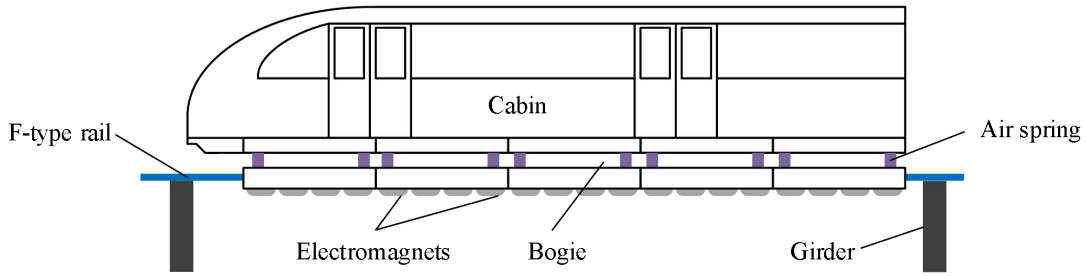
As mentioned in Chapter 3, the levitation function in EMS systems relies on the magnetic attraction force between the guideway and electromagnets. However, this type of levitation is inherently unstable due to the characteristics of the magnetic circuit and external disturbances. Therefore, the design of the control system is crucial to maintaining a stable suspension gap between the maglev train and the guideway. There are two major challenges in controlling the maglev levitation system:

a) Precise modelling: The EMS-type maglev train exhibits highly nonlinear behavior due to the complex interactions between its magnetic field, quality of tracks, and driving conditions. Obtaining an accurate model for the system is challenging due to these nonlinear behaviors.

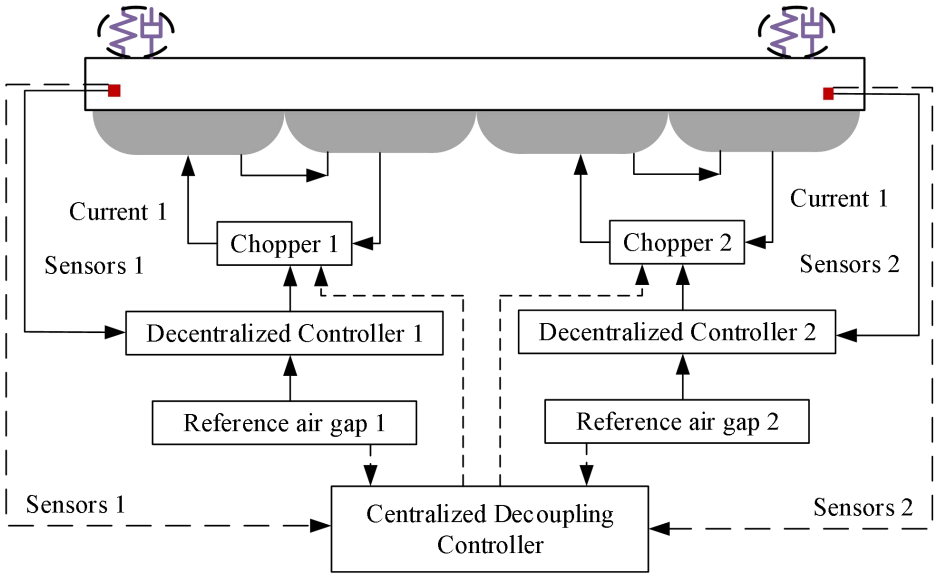
b) Robustness against external disturbances: The maglev system is subject to various external disturbances during operation, including load changes caused by passengers during operation, track irregularity, and disturbances caused by wind or other unpredictable disturbances in the operating environment. The control method needs to be robust enough to handle these disturbances and maintain stability of the suspension system.

The structure of the control system in EMS-type maglev suspension systems can have limitations on control performance (Chen et al., 2018). **Figure 4–1** (a) illustrates the schematic diagram of the current EMS-type maglev levitation system, while

Figure 4–1 (b) presents the centralized and decentralized control structures. It can be observed from **Figure 4–1** that the car body is supported by five bogies, and each of the bogie is consist of two levitation modules. Each module contains two pairs of adjacent electromagnets controlled by decentralized controllers. Specifically in the decentralized control structure, each levitation point on a decentralized maglev levitation bogie has its control loop but all levitation points share the same control parameters. Designing control parameters for all levitation points to ensure robustness of all the levitation points under worst-case condition poses a significant challenge. In current research (Xia et al., 2020; Li and Shen, 2020; Dalwadi et al., 2021), the maglev suspension bogie is typically decomposed into four single levitation units. However, static experiments conducted on the CMS-04 low-speed maglev train (He et al., 2013) have demonstrated that the coupling effect between the two levitation units at one side of the bogie cannot be ignored. Therefore, there is a need to develop a multi-point levitation control method with model uncertainty and control robustness guaranteed.



(a)



(b)

Figure 4–1 Schematic diagram and control structure of EMS-type maglev train: (a) Schematic diagram of EMS-type maglev train, (b) Decentralized and centralized control system

Decoupling control is a commonly employed approach to address the control problem in multi-input-multi-output (MIMO) systems. It involves designing a decoupler that treats the MIMO control system as multiple single-input-single-output (SISO) loops, effectively mitigating the complex couplings between control loops. Several decoupling control strategies have been developed, including static decoupling (Lee et al., 2005) and inverted decoupling (Shinskey, 1996; Wade, 1997). He et al. (2013) proposed a modified adjoint transfer matrix-based decoupler for

MIMO systems, which simplifies the computation involved. Leng et al. (2019) applied a feedback linearization method to design a decoupling controller that reduces the coupling between the two levitation units and minimizes gap fluctuation when encountering track steps. Moreover, extended state observer (ESO) and characteristic modeling have been utilized in addressing MIMO control system problems (Chen et al., 2018; Xu et al., 2020). Chen et al. (2018) introduced a model-guided data-driven decentralized control approach for maglev levitation systems. This method incorporates two extended states, i.e., unmodeled uncertainties and external disturbances, into the state space functions. An ESO is then employed to estimate the unknown part, including uncertainties, disturbances, and coupling between the two levitation units. Xu et al. (2020) proposed a characteristic modeling method, treating the coupling disturbance of the levitation units as an environmental variable. However, the aforementioned methods simplify and decouple the two-point levitation system into a single-point suspension system, disregarding the coupling effect between the two suspension points. Additionally, ensuring coordinated functionality between the two levitation points under external disturbances is challenging. To address these issues and simultaneously control both levitation units in the system, an advanced method called Hamilton-Jacobi-Bellman multi-agent deep reinforcement learning (HJB–MADRL) is proposed in this chapter.

Multi-agent deep reinforcement learning (MADRL) is related to situations that require multiple agents to communicate and cooperate to solve complex tasks and

obtain the best overall results. In a multi-agent system (MAS), these agents interact with each other and the environment to achieve their goals. MADRL can be categorized into four main paradigms based on their underlying paradigm as fully decentralized algorithms, credit assignment (Foerster et al., 2017; Li et al., 2022), communication (Foerster et al., 2016; Sukhbaatar et al., 2016), and algorithms with centralized training and decentralized execution (CTDE) (Lowe et al., 2017; Yu et al., 2022). For the fully decentralized paradigm, all agents learn independently. For one agent, other agents are regarded as a part of the environment, thus the agent cannot distinguish between the noise from the environment and the explorations of other agents. Therefore, convergence of the algorithm is not guaranteed (Hao et al., 2017). The credit assignment converts the global reward into estimated local reward for each agent to train, and the communicating algorithm allows agents to widen their observation space and cooperate (Schmidt et al., 2022). The CTDE is an alternative and popular approach where a group of agents can be trained simultaneously by applying a centralized method via an open communication channel. Specifically, the CTDE algorithm uses an actor-critic architecture to decompose the policy and value estimates into explicit actor and critic networks. The critic is used to improve the policy's performance but it only works in training. Therefore, only independent actors are needed during execution. CTDE paradigm further exploits this by training multiple agents with a shared critic and allows agents to operate without communication after training. The multi-agent deep deterministic policy gradient

(MADDPG) method (Lowe et al., 2017), based on the CTDE algorithm, is specifically adopted in this chapter. MADDPG features the centralized learning and decentralized execution paradigm, in which the critic network utilizes extra information to facilitate the training process while actor networks made decisions based on their own local observations during execution. This chapter adopted the MADDPG as the MADRL controller for its merits as: 1) can learn policies using local information at execution time, and 2) applies to the cooperative interaction for the control purpose in this chapter. During training, however, it is found out that the convergence performance of the MADDPG on a two-point maglev levitation system is not satisfactory. The currently existing algorithms in the reinforcement learning (RL) such as MADDPG use the Bellman optimality equation to update the value function. The Partial differential equation (PDE) of the Bellman optimality equation, known as HJB equation, is combined with the Bellman optimality equation to update the value function in this chapter. The correctness and effectiveness of the developed control strategy are validated through a series of simulations. The main contributions of this work can be summarized as:

1) The HJB method is firstly incorporated into the MADRL method to improve the performance of the MADRL controller. The HJB equation is used in control theory to evaluate the optimality of the value function and to improve the training of the critic network.

2) The proposed control method can control the MIMO system directly without

decoupling the system into the single-point levitation system and deal with uncertainty model parameters. It shows robustness with external disturbances introduced to the system compared to the PID controller, and the fast convergence to the equilibrium point of the two-point levitation system is also realized.

The rest of the chapter is organized as follows. The dynamic model of the levitation system is given in Section 4.2. The HJB–MADRL controller design is described in Section 4.3. Section 4.4 presents numerical results and a discussion of the robustness of the proposed controller by comparing the performance with the PID controller. Section 4.5 presents experiment results and related discussion of the proposed controller. Finally, the conclusion is given in Section 4.6.

4.2 Modelling of two-point maglev levitation system

A typical EMS-type maglev system consists of maglev vehicles and elevated guideway, and each maglev vehicle typically has multiple levitation bogies that work together to bear the weight of the vehicle body by using air springs as buffers and supporters. Each levitation bogie is equipped with four levitation modules that are the basic levitation functional units of a maglev train, and each levitation module contains one pair of adjacent electromagnets controlled by a single-input-single-output (SISO) controller. In **Figure 4–1**, the schematic diagram and control structure of the EMS-type maglev system proposed are depicted. Two types of controller constructions are presented, i.e., the decentralized SISO controller and the centralized MIMO decoupling controller. The decentralized SISO controller is commonly used in the maglev levitation systems. It adjusts the control input based on the local air gap information obtained from the sensor of one levitation unit. However, the stiff structure of the bogie causes the two levitation units to affect each other's control performance. To address the challenges posed by the decentralized SISO controller, the decoupling control is adopted to solve the MIMO control problem. To further address the issues encountered during the decoupling control, the chapter presents a HJB–MADRL controller. The proposed controller utilizes the airgap information from both levitaion units to generate appropriate control signals. The HJB–MADRL controller is discussed in detail in Section 4.3.3 of the thesis.

The dynamic model of the levitation system based on the half bogie in the HJB–MADRL controller design requires several assumptions (He et al., 2015):

1) Homogeneous mass distribution: The mass distribution of the half bogie is assumed to be uniform throughout its structure;

2) Gravity center alignment: The assumption is made that the gravity centre of the half bogie coincides with its geometrical centre;

3) Rigid beam rail: The F-type rail on the guideway is treated as a rigid beam;

4) Neglecting magnetic leakage and edge effect: The model neglects the magnetic leakage and edge effect of the electromagnet;

5) Equating uniformly distributed electromagnet force: The uniformly distributed electromagnet force is approximated as two concentrated forces acting on the centre of each levitation unit.

6) Rigid connections: The vehicle body and bogie are considered to be rigidly connected.

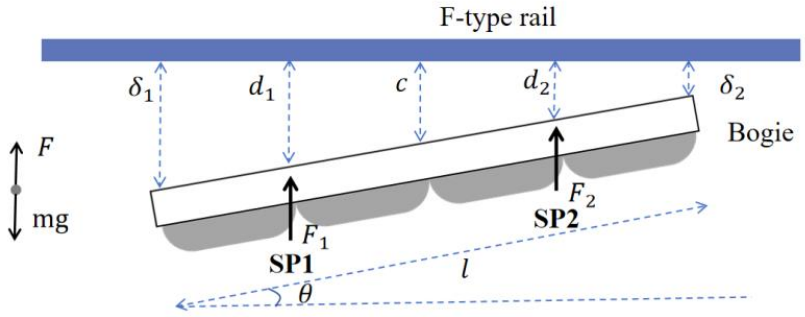


Figure 4–2 Force diagram of a half bogie

With these assumptions, the vertical force diagram of the half bogie is given in **Figure 4–2** (SP1 and SP2 denote two levitation units) , the kinetic equation of the half bogie is written as

$$\begin{cases} m\ddot{c} = mg - F_1 - F_2 \\ J\ddot{\theta} = F_2 \cdot \frac{l}{4}cos(\theta) - F_1 \cdot \frac{l}{4}cos(\theta) \end{cases} \tag{4.1}$$

where m is the total mass of the half bogie and the vehicle body, c is the gap between the center of the half bogie and the F-rail, l is the length of the half bogie, θ is the pitching angle of the half bogie, F_1 and F_2 are the equivalent electromagnetic force of levitation unit 1 and levitation unit 2, J is the rotary inertia of half bogie along z-axis. The geometric relationships are as

$$\begin{cases} c = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_2 \\ d_1 = \frac{3}{4}\delta_1 + \frac{1}{4}\delta_2 \\ d_2 = \frac{1}{4}\delta_1 + \frac{3}{4}\delta_2 \end{cases} \tag{4.2}$$

As the pitch angle is small, giving that

$$\theta \approx \sin(\theta) = \frac{\delta_1 - \delta_2}{l}, \cos(\theta) = 1 \tag{4.3}$$

According to Biot-Savart law and Maxwell equation, the levitation electromagnetic force generated by a single electromagnet coil and the voltage at both ends of the coil can be expressed as

$$F(i, \delta) = \frac{\mu_0 N^2 A}{4}\left(\frac{i}{\delta}\right)^2 \tag{4.4}$$

$$u = Ri + \frac{\mu_0 N^2 A}{2}\dot{i} - \frac{\mu_0 N^2 A}{2\delta^2} i\dot{\delta} \tag{4.5}$$

where N denotes the number of coils turns, A represents the magnetic pole area, δ denotes the airgap, R is the coil resistance, and μ_0 represents the vacuum

permeability. Assuming that the parameters of the front and rear electromagnets are consistent, the system state variables, output variables , and control variables can be define as

$$\begin{cases} z = [z_1 \quad z_2 \quad z_3 \quad z_4]^T = [\delta_1 \quad \dot{\delta}_1 \quad \delta_2 \quad \dot{\delta}_2]^T \\ y = [y_1 \quad y_2]^T = [\delta_1 \quad \delta_2]^T \\ u = [u_1 \quad u_2]^T = [i_1^2 \quad i_2^2]^T \end{cases} \tag{4.6}$$

Then state space expression of two-point levitation system is written as

$$\begin{cases} \dot{\boldsymbol{z}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{z} + \begin{bmatrix} 0 & 0 \\ \left(-\frac{l^2}{8J} - \frac{1}{2m}\right)\frac{2A}{\mu_0} & \left(\frac{l^2}{8J} - \frac{1}{2m}\right)\frac{2A}{\mu_0} \\ 0 & 0 \\ \left(\frac{l^2}{8J} - \frac{1}{2m}\right)\frac{2A}{\mu_0} & \left(-\frac{l^2}{8J} - \frac{1}{2m}\right)\frac{2A}{\mu_0} \end{bmatrix} \boldsymbol{u} + \begin{bmatrix} 0 \\ g \\ 0 \\ g \end{bmatrix} \\ \boldsymbol{y} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \boldsymbol{z} \end{cases} \tag{4.7}$$

It can be observed from Equation (4.7) that the front and rear levitation units are coupled with each other in dynamics. The coupling interference between two levitation units may lead to the decline of levitation control performance, and even lead to safety accidents.

4.3 HJB-MADRL control design

4.3.1 MAMDP for a maglev control system

The two-point maglev levitation control problem (u_1 and u_2) can be formulated as a Multi-agent Markov Decision Processe (MAMDP) (as depicted in **Figure 4–3**) in the RL framework. In the MAMDP formulation, each of two agents makes decision based on their respective functions, denoted as $\pi_1(\cdot)$ and $\pi_2(\cdot)$, which depend on the current states of the maglev levitation system, represented as $a_1 = u_1 = \pi_1(o_1)$ and $a_2 = u_2 = \pi_2(o_2)$. Here, o_1 represents the state of the first levitation unit, and o_2 represents the state of the second levitation unit. The MADRL is a method employed to solve the MAMDP problem. In MADRL, the agents learn policies, denoted as $\pi_1(\cdot)$ and $\pi_2(\cdot)$, by interacting with the maglev levitation system (environment). At each time step, two agents observe the current states o_1 and o_2, and then perform actions, denoted as a_1 and a_2, based on their current policies. After performing the actions, the agents receive scalar rewards, denoted as r_1 and r_2, respectively, from the environment.

The goal of the MADRL is for agents to learn optimal policies, denoted as $\pi_1^*(\cdot)$ and $\pi_2^*(\cdot)$, that maximize their cumulative rewards over time as $U_i(t) = \sum_{k=0}^{\infty} \gamma^k \cdot r_i(t+k), i = 1, 2, 0 \leq \gamma \leq 1$. The discount rate, represented as γ, determines the present value of future rewards, emphasizing the importance of immediate rewards compared to future rewards. To solve the MADRL problem, the actor–critic method, as described in (Li, 2017), is commonly employed. In the actor–critic framework, the actor, also known as policy, selects an action based on the observed state from the

environment. The actor (policy function) aims to improve its policy by maximizing expected rewards. It chooses actions based on the observed state and tries to find the optimal action that leads to higher rewards. The critic (value function) provides feedback to the actor by evaluating the value of being in a particular state following the chosen policy. The critic estimates the expected cumulative rewards from a given state and policy. The actor–critic method strikes a balance between reducing the variance of policy gradients and introducing bias through value function estimation (Konda and Tsitsiklis, 2003; Schulman et al., 2016).

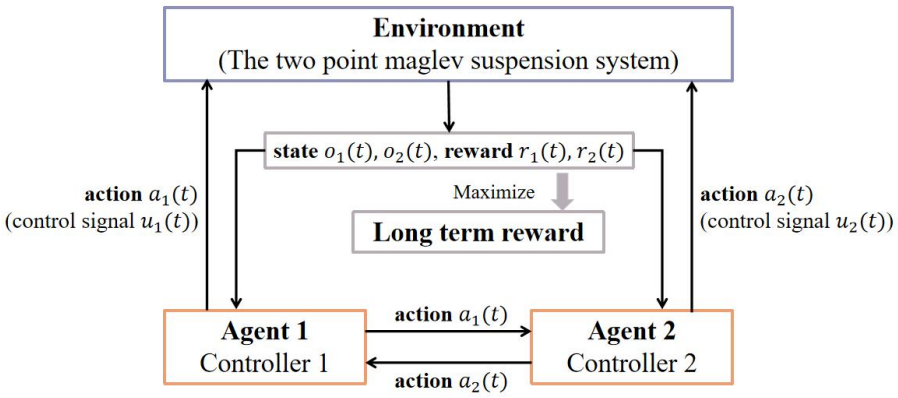


Figure 4–3 Schematic diagram of the MAMDP problem

4.3.2 MADRL for control design

In many actor–critic algorithms, the on-policy policy gradient formulation is used to update the actor (Peters and Schaal, 2008). On-policy training are known for their stability but typically suffer from poor sample complexity (Schulman et al., 2017). To address this issue and improve sample efficiency, researchers have

developed algorithms (Donoghue et al., 2017; Gu et al., 2017) that incorporate off-policy samples and utilize higher-order variance reduction techniques. One notable off-policy actor-critic algorithm is the deterministic policy gradients (DPGs) proposed by Silver et al. (Silver et al., 2014). DPG is specifically designed for high-dimensional continuous control problems and has demonstrated superior performance compared to its stochastic policy gradient counterparts. The DDPG (Lillicrap et al., 2016) algorithm extends the DPG approach to DRL by incorporating a deep neural network architecture. DDPG utilizes a Q-function estimator to enable off-policy learning and employs a deterministic actor to maximize this Q-function. Building upon single-agent DRL algorithms, several multi-agent DRL algorithms have been developed to address cooperative or competitive situations involving multiple agents. One such algorithm is MADDPG, which operates within a CTDE paradigm. MADDPG allows the policies to leverage extra information to ease training to facilitate learning in multi-agent scenarios. Hence, we use the MADDPG to design a cooperative levitation control in this chapter.

The two controllers in the maglev control system can be regarded as two agents with policies parameterized by $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\}$. Assume $\boldsymbol{\pi} = \{\pi(o_1; \boldsymbol{\theta}_1), \pi(o_2; \boldsymbol{\theta}_2)\} = \{\pi_1, \pi_2\}$ to be a set of agent policies, the gradient of the expected reward for agent i ($J(\boldsymbol{\theta}_i) = E[r_i]$) is given by

$$\nabla_{\boldsymbol{\theta}_i} J(\boldsymbol{\theta}_i) = E\left[\nabla_{\boldsymbol{\theta}_i} log\pi_i(a_i|o_i) V_i^{\boldsymbol{\pi}}(\boldsymbol{s}, a_1, a_2)\right], \boldsymbol{s} = \{o_1, o_2\} \tag{4.8}$$

Here $Q_i^{\boldsymbol{\pi}}(\boldsymbol{s}, a_1, a_2)$ is a centralized action-value function that takes the actions of all agents as a_1 and a_2, state information $\boldsymbol{s}$ as input, and the Q-value as outputs for agent i. The standard centralized action-value function $Q_i^{\boldsymbol{\pi}}$ is updated as

$$L_{std}(\boldsymbol{\theta}_i) = E_{\boldsymbol{s},\boldsymbol{a},\boldsymbol{r},\boldsymbol{s}'}[(V_i^{\pi}(\boldsymbol{s}, a_1, a_2) - y)^2] \tag{4.9}$$

where $y = r_i + \gamma V_i^{\boldsymbol{\pi}^-}(\boldsymbol{s}', a'_1, a'_2)|_{a'_j=\pi_j^-(o_j),\, j=1,2}$ and $\boldsymbol{\pi}^- = \{\pi(o_1; \boldsymbol{\theta}_1^-), \pi(o_2; \boldsymbol{\theta}_2^-)\}$ is a set of target policies with delayed parameters $\boldsymbol{\theta}_i^-$; $\boldsymbol{s}'$, a'_1 and a'_2 are state information and actions of two agents at next time step, respectively. In MADDPG, deterministic policies are adopted as they are suitable for the continuous-time control problem of the maglev levitation system. The deterministic policy was proposed by Silver et al. (2014), and can be written as $\boldsymbol{a} = \boldsymbol{\pi}(\boldsymbol{s}; \boldsymbol{\theta})$. The function denotes that for each pair of states $\boldsymbol{s}$ and policy parameters $\boldsymbol{\theta}$, the exact action values $\boldsymbol{a}$ can be determined, instead of probability distribution of actions.

4.3.3 HJB–MADRL for control design

The MADDPG is verified to outperform traditional RL algorithms on a variety of cooperative and competitive multi-agent environments. However, it is found out from the training process that the MADDPG would easily fall into the local optimum. In order to force the convergence of the value function in MADDPG, the partial differential equation (PDE) of the HJB is incorporated with the algorithm.

Reformulate the state space expression (4.7) of the two-point levitation system as

$$\dot{\boldsymbol{z}} = f(\boldsymbol{z}, \boldsymbol{u}), \boldsymbol{z}(t_0) = \boldsymbol{z}_0 \tag{4.10}$$

The goal of the control is to design a control input to assure the states of the system converge to the reference airgap by minimizing a pre-defined performance function of

two agents as

$$J_i(t) = \int_t^{\infty} r_i(\tau)d\tau = \int_t^{\infty} (\mathcal{M}(z_{(i*2-1)}(\tau)) + u_i(\tau)\aleph u_i(\tau))d\tau, i = 1, 2 \quad (4.11)$$

where $\mathcal{M}(\cdot) \geq 0$, $\aleph = \aleph^T > 0$ and the terminal cost is ignored, and $z_{(i*2-1)}, i = 1, 2$ indicates the state variables defined in (4.6). The value function V for any time interval T can be written as

$$V_i(t) = \int_t^{t+T} r_i(\tau)d\tau = \int_t^{t+T} (\mathcal{M}(z_{(i*2-1)}(\tau)) + u_i(\tau)\aleph u_i(\tau))d\tau \quad (4.12)$$

A PDE to (4.12) is

$$r_i(t) + \frac{\partial V_i(t)}{\partial z_{(i*2-1)}}\dot{z}_{(i*2-1)}(t) = 0, V_i(0) = 0 \quad (4.13)$$

where $\dot{z}_{(i*2-1)}$ can be replaced by $z_{(i*2)}$.Then the Hamiltonian is given by

$$H\left(z_{(i*2-1)}, u_i, \frac{\partial V_i}{\partial z_{(i*2-1)}}\right) = r_i(t) + \frac{\partial V_i(t)}{\partial z_{(i*2-1)}(t)} z_{(i*2)}(t) \quad (4.14)$$

The optimal value given by the Bellman optimility equation is defined as

$$r_i(t)|\boldsymbol{u}^* + \frac{\partial V_i^*}{\partial z_{(i*2-1)}} z_{(i*2)} \Big| \boldsymbol{u}^* = 0 \quad (4.15)$$

The equation (4.15) is the HJB equation. Current MADRL algorithms focus on approaching the true total reward through temporal difference (TD) method in critic networks. The proposed HJB–MADDPG adopts the optimal control strategy by incorporating HJB equation into the critic networks. The HJB loss for the proposed algorithm is defined as

$$L_{HJB}(\boldsymbol{\theta}_i) = r_i(t) + \frac{\partial V_i(t)}{\partial z_{(i*2-1)}(t)} z_{(i*2)}(t) \quad (4.16)$$

where $\frac{\partial V_i(t)}{\partial z_{(i*2-1)}(t)}$ is computed using auto-differentiation in deep neural network. The modified loss function for each agent critic network is computed as

$$L(\boldsymbol{\theta}_i) = \alpha L_{std}(\boldsymbol{\theta}_i) + \beta L_{HJB}(\boldsymbol{\theta}_i) \tag{4.17}$$

where α is chosen as 0.5, and β is determined based on the magnitude of the $L_{HJB}(\boldsymbol{\theta}_i)$ loss compared to the $L_{std}(\boldsymbol{\theta}_i)$, so that both kind of loss can have similar weight for network updating. The detailed HJB – MADDPG algorithm is given in **Table 4- 1**, and the schematic diagram of the algorithm is as in **Figure 4–4**.

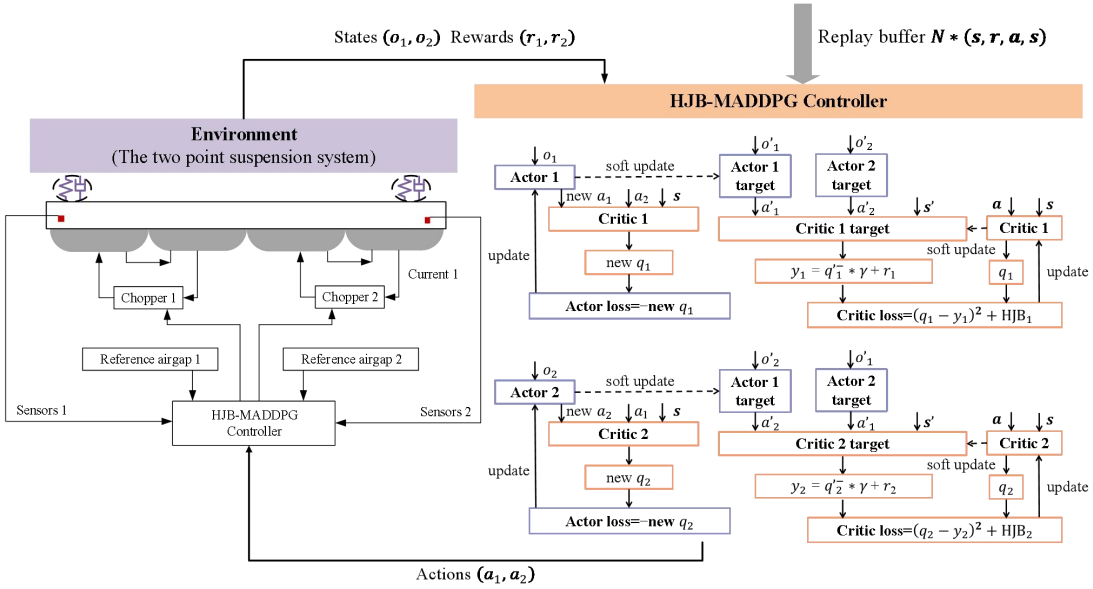


Figure 4–4 Schematic diagram of the HJB–MADDPG algorithm

Table 4- 1 Algorithm of the HJB–MADDPG

Algorithm 1 Pseudo code

Step 1: Initialization

Initialize critic networks $q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_1), q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_2)$, and actor network $\pi(o_1; \boldsymbol{\theta}_1)$, $\pi(o_2; \boldsymbol{\theta}_2)$ with random parameters $\boldsymbol{w}_1, \boldsymbol{w}_2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$.

Initialize target critic networks $q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_1^-), q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_2^-)$, and actor network $\pi(o_1; \boldsymbol{\theta}_1^-), \pi(o_1; \boldsymbol{\theta}_2^-)$ with parameters $\boldsymbol{w}_1^- \leftarrow \boldsymbol{w}_1, \boldsymbol{w}_2^- \leftarrow \boldsymbol{w}_2, \boldsymbol{\theta}_1^- \leftarrow \boldsymbol{\theta}_1, \boldsymbol{\theta}_2^- \leftarrow \boldsymbol{\theta}_2$.

Initialize replay buffer $\mathcal{B}$.

Step 2: Training

For iteration =1, 2, ..., max episode

Select action with exploration noise $a_i = \pi(o_i; \boldsymbol{\theta}_i) + \epsilon$, $\epsilon \sim \mathcal{N}(0, \sigma)$ and obtain reward r_i and next state $\boldsymbol{s}'$, then store the transition $(\boldsymbol{s}, \boldsymbol{a}, \boldsymbol{r}, \boldsymbol{s}')$ in the replay buffer $\mathcal{B}$.

For agent *i*=1, 2, do

Sample mini-batches of M transitions from the replay buffer $\mathcal{B}$

$$a'^-_1 = \pi(o'_1; \boldsymbol{\theta}_1^-) + \epsilon, a'^-_2 = \pi(o'_2; \boldsymbol{\theta}_2^-) + \epsilon, \epsilon \sim clip(\mathcal{N}(0, \sigma), -c, c)$$

$$q'^-_1 = q(\boldsymbol{s}', a'^-_1, a'^-_2; \boldsymbol{w}_1^-), q^-_{2,t+1} = q(\boldsymbol{s}', a'^-_1, a'^-_2; \boldsymbol{w}_2^-)$$

$$q_1 = q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_1), q_2 = q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_2)$$

TD target: $y_i = r_i + \gamma \cdot q'^-_i$

TD error: $\delta_i = q_i - y_i$

TD loss: $L_{std}(\boldsymbol{\theta}_i) = \frac{1}{M} \sum_{j=1}^{M} (\delta_i)^2$

HJB loss: $L_{HJB}(\boldsymbol{\theta}_i) = r_i + \frac{\partial q_i}{\partial z_{(i*2-1)}} z_{(i*2)}$

Update critic networks by minimizing the modified loss:

$$L(\boldsymbol{\theta}_i) = \alpha L_{std}(\boldsymbol{\theta}_i) + \beta L_{HJB}(\boldsymbol{\theta}_i)$$

Update actor networks using the sampled policy gradient:

$$\boldsymbol{\theta}_i \leftarrow \boldsymbol{\theta}_i + \lambda \cdot \nabla_{\boldsymbol{\theta}} \pi(o_i; \boldsymbol{\theta}_i) \cdot \nabla_a q(\boldsymbol{s}, \boldsymbol{a}; \boldsymbol{w}_i)$$

end for

Update target networks for each agent:

$$\boldsymbol{\theta}_i^- \leftarrow \tau \boldsymbol{\theta}_i + (1 - \tau) \boldsymbol{\theta}_i^-$$

$$\boldsymbol{w}_i^- \leftarrow \tau \boldsymbol{w}_i + (1 - \tau) \boldsymbol{w}_i^-$$

end for

4.4 Numerical results and discussion

4.4.1 HJB–MADRL controller training

The objective of the HJB–MADRL controller is to maintain a stable 6 mm airgap between two electromagnets and the F-type rail. The environment of the designed controller is the nonlinear two-point levitation control system which is established in Section 4.2. The initial air gap is 10.5 mm (Leng et al., 2019) and the value of the system model parameters are given in **Table 4- 2.**

Table 4- 2 Parameter values of the maglev levitation system

Physical quantity	Value	Physical quantity	Value
Mass m / kg	21.2	Vacuum permeability μ_0 / (Hm^{-1})	$4\pi \cdot 10^{-7}$
Number of Turns of coil N	540	Area of coil A/m^2	0.0014
Coil resistance R/Ω	2.1	Stable air gap δ_{eq} /m	0.006

For the control algorithm training, the HJB – MADDPG is implemented by modifying the code of the MADDPG. The structure of both the critic and actor networks in the HJB–MADDPG is designed with three hidden layers, as depicted in **Figure 4–5**. It can be observed from the figure that the critic networks take both actions and states as inputs and provide the one estimated result. The output of the critic networks is consist of the action-value function and HJB loss, the former part

represents the expected cumulative reward for a given state-action pair, and the second one is a measure of the deviation from the optimal solution. The actor networks in the HJB–MADDPG receive states as inputs and generate actions for the two levitation units.

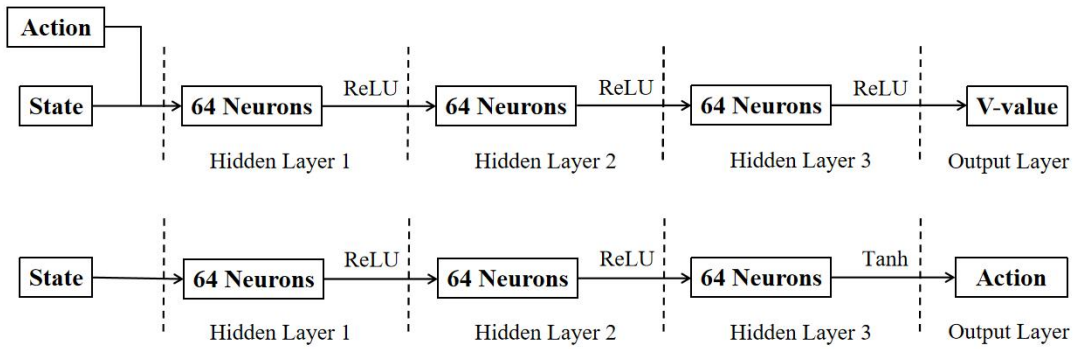


Figure 4–5 Schematic of the critic and actor networks

Specifically, deep neural networks consist of fully connected layers are used for representing actor and critic functions. Both of the actor and critic networks use a rectified linear unit activation function (ReLU) as the activation function, and the output layer of the actor network is processed using a tanh function. The learning rates are set as 1×10^{-3} for the critic networks and 1×10^{-4} for the actor networks. The iteration of the training is set to be 20,000 and the time step in each iteration is 500 ($\Delta t = 0.001s$). In addition, the discounted factor is 0.95 and the update parameter τ is 0.01. The exploration is done by adding noise to the actions. The whole algorithm is trained in Python 3.8 with PyTorch 1.5.1 with a mini-batch of 64 transitions sampled

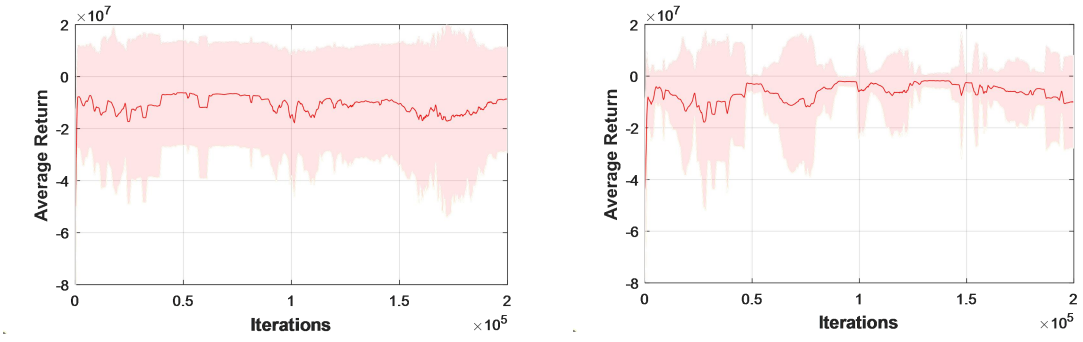
from a replay buffer $\mathcal{B}$ with a size of 1×10^5. All the parameters are fine-tuned based on the reference (Lowe et al., 2017) and the specific characteristics of the magnetic levitation system. Taking the objective of the algorithm into consideration, the reward functions are designed as:

$$r_i = -\zeta\left(\delta_i - \delta_{eq}\right)^2 - \dot{\delta}_i^2, i = 1, 2 \tag{4.18}$$

where ζ is determined based on the magnitude of $\dot{\delta}_i$.

4.4.2 Comparison with MADRL

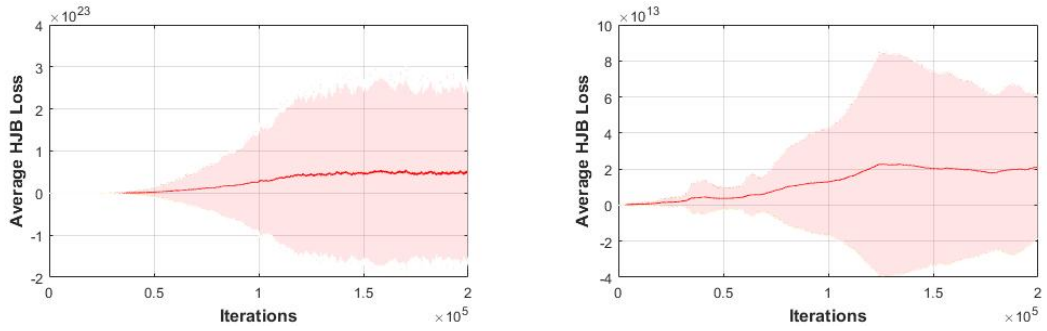
In order to demonstrate the performance and ability of the proposed algorithm, average return curves, HJB loss curves, and Bellman optimality loss curves of the HJB–MADDPG and MADDPG during training are given in **Figure 4–6, Figure 4–7, and Figure 4–8**. In these figures, the average return curves are obtained over 10 consecutive episodes, while the average HJB loss curves and Bellman optimlity loss curves are over 50 episodes. The shaded area in the figures indicate the standard deviation and solid lines indicate the mean values.



(a) (b)

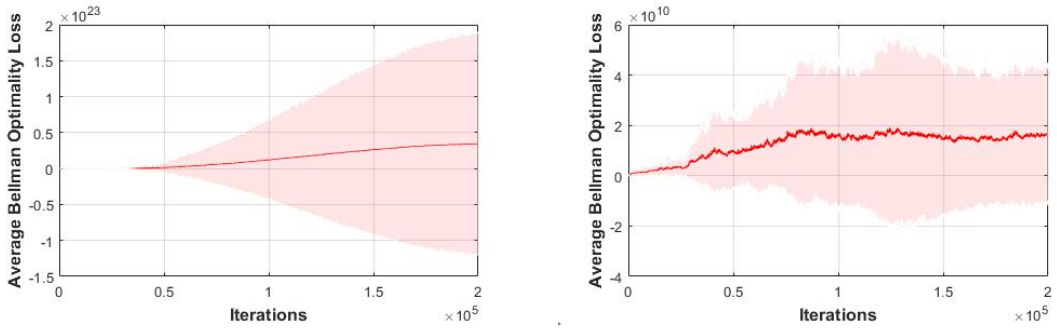
Figure 4–6 Average return curves: (a) MADDPG algorithm, (b) HJB–MADDPG

algorithm



(a) (b)

Figure 4–7 Average HJB loss curves: (a) MADDPG algorithm, (b) HJB–MADDPG algorithm



(a) (b)

Figure 4–8 Average Bellman optimality loss curves: (a) MADDPG algorithm, (b) HJB–MADDPG algorithm

It can be observed from **Figure 4–6** that the average return curve of the HJB–MADDPG approaches 0 more closely than the curve of the MADDPG. This indicates that the HJB–MADDPG algorithm achieves better performance during training. The HJB–MADDPG algorithms leverages finite differences to approximate the underlying governing equation of the two-point maglev levitation system, and

utilizes auto-differentiation of the network to solve the HJB equation, facilitating optimal control. In **Figure 4–7** and **Figure 4–8**, we can see that the HJB loss and Bellman optimlity loss curves of the HJB–MADDPG falls within the range of $0\sim3\times10^{13}$ and $0\sim2\times10^{10}$, respectively, while MADDPG exhibits HJB loss and Bellman optimlity loss within the range of $0\sim1\times10^{23}$ and $0\sim5\times10^{22}$, respectively. The HJB loss represents the deviation from the optimal solution to the HJB equation, and the Bellman optimlity loss measures the deviation from the Bellman optimlity loss, characterizing the optimality of the learned value function. The HJB–MADDPG algorithm′s curves being within a narrower range suggests better convergence towards the optimal solution compared to MADDPG.

In summary, the HJB–MADDPG algorithm demonstrates improved performance compared to MADDPG algorithm. This is evident from the average reward curves, HJB loss curves, and the Bellman optimality loss curve. The incorporation of the HJB loss regularization term and the utilization of an optimal control strategy in HJB–MADDPG contribute to enhanced learning of the value function and improved convergence of the algorithm.

4.4.3 Effectiveness of the HJB–MADRL controller

After training the HJB–MADDPG algorithm, the best neural network parameters obtained during the training process are set as the parameters of the HJB–MADRL controller. This ensures that the learned knowledge and policies are utilized in the control process. To evaluate the control performance of the trained HJB–MADRL controller, PID controller and model-guided controller (Chen et al., 2018) are used for

comparison. The model-guided controller is a control strategy that leverages a mathematical model of the system to inform its control actions. Specifically, in the context of maglev systems, the model-guided controller relies on the physical and dynamic equations governing the magnetic levitation process to predict system behavior and determine appropriate control inputs. This approach enables the controller to achieve improved performance and stability by utilizing system knowledge, serving as a valuable benchmark for assessing advanced control algorithms. The transfer function of the PID controller is presented in (3.23). We set $K_p = 50$, $K_d = 100$ and $K_i = 0.008$ in this section based on trial and error. The model-guided controller was proposed as

$$\begin{cases} u_1 = \frac{1}{b}\left[-g - k_1(\delta_1 - \delta_{eq}) - k_2\dot{\delta}_1\right] \\ u_2 = \frac{1}{b}\left[-g - k_1(\delta_2 - \delta_{eq}) - k_2\dot{\delta}_2\right] \end{cases} \tag{4.19}$$

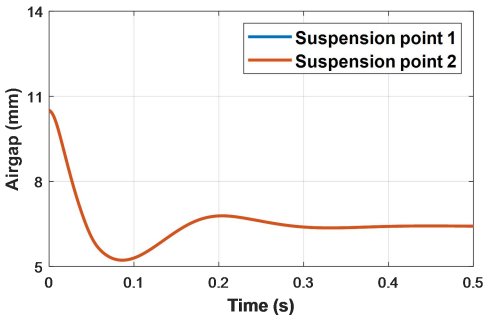
where u_1 and u_2 are signals, b is a parameter determined by the system as $(Al^2/(4J) + A/(m))/\mu_0$, δ_1, $\dot{\delta}_1$, δ_2, and $\dot{\delta}_2$ are defined as in Section 4.2, k_1 and k_2 are set as $k_1 = 9000, k_2 = 30$ to achieve a reasonably good tracking performance.

The air gap curves obtained by using the PID controller, model-guided controller and HJB – MADRL controller are presented in **Figure 4–9**. It can be observed that airgap of the maglev levitation system successfully achieves the desired level using the designed PID controller, the model-guided controller and the HJB – MADRL controller. However, there are differences in their performance:

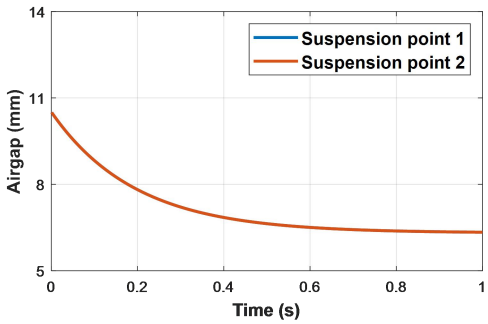
1) Convergence time: The HJB – MADRL controller achieves the desired airgap

level much faster than the PID controller and the model-guided controller. The convergent time for the PID controller and the model-guided controller are approximately 30 s and 0.8 s, whereas for the HJB–MADRL controller, it is around 0.3 s. This indicates that the HJB–MADRL controller exhibits significantly faster response and achieves the desired air gap level in a much shorter time.

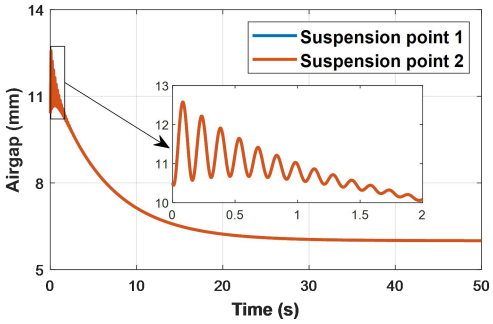
2) Fluctuation and overshooting: The PID controller shows some fluctuations in the air gap curve, particularly during the initial 2 s. Additionally, there is a maximum overshooting of approximately 2 mm observed in the airgap curve for the PID controller. In contrast, the HJB–MADRL controller's airgap exhibits smoother behavior with maximum overshooting less than 1 mm.



(a)



(b)



(c)

Figure 4–9 Control curves of the controllers: (a) HJB–MADRL controller, (b) Model-guided controller, (c) PID controller

4.4.4 Robustness of the HJB－MADRL controller

4.4.4.1 Effect of pitch motion

In order to demonstrate the robustness of the proposed method, the effect of pitch motion on the maglev train is considered. Pitch motion refers to the tilting or rotation of the train caused by the uneven height of the F-type rail or external forces such as wind (Han and Kim, 2016). To simulate pitch motion, a random uniform value between -4 N and 4 N is applied as a torque to the system. The air gap curves of the HJB–MADRL, model-guided and PID controllers under the influence of pitch motion are shown in **Figure 4–10** (a), (b) and (c), respectively. It can be observed from the figures that the effect of pitch motion on the HJB–MADRL controller and the model-guided controller is nearly negligible compared to the previous air gap curve in **Figure 4–10**. The HJB–MADRL controller and the model-guided controller demonstrate relatively stable and close to the desired level. This indicates that the HJB–MADRL controller and the model-guided controller are capable of effectively maintaining an air gap around 6 mm even when pitch motion is present.

On the other hand, the air gap curve of the PID controller under pitch motion shows larger fluctuations compared to the previous case. Even after 30 s, the air gap

of the PID controller continues to fluctuate around the equilibrium point, indicating a less stable control performance under pitch motion.

Based on these observations, it can be concluded that the HJB–MADRL controller and the model-guided controller exhibits robustness in maintaining the desired air gap even in the presence of pitch motion. In contrast, the PID controller shows larger fluctuations and struggles to maintain stable levitation under the same conditions.

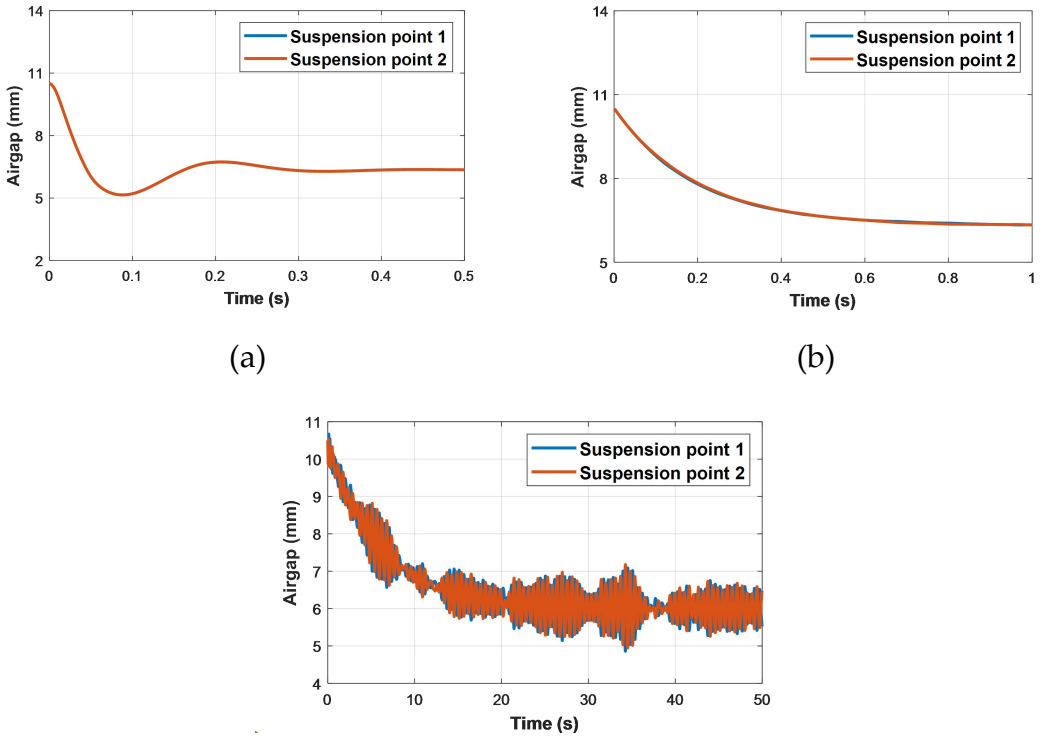


Figure 4–10 Control curves of the controllers with pitch motion: (a) HJB–MADRL controller, (b) Model-guided controller, (c) PID controller

4.4.4.2 Effect of random disturbance

To evaluate the capability of the HJB–MADRL controller under external disturbances, a random uniform value between -10 N and 10 N is considered as a type of disturbance acting on the levitation point 1. The results obtained using the HJB – MADRL controller, the model-guided controller, and the PID controller are presented in **Figure 4–11**. It can be observed that the HJB–MADRL controller quickly reaches equilibrium point at 0.25 s, showcasing its robustness against external disturbances. The model-guided controller reaches the target point at 0.8 s, also shows the robustness to the extra disturbances. On the other hand, the PID controller takes significantly longer, approximately 30 s, to reach the target point under the same disturbances.

Furthermore, it is evident that the HJB–MADRL controller and the model-guided controller can effectively control both levitation units simultaneously, maintaining stable levitation even in the presence of external disturbances. This indicates the capability of the HJB–MADRL controller the model-guided controller to handle multiple control objectives simultaneously. In contrast, the PID controller exhibits larger fluctuations in response to the disturbance compared to the HJB–MADRL controller. The fluctuations in the air gap curve of the PID controller are more pronounced, indicating a less stable control performance.

Based on these observations, it can be concluded that the HJB–MADRL controller and the model-guided controller demonstrate superior capability in handling external disturbances compared to the PID controller. The HJB–MADRL controller achieves faster convergence, maintains stable levitation under disturbances,

and effectively controls both levitation units simultaneously. Though the model-guided controller takes a little more time to convergent, it also demonstrates robustness to the disturbances. While the PID controller shows slower convergence, larger fluctuations, and struggles to handle the disturbance effectively.

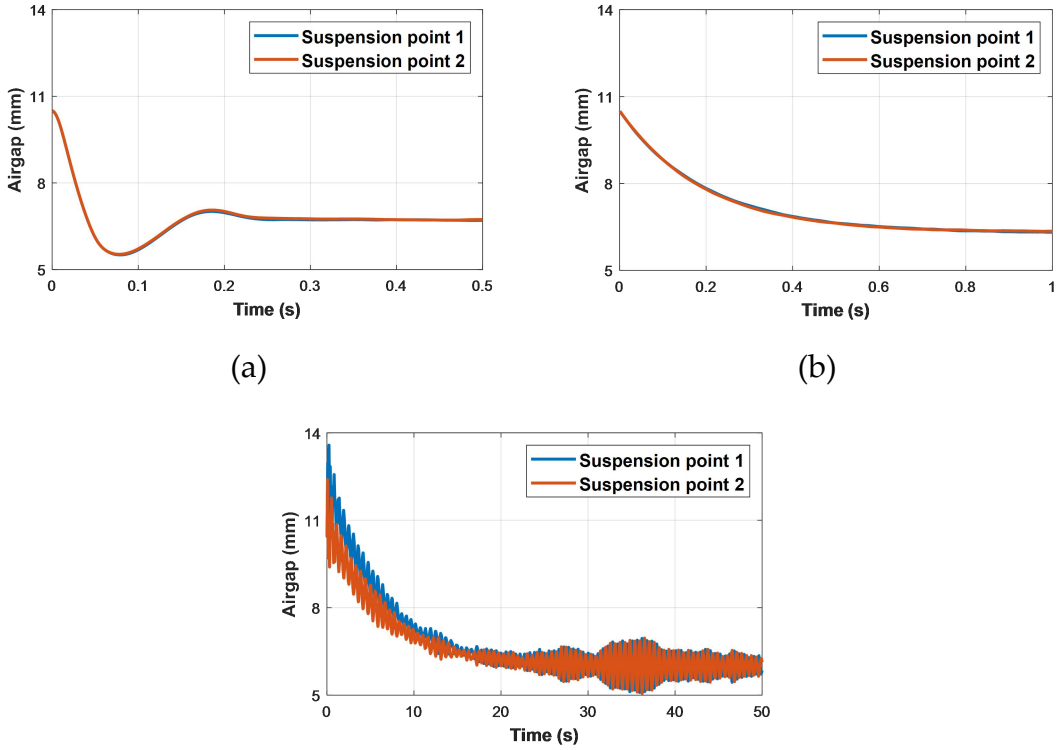


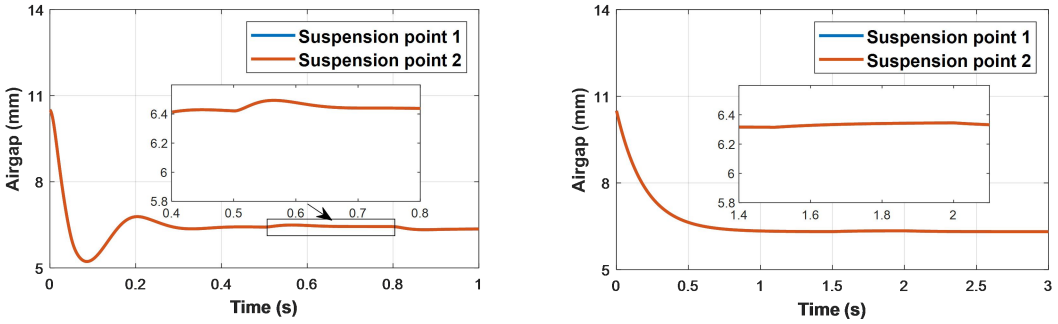
Figure 4–11 Control curves of the controllers with disturbance at levitation point 1: (a) HJB–MADRL controller, (b) Model-guided controller, (c) PID controller

4.4.4.3 Effect of change in train load

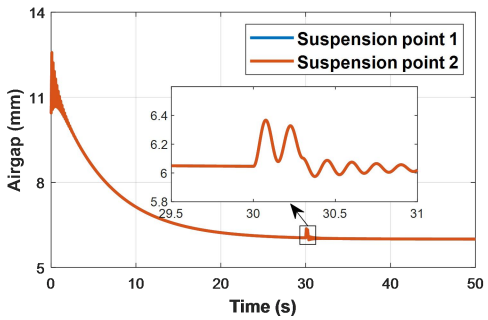
To further evaluate the robustness of the HJB–MADRL controller, the study sudden changes in the train load, which can occur due to variations in the number of passengers. These sudden changes in train load can pose challenges to the levitation

system, requiring it to adapt quickly to ensure comfort and safety. In the simulation, the train load is changed abruptly from 0 N to 50 N, and lasts for 0.5 s. The airgap curves obtained using the HJB–MADRL controller, the model-guided controller, and PID controller are given in **Figure 4–12**. As can be seen, there is only a slight oscillation in the air gap of the maglev levitation system using the HJB–MADDPG controller and the model-guided controller. This indicates that the HJB–MADDPG controller and the model-guided controller can quickly adapt to the sudden change in train load and maintain a stable air gap. In the contrast, the air gap fluctuations in the PID controller are much larger and continues for more than 1 s. This suggests that the PID controller struggles to handle the sudden change in train load and maintain a stable air gap within a shorter time frame.

These findings confirm that the HJB–MADDPG controller and the model-guided controller are much more stable and robust than the PID controller when it comes to coping with changes in train load. The HJB–MADDPG controller and the model-guided controller also demonstrate the ability to adapt quickly and maintain a stable air gap, ensuring the comfort and safety of the maglev train even in the presence of sudden changes in passenger load.



(a) (b)



(c)

Figure 4–12 Control curves of the controllers under changes in load: (a) HJB–MADRL controller, (b) Model-guided controller, (c) PID controller

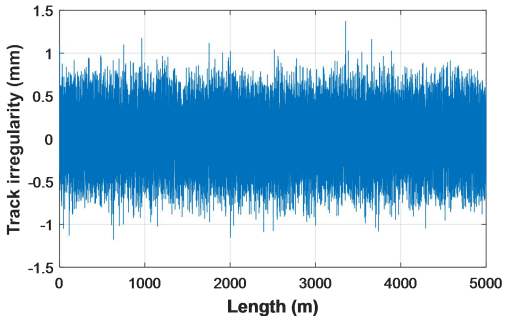
4.4.4.4 Effect of track irregularity

Track irregularity is the main source of excitation in maglev control systems. It can cause a substantial instability in a levitation controller as it leads directly to fluctuations in the levitation air gap. To evaluate the effects of the random nature and characteristics of rail irregularity on the designed controller, the power spectrum density function as in (3.26) (in Chapter 3) is adopted to simulate the vertical profile of variations in the guideway geometry (Yang and Lin, 2005). The vertical profile of the track irregularity is given in **Figure 4–13** (a).

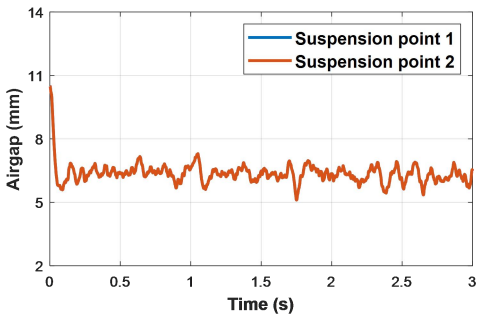
The airgap curves associated with the HJB–MADDPG controller, the model-guided controller, and PID controller are presented in **Figure 4–13** (b), (c), and (d). From figures, it can be observed that both controllers exhibit oscillations in the air gap, caused by the fluctuations caused by the track irregularity. However, there are

notable differences in the behavior of the two controllers. The air gap curve of the HJB–MADRL controller, after 0.2 s, shows relatively small fluctuations within a narrow range (less than 1.5 mm) around the equilibrium point. This indicates that the HJB–MADDPG controller can effectively maintain the air gap at approximately 6 mm despite the presence of track irregularities. On the other hand, the air gap curve of the model-guided controller and the PID controller demonstrates larger oscillations with a range of over 2.5 mm and 5 mm, respectively. This suggests that the model-guided controller and the PID controller struggle to maintain stable air gaps under the influence of track irregularities.

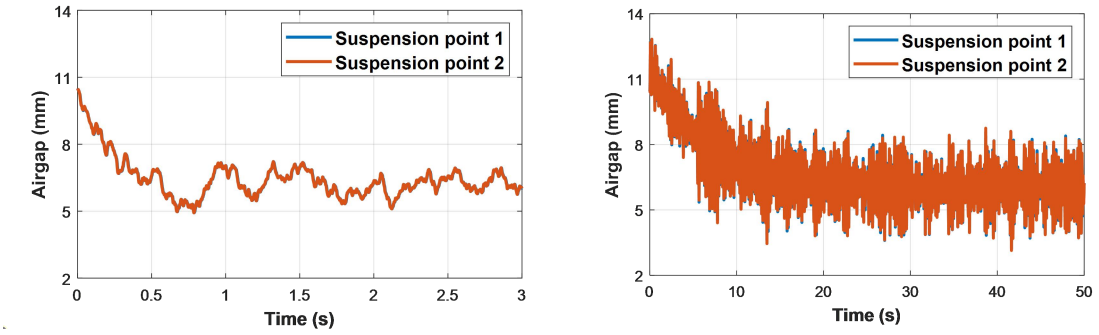
Thus, it can be concluded that the HJB–MADDPG controller is capable of maintaining a consistent air gap around 6 mm under track irregularities. In contrast, the the model-guided controller and the PID controller exhibits larger fluctuations, indicating a less stable control performance in maintaining the desired air gap.



(a)



(b)



(c) (d)

Figure 4–13 Control curves of the controllers with track irregularity: (a) Track irregularity, (b) HJB–MADRL controller, (c) Model-guided controller, (d) PID controller

4.5 Experiment on a full-scale maglev bogie

4.5.1 Experiment setting

The experimental setup of the 1:1 scale maglev bogie levitation system is illustrated in **Figure 4–14**, showcasing its physical structure and key components. The central bogie frame is supported by two air springs and equipped with four evenly distributed electromagnets. These electromagnets are organized into two groups: the front and rear, with each pair connected in series to ensure synchronized current flow within the same group.

The **Figure 4–14** also shows the electrical cabinet, which supplies the system with DC 330V and DC 110V power, ensuring stable and adequate energy for the electromagnets' operation. The operating console serves as the interface for real-time system control, allowing users to send commands to the controller and monitor critical parameters during experiments. Specifically, the dSPACE-related software installed in the computer was utilized to generate the control signals. The control box directs current to the electromagnets based on control signals generated by the control algorithm. The air springs, located beneath the suspension frame, are connected to an air source system that adjusts their pressure to simulate various loading conditions, such as different passenger loads. These components work together to create a comprehensive experimental platform for testing and validating control algorithms under realistic conditions.

The operational workflow begins with system initialization via the electrical cabinet and control box, which energizes the electromagnets to establish baseline suspension gaps. Real-time data from gap sensors at both ends of the bogie frame is processed by the controller, which then applies corresponding control actions to the electromagnets. During experiments, the air springs introduce dynamic loads to the suspension frame, allowing for the evaluation of the control algorithms' performance and robustness. The data acquisition system records suspension gaps, electromagnet currents, and air spring pressures, providing detailed insights into system dynamics and control effectiveness. The parameter values of the full-scale maglev bogie, as presented in **Table 4-3**, are sourced from the technical documentation provided by the manufacturer.

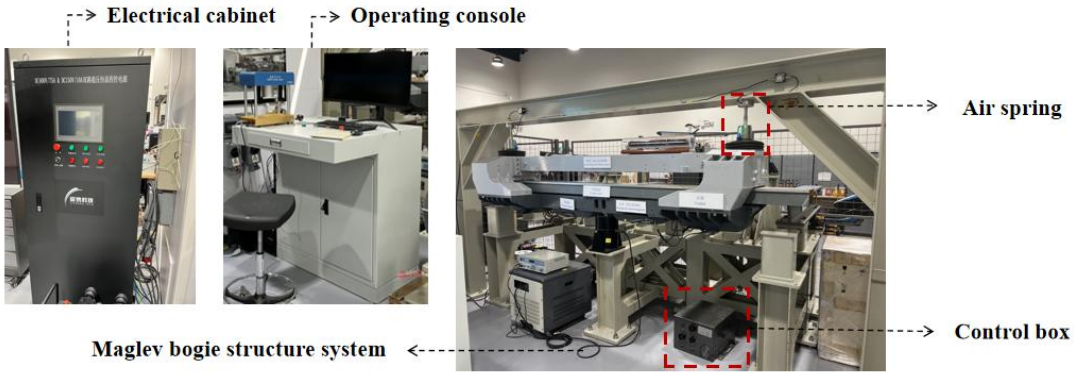


Figure 4–14 The experimental setup of the full-scale 1:1 maglev bogie levitation system

Table 4-3 Parameter values of the full-scale maglev bogie

Physical quantity	Value	Physical quantity	Value
Mass m / kg	750	Vacuum permeability μ_0 / (Hm^{-1})	$4\pi \cdot 10^{-7}$
Number of Turns of coil N_m	450	Area of coil A_m/m^2	0.024
Coil resistance R/Ω	1.2	Stable air gap x_{eq} /m	0.008

4.5.2 Experiment results and discussion

In this part, some experimental results of the full-scale maglev bogie levitation system are provided to verify the practical control performance of the designed HJB–MADRL controller. The PID controller is adopted to compare with the proposed controller, and the coefficients of the PID controller are chosen as $k_p = 12000, k_i = 10000, k_d = 1000$. In **Figure 4–15**, the control performance and control signals of the PID and HJB–MADRL controllers are displayed. Notably, the HJB–MADRL method achieves the target air gap in just 2.3 s, a significant improvement over the PID controller, which requires 4.6 s to reach the same target air gap. Furthermore, the HJB–MADRL controller can well control the two levitation units converge to 0.008 mm target, while the levitation point 2 of the PID controller fails to reach 0.008 mm in this scenario.

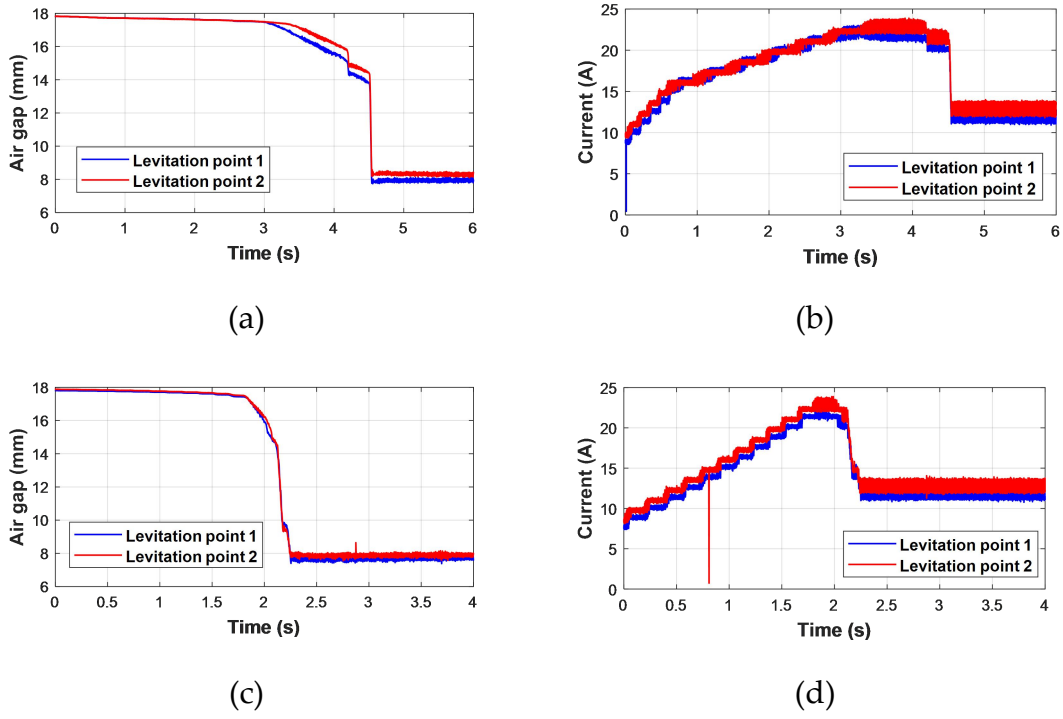


Figure 4–15 Control curves of the controllers: (a) Air gap curves of the PID controller, (b) Control signal curves of the PID controller, (c) Air gap curves of the HJB－MADRL controller, (d) Control signal curves of the HJB－MADRL controller

4.6 Conclusion

In this section, a new algorithm that integrates PDE of the Bellman optimality equation into the MADDPG is proposed. This algorithm, termed HJB–MADRL, is designed to address the nonlinear control problem in maglev levitation systems. With an appropriate reward and state design, the HJB–MADRL algorithm can provide the optimal control solution for a two-point maglev levitation system. The effectiveness of the HJB–MADRL algorithm is verified by comparing its average return curves, HJB loss curves, and Bellman optimality loss curves during training with the original MADRL algorithm. The control performance of the HJB–MADRL controller is compared with the traditional PID controller and the newly proposed model-guided controller. In addition, the robustness of the proposed controller is assessed by evaluating its response to pitch motion of the train, disturbance force, load change, and rail irregularity. The main results are as follows.

1) The convergence performance of the HJB–MADRL algorithm is more stable than that of the original MADRL algorithm in average return curves, HJB loss curves, and Bellman optimality loss curves.

2) From simulation, the convergent time of the HJB–MADRL controller is 0.3 s, whereas the conventional PID controller is approximately 30 s, and the model-guided controller is around 0.8 s.

3) From the experiment on the full-scale maglev bogie levitation system, the convergent time of the HJB–MADRL controller is 2 s, while the conventional PID controller is approximately 6.2 s. Besides, the two levitation units controlled by PID controller can not converge to the target air gap well.

4) The HJB–MADRL controller is more robust than the PID controller when the system is disturbed by changes in pitch motion of the train, disturbance force, change in train load, and rail irregularity. When compared with the model-guided controller, the HJB–MADRL controller shows superior robust performance under rail irregularity condition.

Future work studies could consider the following issues to improve the HJB–MADRL method: data processing speed and data storage; coupling of maglev train system and F-type rail; and experimental verification of the DRL method.

CHAPTER 5 RECIPROCAL CONTROL BARRIER FUNCTION INCORPORATED SAFE DEEP REINFORCEMENT LEARNING CONTROL OF MAGLEV TRAIN–GUIDEWAY COUPLING SYSTEM

During long-term operation, even small perturbations or disturbances can cause maglev train system to become unstable as the levitation air gap between the maglev train and guideway is only within millimeters. Specially, as the guideway is highly susceptible to deformation, making the maglev control system more easily to strike the guideway and resulting in damage of the system. Thus, the coupling effect between maglev train and flexible guideway cannot be ignored when designing the levitation controller. Moreover, ensuring the air gap remains within a safe range is crucial to prevent collisions between the maglev system and the guideway. To address the mentioned issue, a safe deep reinforcement learning (SDRL) controller is proposed for the magnetic levitation system considering the deformation of the flexible guideway. Notably, a reciprocal control barrier function (RCBF) is augmented in the reward function of the DRL to ensure safety and optimality of the controller. Furthermore, the designed RCBF includes a damping coefficient to balance the safety and optimality of the system. The improved performance of the proposed SDRL algorithm is verified by comparing to original DRL algorithm. The superiority of the proposed controller in terms of efficiency and accuracy is validated through a comparative analysis with a traditional proportional – integral – derivative (PID)

controller and a novel genetic algorithm tuned super twisting sliding mode controller (GA–ST–SMC) via simulations. Additionally, the robustness of the RCBF–SDRL controller is assessed under conditions of changing train loads, load fluctuations, external disturbances, and track irregularities. Furthermore, experiments have also been carried out to validate the control performance of the proposed RCBF–SDRL controller in comparison to the PID controller on a magnetic levitation system.

5.1 Introduction

As mentioned in Chapters 3 and 4, the EMS-type maglev system is inherently unstable due to the its strong system nonlinearity, its sensitivity to disturbances, and characteristics of the magnetic circuit. Thus, the control design is a pivotal component to ensure the air gap between the maglev train and its guideway at a stable value.

Recently, many research has been carried out to design nonlinear controllers for the maglev trains to effectively handle external disturbances. Yang et al. (2004) designed a robust nonlinear controller with improved position-tracking performance in presence of uncertain model parameters. Huang et al. (2000) designed an adaptive backstepping controller to improve system stability in the presence of model uncertainty. Yaseen et al. (2022) combined the adaptive control and sliding mode control (SMC) approaches to devise three nonlinear suspension controllers, i.e., adaptive terminal SMC, adaptive backstepping SMC and adaptive integral backstepping SMC, to ensure the air gap stayed within a desired range. However, these novel controllers are all designed under the assumption of rigid guideway without considering the guideway deformation. Research has demonstrated a strong coupling interaction between the maglev train, levitation control, and guideway systems, forming a time-varying system problem. This interaction introduces additional damping or mass effects to the guideway's vibration, causing temporal changes in its natural frequency and modal damping ratio. The guideway's flexibility,

acting as an external force of excitation on the train, can lead to dynamic instability and unacceptable vibrations, such as self-excited and resonant vibrations (Han et al., 2009; Li et al., 2015). When the frequency of external excitation on the train nears the train vehicles' natural frequency, the train's dynamic response can significantly increase, resulting in resonance (Yang and Yau, 2015; Li et al., 2016). Resonance-induced instability of the levitation gap, potentially causing levitation control failure, has been observed in maglev trains on both test and commercial lines. Therefore, to ensure the operational stability and safety of maglev trains, it is crucial to develop advanced control methods that account for the complex dynamic interaction mechanism of the maglev train–guideway system.

Currently, several studies have been conducted out to evaluate the dynamic performance of the maglev train–guideway coupling system (Lengyel and Kocsis, 2014; Kim et al., 2015; Min et al., 2017) and develop appropriate controllers for its regulation. Wang et al. (Wang et al., 2014) developed a full-state feedback controller and used the particle swarm optimization (PSO) algorithm to optimize the control gains; then, they verified the performance of the proposed controller through simulation and test rig even under violent external disturbances. Zhou et al. (2017) presented an effective novel approach by using a pair of mirror FIR filters alongside an adaptation mechanism controller for vibration suppression under various common track irregularities, including sinusoidal track profiles, random track irregularities, and track steps. Sun et al. (2019) developed a fuzzy adaptive tuning PID controller, in

which the controller gains were adjusted according to the identified disturbance or changes in the system parameters. In (Sun et al., 2020), Sun et al. employed an enhanced version of the Apriori algorithm to extract and process data from a stored historical database. This process facilitated the establishment of a reliable database, based on which the researchers developed an improved fuzzy adaptive controller. The proposed controller was verified on a full-scale Internet of Things (IoT) maglev train system at Tongji University. In addition, sliding mode robust adaptive control (Chen et al., 2019), robust control (Li and Shen, 2020), feedback linearization control (Zhang et al., 2022), double loop PID considering control gain perturbation (Sun et al., 2023), and genetic algorithm tuned Super Twisting sliding mode controller (GA–ST–SMC) (Teklu et al., 2023), have been proposed. Nevertheless, the controllers mentioned above are unable to ensure safety when implemented in the practical control of maglev train-guideway coupling systems. Particularly, in the presence of resonance-induced instability induced by the coupling effect between the maglev train and the guideway, there exists a risk of the maglev system coming into strike the guideway, potentially leading to system damage. Besides, the majority of these control strategies exhibit degraded performance in the presence of disturbances because their performance relies on a precise model and detailed system information that is difficult to obtain in practice. To address these issues, an advanced method named reciprocal control barrier function incorporated safe reinforcement learning (RCBF–SDRL) is proposed in this chapter.

Over the past decades, reinforcement learning (RL) algorithms have been widely adopted in control areas (Bellemare et al., 2013; Koutník et al., 2013; Levine et al., 2016), as they can automatically learn a control policy. Researchers have also used the DRL algorithms to solve the maglev levitation control problems (Zhao et al., 2021; Zhu et al., 2024), and these DRL controllers have less overshoot and more robust than conventional PID and LQR controllers. In DRL, the agent adopts "trial and error" mechanism to explore possible operations based on the current state. the control problem is firstly reformulated as Markov Decision Process (MDP) in DRL. Then, a deep neural network (DNN) is trained to solve for optimal solutions. With sufficient training, the trained DNN is capable to make online decisions in levitation control. The mechanism for DRL to control is to maximize a predefined long term reward regardless of any practical constraints (Zhou et al., 2022). Thus, the conventional DRL algorithms can not ensure the safety of the trained controller in real application. In this chapter, a constrained MDP (CMDP) (Altman, 1999) framework is adopted to describe the magnetic levitation control problem with safe constraints to ensure that the air gap of the maglev system and the guideway is in safe range, and a RCBF–SDRL method is proposed to solve it. The proposed RCBF–SDRL method integrate RCBF in the SDRL algorithm to convert the constrained optimal problem to a unconstrained one. RCBF is a type of control barrier function (CBF), a commonly used approach to ensure the safety of control systems (Tee et al., 2009; Srinivasan et al., 2018; Agrawal and Sreenath, 2017). The main idea behind CBF is to develop an

interior penalty method for converting constrained optimal control methods into unconstrained ones (Malisani et al., 2016). For the RCBF used in this chapter, it is associated with a safe set C that is unbounded on the set boundary, i.e., $B(x) \to \infty$ a s $x \to \partial C$. The effectiveness of the proposed RCBF–SDRL control method is validated through a series of simulations and experiments. The main contributions of this work can be summarized as:

1)The RCBF is firstly integrated into the SDRL method to solve for optimal control problem. The RCBF is used in the SDRL to transform the constrained safe control problem into unconstrained one.

2)The proposed RCBF–SDRL control method can control the magnetic levitation system coupling with flexible guideway and deal with uncertain model parameters. It shows robustness with external disturbances introduced to the system compared to the PID controller and GA–ST–SMC.

The rest of the chapter is organized as follows. The dynamic model of the magnetic system with flexible guideway is given in Section 5.2. The RCBF–SDRL controller design is described in Section 5.3. Section 5.4 and 5.5 present numerical and experimental results and a discussion of robustness of the proposed controller by comparing the performance with the PID controller, respectively. Finally, the conclusion is given in Section 5.6.

5.2 Modeling of magnetic levitation system with flexible guideway

The configuration of a typical EMS-type maglev train is as shown in **Figure 5–1**, which consists of maglev vehicles and elevated guideway. Each maglev vehicle has multiple levitation bogies working together to bear the weight of the vehicle body through air springs. Each levitation bogie is equipped with four levitation modules which are the basic levitation functional units of a maglev train. The levitation function is realized via the levitation controllers by controlling the currents of the electromagnets to achieve the stable levitation of the bogie. The two sides of bogie are decoupled by anti-rolling beams, and the mechanical decoupling strategy allows each levitation module to be controlled independently. Thus, the levitation performance of a magnetic levitation system relies on a single levitation module. Taking the flexible guideway into consideration, the levitation control problem for a magnetic levitation system can be simplified as a problem consisting of a single levitation unit, a flexible guideway, and a levitation controller. The control scheme is illustrated in **Figure 5–1**. as can be seen, the levitation controller receives signals from the current and air gap sensors. Then, the control signals are amplified by a power chopper and sent to the levitation electromagnets. In proper control mode, the levitation electromagnets are commanded to generate an appropriate attractive force to adjust the air gap between the electromagnets and the guideway around a reference value, e.g., 9 mm.

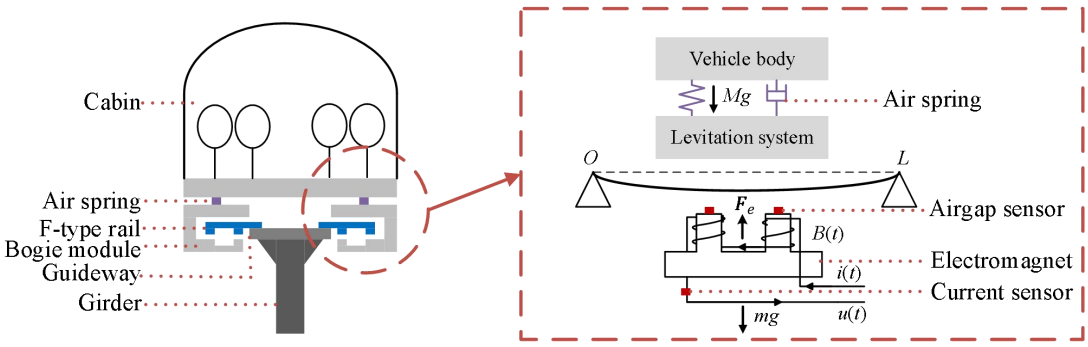


Figure 5–1 Cross-section of an EMS-type maglev system and a schematic of a single EMS module with flexible guideway

5.2.1 Modeling of guideway system for vertical motion

The guideway of the maglev train is normally supported by a viaduct or pier as in **Figure 5–1**, and it can be simplified to a simply supported beam for analysis. The vertical motion of the guideway is described by the following partial differential equation of Euler-Bernoulli beam vibration

$$E_r I_r \frac{\partial^4 x_r}{\partial y^4} - T_r \frac{\partial^2 x_r}{\partial y^2} + \rho_r \frac{\partial^2 x_r}{\partial t^2} + c\frac{\partial x_r}{\partial t} + k_r x_r = f(y,t) \tag{5.1}$$

$$\text{where } f(y,t) = \begin{cases} F_e/l_e, y_0 \le y \le y_0 + l_e \\ 0, else \end{cases}$$

where x_r is the vertical displacement of the guideway along y-axis; E_r and I_r denote the Young's modulus of elasticity and the cross-sectional inertia of the guideway beam, respectively; T_r represents the tension generated when the beam is deformed; ρ_r is the beam mass linear density; c denotes the damping coefficient of the beam; k_r is the elasticity coefficient when the beam is elastically deformed; $f(y,t)$ is the distribution load density when the maglev vehicle passes through; F_e

denotes the electromagnetic levitation force; and l_e is the effective length of single levitation module.

According to the Equation (5.1), the mathematical model of the flexible guideway's vertical displacement can be obtained as

$$x_r = \sum_n \varphi_n(y) q_n(t), n = 1, 2, \ldots, \infty \tag{5.2}$$

$$\varphi_n(y) = \sqrt{\frac{2}{M_r}} sin\left(\frac{n\pi y_0}{L}\right) \tag{5.3}$$

$$\ddot{q}_n(t) + 2\xi_n\left(\frac{n\pi}{L}\right)^2 \sqrt{\frac{E_r I_r}{\rho_r}} \dot{q}_n(t) + \left(\frac{n\pi}{L}\right)^4 \frac{E_r I_r}{\rho_r} q_n(t) = \sqrt{\frac{2}{M_r}} sin\left(\frac{n\pi y_0}{L}\right) F_e \tag{5.4}$$

where $\varphi_n(y)$ denotes the *n* th order function corresponding to the simply supported beam; $q_n(t)$ is the generalized coordinate corresponding to the function at time *t*; M_r represents mass of the guideway, L is the span length of the guideway beam, and ξ_n is the damping ratio of the *n*-th order function.

As the first mode occupies the majority of the response, it is often the dominant factor in the overall behavior of the system.Hence, x_r can be regarded as the displacement of the first mode as

$$x_r = \sqrt{\frac{2}{M_r}} sin\left(\frac{\pi y_0}{L}\right) q_1(t) \tag{5.5}$$

Then the Equation (5.4) can be rewritten as follows

$$\ddot{q}_1(t) + 2\xi_1\left(\frac{\pi}{L}\right)^2 \sqrt{\frac{E_r I_r}{\rho_r}} \dot{q}_1(t) + \left(\frac{\pi}{L}\right)^4 \frac{E_r I_r}{\rho_r} q_1(t) = \sqrt{\frac{2}{M_r}} sin\left(\frac{\pi y_0}{L}\right) F_e \tag{5.6}$$

By replacing the q_1 in the Equation (6) using x_r in the Equation (5.5), the Equation (5.6) can be expressed as

$$\ddot{x}_r(t) + 2\xi_1\left(\frac{\pi}{L}\right)^2\sqrt{\frac{E_r I_r}{\rho_r}}\dot{x}_r(t) + \left(\frac{\pi}{L}\right)^4\frac{E_r I_r}{\rho_r}x_r(t) = \frac{2}{M_r}sin^2\left(\frac{\pi y_0}{L}\right)F_e \tag{5.7}$$

The vertical displacement of the guideway can be obtained by solving Equation (5.7).

5.2.2 Modeling of magnetic levitation system for vertical motion

It can be observed from **Figure 5–1** that the levitation air gap x_e with the flexible guideway can be obtained as

$$x_e = x_m - x_r \tag{5.8}$$

where x_m denotes the vertical displacement of the electromagnet.

The magnetic force $F_e\,(i, x_e)$ according to Maxwell's equation and Biot-Savart's theorem is as

$$F_e\,(i, x_e) = \frac{\int_0^t \psi_e(i,\ x_e)dt}{\partial x_e} \tag{5.9}$$

where i denotes electromagnet current. According to Kirchhoff magnetic-circuit law, $\psi_e(i, x_e)$ can be obtained as

$$\psi_e(i, x_e) = \frac{N^2 i(t)}{R(x_e)} \tag{5.10}$$

where reluctance $R(x_e) = 2x_e(t)/(\mu_0 A_e)$, N denotes the coil number of turns, μ_0 is air permeability, and A_e denotes the effective magnetic pole area. The air gap magnetic flux density of the levitation electromagnet is as follows

$$B(t) = \frac{\mu_0 N i(t)}{2x_e} \tag{5.11}$$

Then the expression of the magnetic force can be obtained as

$$F_e\,(i, x_e) = \frac{B(t)^2 A_e}{\mu_0} \tag{5.12}$$

Using Newton's second law, the dynamics equation of the levitation system can be expressed as

$$m\ddot{x}_m(t) = (m+M)g - \frac{B(t)^2 A_e}{\mu_0} \tag{5.13}$$

where M is mass of the cabin and m is mass of the electromagnet.

5.2.3 Modeling of magnetic levitation system for vertical motion

The magnetic levitation system model with flexible guideway based on the first-order vibration of the flexible guideway is generated by integrating the vertical motion model of the guideway structure with the vertical motion of the electromagnet as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\dfrac{A_e}{m\mu_0}B(t)^2 + \dfrac{m+M}{m}g \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = \dfrac{2}{M_r}sin^2\left(\dfrac{\pi y_0}{L}\right)\dfrac{A_e}{\mu_0}B(t)^2 - 2\xi_1\left(\dfrac{\pi}{L}\right)^2\sqrt{\dfrac{E_r I_r}{\rho_r}}x_4 - \left(\dfrac{\pi}{L}\right)^4\dfrac{E_r I_r}{\rho_r}x_3 \end{cases} \tag{5.14}$$

where $[x_1, x_2, x_3, x_4] = [x_m, \dot{x}_m, x_r, \dot{x}_r]$, and control signal is $B(t)$. Assuming that

$$k_1 = \frac{A_e}{\mu_0}, k_2 = \frac{2}{M_r}sin^2\left(\frac{\pi y_0}{L}\right), k_3 = \left(\frac{\pi}{L}\right)^2\sqrt{\frac{E_r I_r}{\rho_r}}$$

Then the nonlinear state space function can be expressed as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\dfrac{k_1}{m}B(t)^2 + \dfrac{m+M}{m}g \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = k_1 k_2 B(t)^2 - 2\xi_1 k_3 x_4 - {k_3}^2 x_3 \end{cases} \tag{5.15}$$

5.3 RCBF-SDRL controller design

5.3.1 CMDP for a magnetic levitation system

The safe control objective of maglev levitation system with flexible guideway is to assure the air gap between the electromagnets and guideway converge to the reference value using minimum energy with air gap constrained. In this chapter, SDRL is leveraged to solve this problem with low complexity, and a Constrained Markov Decision Process (CMDP) is firstly built to formulate the problem. A CMDP has five elements: a state space S; an action space A; an immediate/instantaneous reward R; transition dynamics P that maps a state-action pair at time t into a distribution of state at time $t + 1$; and a cost function C. The state of an established CMDP should fully capture the system behaviors and be adequate for calculating new states. We set vertical displacement of the electromagnet x_1 and velocity of the displacement x_2 as state, i.e., $s_t = [x_1(t), x_2(t)]$. The control signal $B(t)$, which is the input in the control system, is regarded as the action a_t. In line with the objective, the reward r_t is defined as $r_t = -(x_1(t) - x_{eq})^2 - a_t{}^2$, where x_{eq} denotes reference air gap. As for constraint of the problem, it is represented as $x_{max} \geq x_1(t) \geq x_{min}$. With regard to the constraint, the cost functions are defined as $c_t^1 = x_{max} - x_1(t)$, and $c_t^2 = x_1(t) - x_{min}$.

The maglev levitation system with flexible guideway is considered as an agent and makes decision as function $\pi(\cdot)$, based on the current state of the system s_t, i.e. $a_t = \pi(s_t)$. At each time step, the agent observes the state s_t, and performs action a_t based on the current policy, and receive scalar rewards r_t as well as costs c_t^1 and c_t^2 from the environment after the system transition occurs. The tuple

$\left(s_t, a_t, r_t, c_t^{i\,(i=1,2)}, s_{t+1}\right)$ is then stored in a replay buffer $\mathcal{B}$. During training, the SDRL algorithm samples from $\mathcal{B}$ and updates the policy, the controller will finally find an optimal policy π^* that maximizes the the total amount of the received reward (return) $U(t) = \sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k}, 0 \leq \gamma \leq 1$ while satisfying the constraints, where γ is the discount rate determining the present value of future rewards.

5.3.2 RCBF-SDRL for control design

To simplify the CMDP problem, a reciprocal control barrier function (RCBF) is integrated in the reward function to eliminate the cost terms. Given a safe closed set $\mathcal{S} \subset \mathrm{R}^n$ defined as

$$\mathcal{S} = \{x(t) | h(x(t)) \geq 0\} \tag{5.16a}$$

$$\partial\mathcal{S} = \{x(t) | h(x(t)) = 0\} \tag{5.16b}$$

$$\mathrm{Int}(\mathcal{S}) = \{x(t) | h(x(t)) > 0\} \tag{5.16c}$$

where $h(x(t))$ is a continuously differentiable function of $x(t)$.

In the RCBF, the barrier function (BF) $B_\gamma(x) \to \infty$ as $x \to \partial\mathcal{S}$, and the value of the BF can grow when it is far away from the boundary of $\mathcal{S}$. The BF only needs to fulfill a requirement that $\dot{B}_\gamma(x) \leq \alpha\left(1/B_\gamma(x)\right)$, where α is a class $\mathcal{K}$ function stated in **Definition 1**.

Definition 1. Class $\mathcal{K}$ function

A continuous function $\alpha: [0, a) \to [0, \infty)$ is a class $\mathcal{K}$ function if it strictly increasing and $\alpha(0) = 0$.

In conventional BFs, $\dot{B}_\gamma(x) \leq 0$ is enforced (Tee et al., 2009; Prajna et al., 2007), but this may not be desirable since it requires all sub-level sets of $\mathcal{S}$ to be

invariant. Thus, the condition is relaxed to be $\dot{B}_\gamma(x) \leq \gamma / B_\gamma(x)$, where γ is positive. In a more general context, RCBF can be formulated as in **Definition 2**.

Definition 2. Reciprocal control barrier function (RCBF)

For a nonlinear dynamic system, a continuously differentiable function $B_\gamma(x)$: $\mathrm{Int}(\mathcal{S}) \rightarrow \mathrm{R}$ is a reciprocal control barrier function (RCBF) for the safe set $\mathcal{S}$ defined in Equation (16) for a continuously differentiable function $h(x)$, if there exist class $\mathcal{K}$ functions α_1, α_2, and α_3 such that for all $x \in \mathrm{Int}(\mathcal{S})$:

$$\frac{1}{\alpha_1(h(x))} \leq B_\gamma(x) \leq \frac{1}{\alpha_2(h(x))} \tag{5.17a}$$

$$\dot{B}_\gamma(x) \leq \alpha_3(h(x)) \tag{5.17b}$$

A logarithmic barrier function candidate $B_\gamma(\boldsymbol{x}) = -\log\left(h(\boldsymbol{x})/(1+h(\boldsymbol{x}))\right)$ is used in this chapter, and it satisfies the important properties as

$$\inf_{x \in \mathrm{Int}\mathcal{S}} B_\gamma(x) \geq 0 \tag{5.18a}$$

$$\lim_{x \rightarrow \partial\mathcal{S}} B_\gamma(x) = \infty \tag{5.18b}$$

To determine the relative dominance of the RCBF over the reward function r_t, $B_\gamma(x)$ is modified as

$$B_\gamma(x) = -\log\left(\beta h(x)/(1+\beta h(x))\right) \tag{5.19}$$

where coefficient β balances safety and optimality by defining the extent to which safety takes precedence over other control objectives. The origin reward function r_t is modified to be $r_t = -(x_1(t) - x_{eq})^2 - {a_t}^2 - B_\gamma(x)$.

In proposed formulation, safety is ensured while a desired performance is maintained within the safe region. The incorporated RCBF $B_\gamma(x)$ acts as a safety component alongside the optimization of the other objectives. All of these goals,

including safety and optimization of other objectives, should be pursued through an iterative approach since the value function can not be solved directly. Furthermore. the safety of the system, along with the optimal solution within the safe region, will be examined subsequently.

As introduced earlier of the $B_\gamma(\boldsymbol{x})$ term, it is the primary factor near the risky area. Consequently, samples from both the safe region and the area near the safety boundary should be collected for training. For the SDRL algorithm, an actor-critic DRL algorithm named twin delayed deep deterministic policy gradient (TD3) (Fujimoto et al., 2007) method is adopted, which has advantage in increasing the stability and performance while considering function approximation error. The detailed RCBF incorporated SDRL algorithm is given in **Table 5-1**, and the schematic diagram of the algorithm is shown in **Figure 5–2**.

Table 5-1 Pseudo code of RCBF incorporated SDRL algorithm

Algorithm 1
Initialize critic networks $Q(s,a;\boldsymbol{w}_1), Q(s,a;\boldsymbol{w}_2)$, and actor network $\mu(s;\boldsymbol{\theta})$ with random parameters $\mathrm{w}_1, \mathrm{w}_2, \theta$.
Initialize target critic networks $Q(s,a;\boldsymbol{w}_1^-), Q(s,a;\boldsymbol{w}_2^-)$, and actor network $\mu(s;\boldsymbol{\theta}^-)$ with parameters $\boldsymbol{w}_1^- \leftarrow \boldsymbol{w}_1, \boldsymbol{w}_2^- \leftarrow \boldsymbol{w}_2, \boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$.
Initialize Replay buffer $\mathcal{B}$.

For iteration =1, 2, ..., T

Select action with exploration noise $a_t = \mu(s_t; \boldsymbol{\theta}) + \epsilon,\ \epsilon \sim \mathcal{N}(0, \sigma)$ and obtain RCBF incorporated reward r_t and next state s_{t+1}, then store the transition (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$.

Sample mini-batches of N transitions from the replay buffer $\mathcal{B}$

$$a_{j+1}^{-} = \mu(s_{j+1}; \boldsymbol{\theta}^{-}) + \epsilon, \epsilon \sim clip(\mathcal{N}(0, \sigma), -c, c)$$

$$Q_{1,j+1}^{-} = Q(s_{j+1}, a_{j+1}^{-}; \boldsymbol{w}_1^{-}), Q_{2,j+1}^{-} = Q(s_{j+1}, a_{j+1}^{-}; \boldsymbol{w}_2^{-})$$

$$Q_{1,j} = Q(s_j, a_j; \boldsymbol{w}_1), Q_{2,j} = Q(s_j, a_j; \boldsymbol{w}_2)$$

TD target: $\eta_j = r_j + \gamma \cdot min\{Q_{1,j+1}^{-}, Q_{2,j+1}^{-}\}$

TD error: $\delta_{1,j} = Q_{1,j} - \eta_j, \delta_{2,j} = Q_{2,j} - \eta_j$

Update of critic networks:

$$\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_1 - \alpha \cdot \delta_{1,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_1)$$

$$\boldsymbol{w}_2 \leftarrow \boldsymbol{w}_2 - \alpha \cdot \delta_{2,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_2)$$

If iteration mod k, **then**

Update actor network:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \beta \cdot \nabla_{\boldsymbol{\theta}} \mu(s_j; \boldsymbol{\theta}) \cdot \nabla_a Q(s_j, a_j; \boldsymbol{w}_1)$$

Update target networks:

$$\boldsymbol{\theta}^{-} \leftarrow \tau \boldsymbol{\theta} + (1 - \tau) \boldsymbol{\theta}^{-}$$

$$\boldsymbol{w}_1^- \leftarrow \tau \boldsymbol{w}_1 + (1-\tau)\boldsymbol{w}_1^-$$

$$\boldsymbol{w}_2^- \leftarrow \tau \boldsymbol{w}_2 + (1-\tau)\boldsymbol{w}_2^-$$

End if

End for

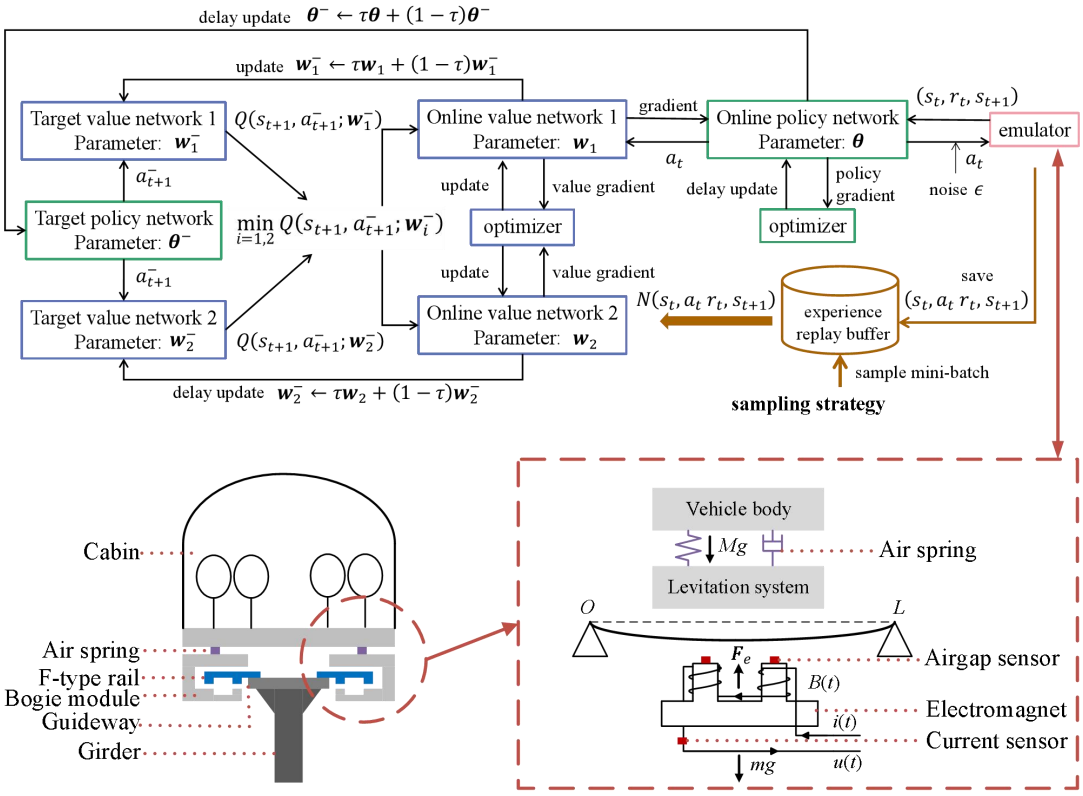


Figure 5–2 Schematic diagram of the RCBF–SDRL algorithm

5.3.3 Safety and stability analysis of the controller

In this section, we will analyze the safety of the system, as well as the stability and optimality of the solution within the safe region.

1) Safety analysis

To analyze the safety of the system, we first need to verify the existence of the value function and demonstrate the boundedness of the RCBF. Based on the proved results, safety is guaranteed. Before proceeding, admissible feedback control policy needs to be defined as in **Definition 3**.

Definition 3. Admissible feedback control policy

A control policy is deemed admissible for a safe optimal control problem if it stabilizes the system and its associated cost is bounded. it is defined as follow:

$$\boldsymbol{u} \in \mathcal{U}_a = \mathcal{U} \cap \mathcal{U}_{safe}$$

where $\mathcal{U}$ is the admissible control policy for the optimal control problem and $\mathcal{U}_{safe}$ is a set of safe inputs as $\mathcal{U}_{safe} = \{\boldsymbol{u} \in \mathrm{R}^m | \boldsymbol{x}^{\boldsymbol{u}} \in \mathrm{Int}(\mathcal{S})\}$, and $\boldsymbol{x}^{\boldsymbol{u}}$ is the state of the system evolved by the input $\boldsymbol{u}$.

Lemma 1.

Consider an admissible feedback control policy $\boldsymbol{u}_1 \in \mathcal{U}_a$. If there exists a time invariant positive function $V \in \mathrm{C}^1$ such that

$$\frac{\partial V^T}{\partial \boldsymbol{x}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}_1) + Q(\boldsymbol{x}) + B_\gamma(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 = 0, V(\boldsymbol{x}_0, \boldsymbol{u}_1) = J(\boldsymbol{x}_0, \boldsymbol{u}_1) \quad (5.20)$$

Then V can be regarded as the value function of the system for all $t \in [0, \infty)$, $V(\boldsymbol{x}, \boldsymbol{u}) = J(\boldsymbol{x}, \boldsymbol{u})$.

Proof for Lemma 1.

Assume $V(\boldsymbol{x}, \boldsymbol{u}_1) > 0$ exists, then we have

$$\mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}_0, \boldsymbol{u}_1) = \int_0^t \frac{\partial V}{\partial \boldsymbol{x}(\tau)}(f(\boldsymbol{x}(\tau)) + g(\boldsymbol{x}(\tau))\boldsymbol{u}_1)d\tau \quad (5.21)$$

Using $J(\boldsymbol{x}(t), \boldsymbol{u}) = \int_t^\infty Q(\boldsymbol{x}) + \boldsymbol{u}^T R \boldsymbol{U} + B_\gamma(\boldsymbol{x})d\tau$, we have

$$J(\boldsymbol{x}(t), \boldsymbol{u}_1) - J(\boldsymbol{x}_0, \boldsymbol{u}_1) = -\int_0^t Q(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 + B_\gamma(\boldsymbol{x}) d\tau \tag{5.22}$$

Based on (19) and (20),

$$\begin{aligned} J(\boldsymbol{x}(t), \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1) - (J(\boldsymbol{x}_0, \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}_0, \boldsymbol{u}_1)) \\ = -\int_0^t Q(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 + B_\gamma(\boldsymbol{x}) + \frac{\partial V}{\partial \boldsymbol{x}(\tau)} (f(\boldsymbol{x}(\tau)) \\ + g(\boldsymbol{x}(\tau)) \boldsymbol{u}_1) d\tau \end{aligned} \tag{5.23}$$

Using (5.20) in **Lemma 1**, we have $J(\boldsymbol{x}(t), \boldsymbol{u}_1) = \mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1)$. The proof is completed.

Lemma 2.

Assume positive value functions $V(\boldsymbol{x}, \boldsymbol{u}_1)$, $V(\boldsymbol{x}, \boldsymbol{u}_2)$, ..., $V(\boldsymbol{x}, \boldsymbol{u}_n)$ are associated with admissible control policy sequence $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ..., $\boldsymbol{u}_n \in \mathcal{U}_a$. If corresponding minimized Hamiltonian values satisfy $H_{min1} \leq H_{min2} \leq \ldots \leq H_{minn}$, then the RCBF term at each time step is bounded. The Hamiltonian function is defined as

$$H_i(x, \boldsymbol{u}_i, \nabla V_i) = r(\boldsymbol{x}, \boldsymbol{u}_i) + (\nabla V_i)^T (f(\boldsymbol{x}) + g(\boldsymbol{x}) \cdot \boldsymbol{u}_i) \tag{5.24}$$

Then the minimized Hamiltonian function is given as $H_{mini} = H_i(x, \boldsymbol{u}_i^*, \nabla V_i)$.

Proof for Lemma 2.

For any *i* and *j* that fulfill $0 \leq i \leq j \leq n$, assume that $H_{mini} \leq H_{minj}$, given

$$V(\boldsymbol{x}, \boldsymbol{u}_j) = V(\boldsymbol{x}, \boldsymbol{u}_i) + V_d(\boldsymbol{x}, \boldsymbol{u}_i) \tag{5.25}$$

Then, optimal policy $\boldsymbol{u}_j = -0.5 R^{-1} g^T \nabla V_j$ is adopted to replace the safe input term in the minimized Hamiltonian function

$$\begin{aligned} H_{minj} = Q(\boldsymbol{x}) + B_\gamma(\boldsymbol{x}) + 0.25 \nabla V_j^T g R^{-1} g^T \nabla V_j + \left(\nabla V_j\right)^T (f(\boldsymbol{x}) + g(\boldsymbol{x}) \\ \cdot \left(-0.5 R^{-1} g^T \nabla V_j\right)) = H_{mini} + \nabla V_d^T (f + g u_i^*) - (u_d^*)^T R u_d^* \end{aligned} \tag{5.26}$$

Since $H_{minj} - H_{mini} + (u_d^*)^T R u_d^* \geq 0$, then $\nabla V_d^T (f + g u_i^*) = \frac{d \nabla V_d^T}{dt} \geq 0$. In addition, $\lim_{t \to \infty} V_d(\boldsymbol{x}, \boldsymbol{u}_i) = 0$, thus $V_d(\boldsymbol{x}, \boldsymbol{u}_i)$ is verified to be less than 0. As a result, $V(\boldsymbol{x}, \boldsymbol{u}_j) \leq V(\boldsymbol{x}, \boldsymbol{u}_i), 0 \leq i \leq j \leq n$, $J(\boldsymbol{x}, \boldsymbol{u}_j) \leq J(\boldsymbol{x}, \boldsymbol{u}_i) \leq J(\boldsymbol{x}, \boldsymbol{u}_1)$. Since $J(\boldsymbol{x}(t), \boldsymbol{u})$ is bounded, then $r(\boldsymbol{x}, \boldsymbol{u})$ and $B_\gamma(\boldsymbol{x})$ are bounded.

Lemma 2 indicates that the $B_\gamma(\boldsymbol{x})$ remains bounded after each policy improvement step using the optimal policy $\boldsymbol{u}_j = -0.5 R^{-1} g^T \nabla V_j$, with the initial condition $\boldsymbol{x}_0 \in \text{Int}(\mathcal{S})$, and admissible feedback control policy exists. As aforementioned, the value of the RCBF function becomes infinity only at the boundary of the safe set. Therefore, it guarantees that the system states never reach the boundary of the safe set.

2) Stability analysis

The proposed safe controller should also ensure stability within the safe region defined in **Definition 4**.

Definition 4. Safe region

The safe region for the safe optimal control problem is defined as

$$D = \{\boldsymbol{x} | \boldsymbol{x} \in \text{Int}(\mathcal{S}) - \beta(\boldsymbol{x}_h^*, r_0)\} \tag{5.27}$$

where $\boldsymbol{x}_h^* = \{\boldsymbol{x} | h(\boldsymbol{x}) = 0\}$, and β is the ball around the boundary with radius of r_0 and $\boldsymbol{x}^*$ is the equilibrium point of the system.

The coefficient γ in RCBF is chosen that $B_\gamma(\boldsymbol{x}) / \big(B_\gamma(\boldsymbol{x}) + Q(\boldsymbol{x})\big) \leq 0.5$, $\boldsymbol{x} \in D$. Thus, within the safe region, $Q(\boldsymbol{x})$ is a dominant term in the optimal control problem.

Lemma 3.

Assume that $\boldsymbol{x} = 0$ is the equilibrium of the nonlinear system, and safe region

contains the origin. Let $W(t, \boldsymbol{x})$: $[0, \infty) \times \mathrm{D}$ be a continuously differentiable function such that

$$N_1 \leq W(t, \boldsymbol{x}) \leq N_2 \tag{5.28 a}$$

$$\frac{\partial W}{\partial t} + \frac{\partial W}{\partial \boldsymbol{x}}(f + g\boldsymbol{u}) \leq 0, \ \boldsymbol{x} \in D \tag{5.29 b}$$

where N_1 and N_2 are continuous positive-definite functions in safe region D. Then, the origin is uniformly stable.

From **Lemma 1** and **Lemma 2**, it can be obtained that $V(\boldsymbol{x}, \boldsymbol{u}_j) \leq V(\boldsymbol{x}, \boldsymbol{u}_1)$, $1 \leq j$, and $V(\boldsymbol{x}, \boldsymbol{u}_1)$ is bounded. Thus, N can be defined as $N = \max_{\mathrm{t}} V(\boldsymbol{x}, \boldsymbol{u}_1)$. Besides, proof for **Lemma 2** indicates that $V(\boldsymbol{x}, \boldsymbol{u}_j)$ is decreasing. Thus, using **Lemma 3**, the control system is uniformly stable.

5.4 Numerical results and discussion

5.4.1 RCBF–SDRL controller training

The objective of the RCBF incorporated SDRL controller is to maintain a stable 9 mm air gap between electromagnets and the guideway. The environment of the designed controller is the nonlinear magnetic levitation system with flexible guideway as established in Section 5.2. The initial airgap is 16 mm (Teklu and Abdissa, 2023) and the value of the system model parameters are given in **Table 5- 2**.

Table 5- 2 Parameter values of the magnetic levitation system with flexible guideway

Physical quantity	Value	Physical quantity	Value
Mass m / kg	700	Vacuum permeability μ_0 / (Hm^{-1})	$4\pi \cdot 10^{-7}$
Number of Turns of coil N	700	Area of coil A/m^2	0.024
Coil resistance R/Ω	1.2	Stable air gap x_{1eq} /m	0.009
Mass of guideway M / kg	8937.755	First mode natural frequency of guideway w	188.49564
First mode damping ratio of guideway ξ_1	0.005	-	-

For the control algorithm training, the RCBF incorporated SDRL algorithm is implemented by modifying the code of the TD3 algorithm. The structure of both the critic and actor networks in the SDRL are designed with three hidden layers. Both of the actor and critic networks use a rectified linear unit activation function (ReLU) as

the activation function, and the output layer of the actor network is processed using a tanh function. The learning rates are set as 1×10^{-3} for the critic networks and 1×10^{-4} for the actor networks. The iteration of the training is set to be 20,000 and the time step in each iteration is 500 ($\Delta t = 0.001s$). In addition, the discounted factor is 0.99 and the update parameter τ is 0.01. The exploration is done by adding noise to the actions. The whole algorithm is trained in Python 3.8 with PyTorch 1.5.1 with a mini-batch of 64 transitions sampled from a replay buffer $\mathcal{B}$ with a size of 1×10^{5}. The reward function is designed as

$$r_{sd}(\boldsymbol{x}) = -\zeta\left(x_1 - x_{1eq}\right)^2 - \dot{x}_1^2 - \log\left(\frac{\gamma_1(x_1 - x_{1min})}{1+\gamma_1(x_1 - x_{1min})}\right) - \log\left(\frac{\gamma_2(x_{1max} - x_1)}{1+\gamma_2(x_{1max} - x_1)}\right) \tag{5.30}$$

where ζ is a weight coefficient selected as 10000, $x_{1max} = 0.016$ and $x_{1min} = 6$, γ_1 and γ_2 are designed coefficients selected as 2 for this problem. The safety constraints for the levitation air gap are set between 6 mm and 16 mm in this chapter given the initial air gap of the simulation model.

In order to demonstrate the performance and ability of the proposed RCBF–SDRL algorithm, average return curves of RCBF–SDRL and normal DRL algorithms during training are given in **Figure 5–3**. As can be seen, the average return curves are obtained from 10 random seeds, and averaged over 10 consecutive episodes. It can be obviously detected from the figure that with safe constraints integrated, the convergence of the algorithm comes to be faster and more stable.

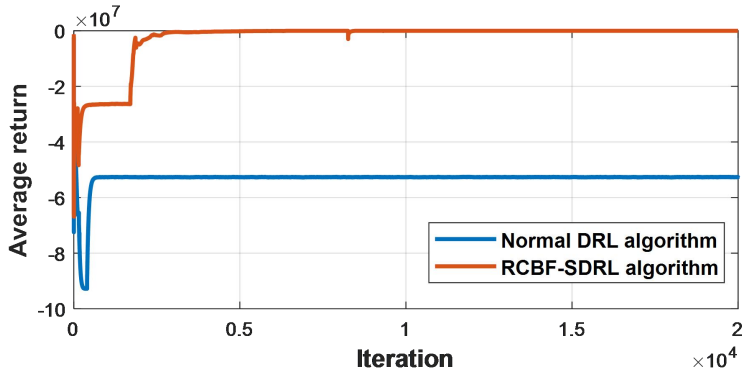


Figure 5–3 Average return curves of RCBF–SDRL and normal DRL algorithms

5.4.2 Effectiveness of the RCBF–SDRL controller

After training the RCBF–SDRL algorithm, the optimal neural network parameters obtained during the training process are set as the parameters of the RCBF–SDRL controller. This ensures that the learned knowledge and policies are utilized in the control process. To evaluate the control performance of the trained RCBF–SDRL controller, a PID controller and a GA–ST–SMC (Teklu and Abdissa, 2023) are used for comparison. In this chapter, coefficients of the PID controller are set as $K_p = 190$, $K_d = 400$ and $K_i = 1$ based on trial and error.

The GA–ST–SMC proposed by Teklu and Abdissa is designed to overcome the chattering problem of the conventional SMC. Compared with conventional one, the designed ST–SMC integrated a discontinuous controller given under SMC. The general ST–SMC is expressed as

$$u(t) = u_e(t) + u_s(t) \tag{5.31}$$

where $u_e(t)$ is obtained using conventional SMC, and $u_s(t)$ is given by

$$u_s(t) = -k_1\sqrt{|s|}sign(s) - k_2 sign(s) \tag{5.32}$$

To solve $u_e(t)$, the generalized sliding surface of magnetic levitation system is obtained as

$$s = c_1 e_1 + c_2 e_2 + c_3 e_3 + c_4 e_4 + e_5 \tag{5.33}$$

where e_1, e_2, e_3, e_4, and e_5 denote errors in vertical displacement of the electromagnet, velocity of electromagnet, vertical displacement of the guideway, velocity of the guideway, and current, respectively. Since Teklu and Abdissa still used current as control signal, the state space function of the original chapter is adopted for ST–SMC. Using the derivative of (24)

$$\dot{s} = c_1 \dot{e}_1 + c_2 \dot{e}_2 + c_3 \dot{e}_3 + c_4 \dot{e}_4 + \dot{e}_5 = 0 = \gamma(x,t) + \delta(x,t) u_e \tag{5.34}$$

Then u_e can be obtained as $u_e = -\delta(x,t)^{-1}\gamma(x,t)$. Then coefficients in the controller as k_1, k_2, c_1,... are solved using genetic algorithm (GA).

In **Figure 5–4**, the control curves of the PID, the GA–ST–SMC and the proposed RCBF–SDRL controllers are depicted. A distinct observation is the smoother convergence of the RCBF–SDRL and GA–ST–SMC controllers towards the reference air gap in contrast to the PID controller. The RCBF–SDRL and GA–ST–SMC controllers achieve convergence in approximately 0.15 s, notably faster than the PID controller which takes about 0.8 s. Moreover, both the RCBF–SDRL and PID controllers exhibit an overshoot value of 0 mm, while the GA–ST–SMC controller shows a maximum overshoot of around 0.8 mm. The comparison of these three controllers is listed in **Table 5-3**.

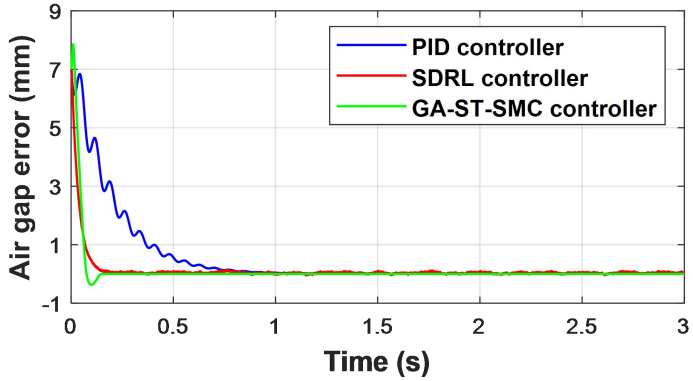


Figure 5–4 Control curves of the PID, the GA–ST–SMC, and the proposed RCBF–SDRL controllers

Table 5- 3 Comparison of control performance of three controllers

Performance criteria	RCBF–SDRL	PID	GA–ST–SMC
Settling time	**0.15 s**	0.8 s	**0.15 s**
Overshoot value	**0 mm**	**0 mm**	0.8 mm

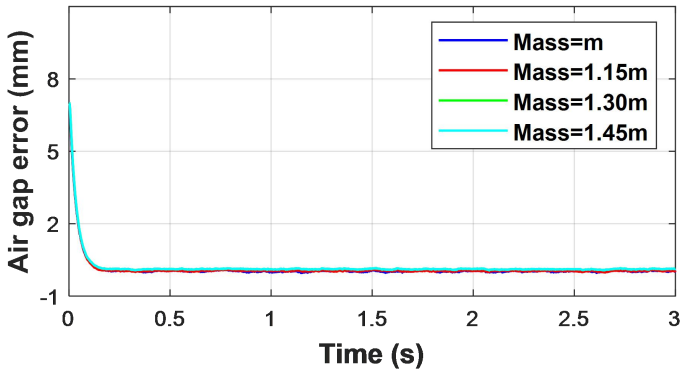
5.4.3 Robustness of the RCBF–SDRL controller

In this section, the robustness of the proposed RCBF–SDRL controller is verified under the different train load, fluctuation of the train load, random disturbance, and guideway irregularity.

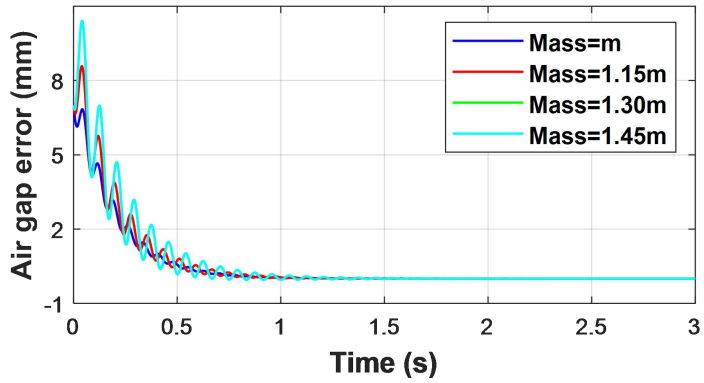
5.4.3.1 Effect of different train load

During field testing, the levitation controller is evaluated under four train load operating conditions referred to as AW0, AW1, AW2, and AW3 (DBJ50T-347-2020, 2020). To replicate these varying train load conditions, this study assigns masses for

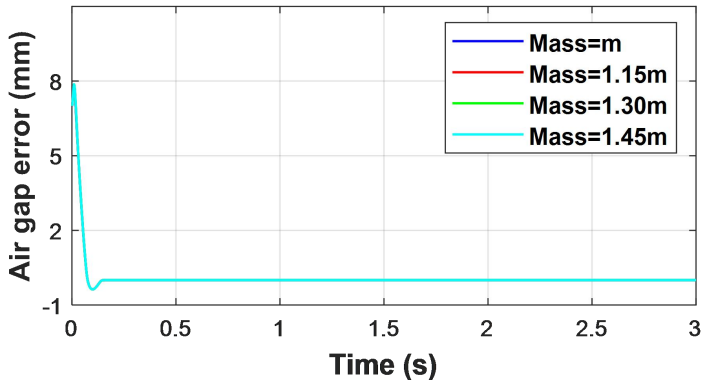
AW1, AW2, and AW3 at 115%, 130%, and 145% of AW0, respectively. The control curves depicting air gap errors for the RCBF–SDRL, GA–ST–SMC, and PID controllers under the influence of these four train loads are illustrated in **Figure 5–5**. In **Figure 5–5** (a) and (c), it is evident that the air gap error of the RCBF–SDRL and GA–ST–SMC controllers show a slight increase as the train load grows, while in **Figure 5–5** (b), the air gap error oscillations of the PID controller escalate with the train load. Additionally, the overshoot value of the PID controller spikes to approximately 2 mm, contrasting with the GA–ST–SMC controller that maintains an overshoot of around 0.9 mm. Notably, the RCBF–SDRL controller maintains 0 mm overshoot throughout. Since the PID coefficients are derived from the AW0 train load, the controller's performance may falter when the system undergoes changes. These findings suggest that both the RCBF–SDRL and GA–ST–SMC controllers exhibit greater resilience to variations in train load compared to the PID controller. However, the overshoot associated with the GA–ST–SMC controller cannot be disregarded.



(a)



(b)



(c)

Figure 5–5 Control curves of the RCBF–SDRL and PID controllers under the four train loads: (a) The RCBF–SDRL controller, (b) The PID controller, (c) The GA–ST–SMC controller

5.4.3.2 Effect of fluctuation of the train load

During operation, changes of passengers may cause sudden changes in the train load. To ensure the comfort and safety of a maglev train, the maglev system is required to adapt to these sudden changes. To further verify the robustness of the proposed RCBF–SDRL controller, this study simulates sudden changes in the train mass from 700 kg to 900 kg at 2 s that last for 0.5 s. The air gap error of the levitation

system with RCBF–SDRL, GA–ST–SMC and PID controllers are shown in **Figure 5–6** As can be seen, when the train load increases at 2 s, the change in air gap error curve using the RCBF–SDRL and GA–ST–SMC controllers can be ignored. However, there is a huge oscillation in the air gap error curve of the system using PID controller and this oscillation continues for more than 1 s. This confirms that the proposed controller and GA–ST–SMC controller are much more stable than the PID controller to fluctuations in train load.

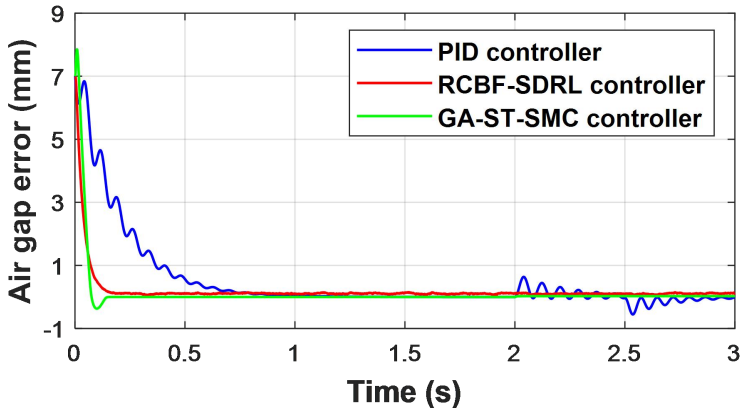


Figure 5–6 Control curves of the PID, the GA–ST–SMC, and the proposed RCBF–SDRL controllers under train load fluctuation

5.4.3.3 Effect of random disturbance

To analyze the effect of disturbance forces on the controllers' performance, multi-amplitude and multi-period sine curves are used to simulate disturbance forces, as follows:

$$f_d = 1000sin(2t + \frac{\pi}{2}) + 500sin(4t + \frac{\pi}{2}) \quad (5.35)$$

The control curves of the RCBF–SDRL, GA–ST–SMC and the PID controllers

are as in **Figure 5–7**. It indicates that the overshoot values of the PID and GA–ST–SMC controllers increase about 0.3 mm and 0.1 mm, respectively, whereas the RCBF–SDRL controller is near-negligibly affected.

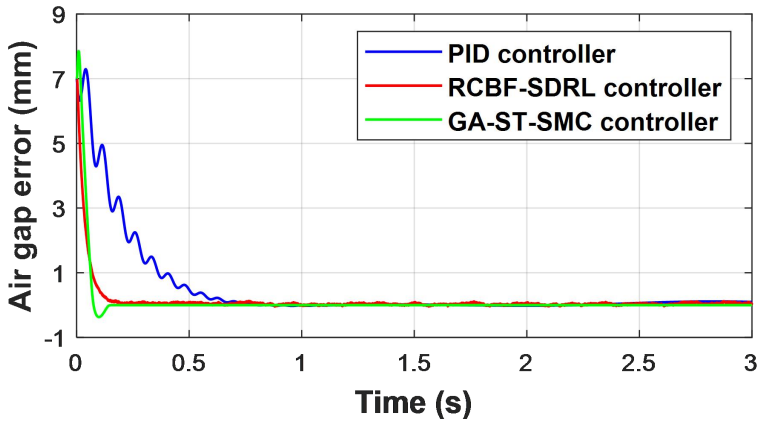


Figure 5–7 Control curves of the PID, the GA–ST–SMC, and the proposed RCBF–SDRL controllers under random disturbance

5.4.3.4 Effect of track irregularity

Track irregularity stands as a primary source of excitation in maglev systems, capable of inducing significant instability in levitation controllers by directly causing fluctuations in the air gap. In order to assess the impact of the random nature and specific features of rail irregularities on the developed RCBF–SDRL controller, the power spectrum density function introduced by Yang et al.(2004) is utilized to replicate the vertical profile variations in the guideway geometry. The control curves of the RCBF–SDRL, GA–ST–SMC and PID controllers are displayed in **Figure 5–8**. Notably, the introduction of track irregularities accentuates the differences in performance between the two controllers. The PID controller's curve shows

pronounced fluctuations, indicating challenges in maintaining stability, while the RCBF–SDRL and GA–ST–SMC controllers' curves remains notably smoother and demonstrates better resilience to the introduced track irregularities. Additionally, the overshoot value of the GA–ST–SMC controller maintains an overshoot of around 0.9 mm. Notably, the RCBF–SDRL controller maintains 0 mm overshoot throughout.

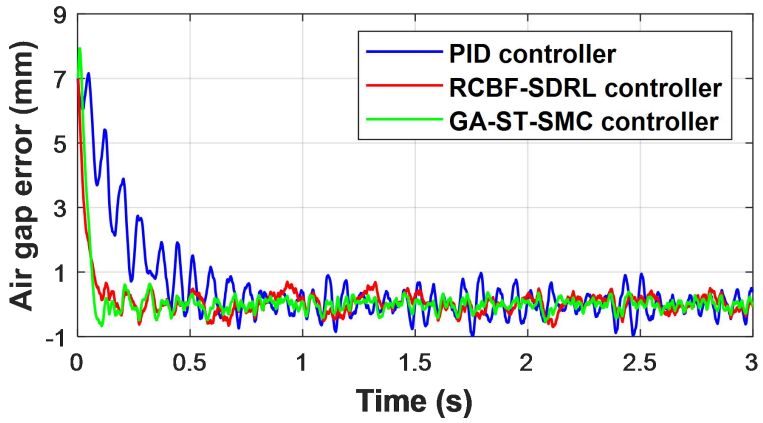


Figure 5–8 Control curves of the PID, the GA–ST–SMC, and the proposed RCBF–SDRL controllers under track irregularity

5.5 Experiment results and discussion

The effectiveness of the proposed RCBF–SDRL controller is preliminary verified through the simulation. It can also be observed that the proposed controller has better control performance than the PID controller and newly proposed GA–ST–SMC. To further verify the proposed method, experiments on the GML1001 magnetic levitation system as introduced in 3.7.1 in Chapter 3 are conducted. The safe constraints are set between 38.7 mm (initial air gap) and 34 mm for this experiment given the information of the GML1001 magnetic levitation system.

5.5.1 Comparison with PID controller

The proposed controller is firstly compared with the PID controller. The coefficient parameters of the PID control are set as $k_p = 4.5, k_i = 0.01, k_d = 50$. The experimental data from **Figure 5–9** illustrates the performance of both PID and RCBF–SDRL control methods under conditions where track irregularities are not considered. Notably, the RCBF–SDRL method achieves the target air gap in just 0.6 s, a significant improvement over the PID controller, which requires 2.5s to reach the same target air gap. Furthermore, the RCBF–SDRL controller exhibits a smaller overshoot compared to the PID controller in this scenario.

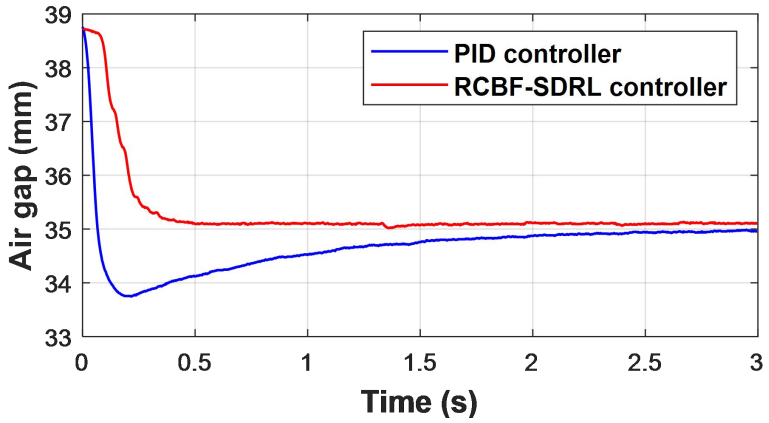


Figure 5–9 Control curves of PID and RCBF–SDRL controllers

In **Figure 5–10**, the air gap performance of the PID and RCBF–SDRL methods are depicted under track irregularities in the maglev system. In this experimental setup, a random value between -0.1 mm and 0.1 mm is introduced to the measured air gap to simulate the track irregularity. It is evident that the RCBF–SDRL control method achieves the target air gap more swiftly compared to the PID controller. Additionally, the RCBF–SDRL control demonstrates a smaller overshoot than the PID controller. These outcomes underscore the superior efficiency and robustness of the RCBF–SDRL control method over PID control.

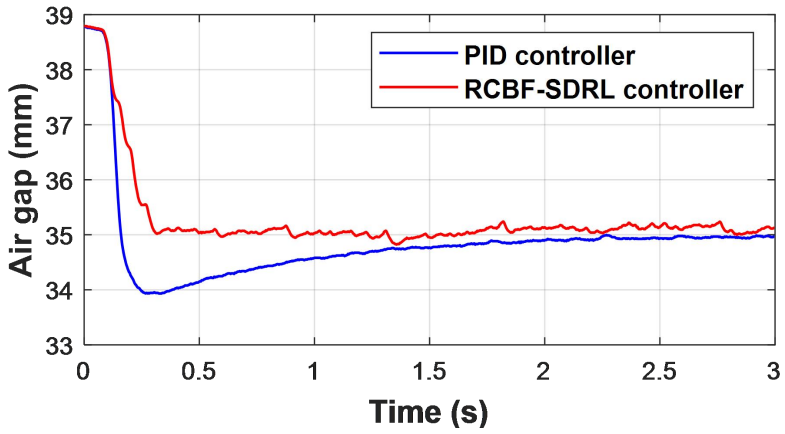


Figure 5–10 Control curves of PID and RCBF–SDRL controllers under track irregularity

5.5.2 Robustness of the RCBF–SDRL controller

In order to assess the robustness of the newly proposed RCBF–SDRL controller, experiments were conducted using balls of varying masses: 100 g, 80 g, and 150 g, mimicking different train load scenarios. The air gap curves generated by the proposed controller under these conditions are depicted in **Figure 5–11**. It can be observed that the air gap curve for the ball with a mass of 80 g exhibits greater fluctuations compared to those of the other two balls, which is attributed to the application of larger voltages, as expected. Additionally, the convergence time for the ball with a mass of 150 g is the longest, due to the smaller voltage supplied by the controller. Nevertheless, all three balls with different masses are successfully levitated to the equilibrium point within 0.5 s. These results demonstrate that, regardless of mass differences, the proposed controller consistently achieves excellent control performance. Besides, track irregularity is also introduced in these three mass scenarios, and the air gap curves are plotted in **Figure 5–12**. The conclusion can be

drawn that the proposed controller has strong robustness under different masses as well as track irregularity.

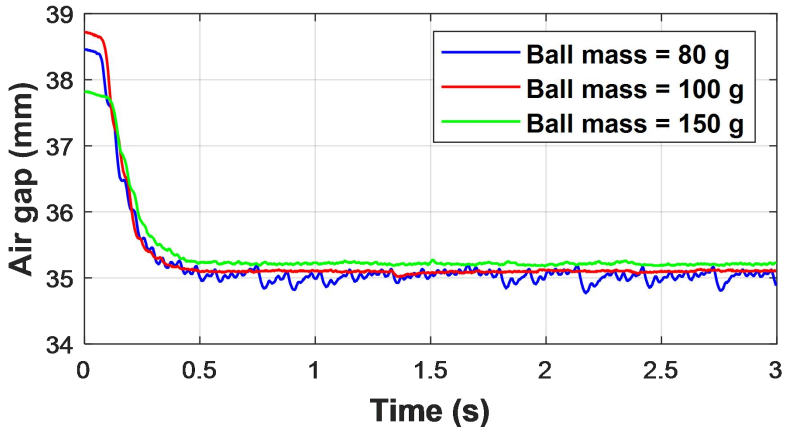


Figure 5–11 Control curves of RCBF–SDRL controllers using balls of varying masses

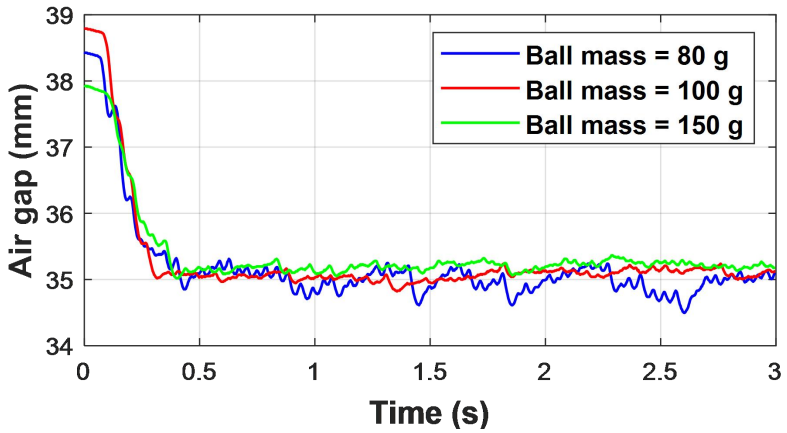


Figure 5–12 Control curves of RCBF–SDRL controllers using balls of varying masses under track irregularity

5.6 Conclusion

In this chapter, a new algorithm that integrates safety boundary RCBF into the DRL is proposed. This algorithm, named RCBF–SDRL, is designed to address the nonlinear control problem of the magnetic levitation system with flexible guideway. With the RCBF incorporated into the reward design, the RCBF–SDRL can provide the optimal and safe control for the flexible guideway coupled maglev system. The efficacy of the proposed algorithm is validated by contrasting its average return curve with that of the original DRL algorithm. The control performance of the RCBF–SDRL controller is compared with the traditional PID controller and a novel GA–ST–SMC. Additionally, the robustness of the proposed controller is evaluated through simulations that analyze its response to varying train loads, load fluctuations, external disturbances, and track irregularities. Experimental validation of the proposed method has been carried out on a magnetic levitation system to corroborate its effectiveness. The main results are as follows:

1) The convergence of the RCBF–SDRL algorithm demonstrates greater stability in average return curves compared to the original DRL algorithm.

2) In simulations, RCBF–SDRL controller achieves convergence in approximately 0.1 s, whereas the conventional PID controller takes around 0.8 s. Both the RCBF–SDRL and PID controllers exhibit zero overshoot, while the GA–ST–SMC shows an overshoot of about 0.4 mm. Moreover, the RCBF–SDRL demonstrates superior robustness compared to the PID controller when subjected to disturbances such as changing train loads, load fluctuations, external disturbances, and track irregularities.

3) During the experiment, the RCBF–SDRL controller achieves convergence in

approximately 0.6 s, whereas the PID controller takes about 2.5 s. Additionally, the robustness of the RCBF–SDRL controller is tested by varying the mass of the ball and introducing track irregularities.

CHAPTER 6 ENHANCED DEEP REINFORCEMENT LEARNING CONTROLLER FOR MAGLEV TRAIN-GUIDEWAY COUPLING SYSTEMS IN CROSSWIND CONDITIONS

The magnetic levitation (maglev) control system is crucial for maintaining the stability of the air gap between a maglev train and its guideway. Although current levitation controllers satisfy basic engineering requirements, performance issues often arise during long-term operation especially when the maglev train is exposed to crosswind conditions. To analyze the impact of crosswinds on maglev trains, a numerical model is developed in ANSYS Fluent Meshing, accounting for the complexities of the surrounding environment. This model is validated using wind tunnel experiments, and the principles of fluid mechanics similarity are applied to scale wind forces for real-world maglev train scenarios. To mitigate the effects of crosswinds on maglev trains, a safe deep reinforcement learning (SDRL) controller is proposed. In this context, safe control refers to the ability of the controller to ensure system stability and prevent unsafe states—such as excessive air gap deviations or collisions between the train and guideway—while achieving control objectives. The SDRL controller dynamically adjusts the control signals for the maglev train–guideway coupling system, thereby enhancing stability and preventing the collapse of the maglev train under adverse wind conditions. Notably, a reciprocal control barrier function (RCBF) is incorporated into the reward function of the deep reinforcement

learning (DRL) to ensure both the safety and optimality of the controller. The effectiveness of the proposed SDRL controller is demonstrated through a comparative analysis against a traditional proportional–integral–derivative (PID) controller and a genetic algorithm tuned super twisting sliding mode controller (GA–ST–SMC). This evaluation, conducted under varying crosswind and train speeds, highlights the superiority of the SDRL controller in terms of efficiency and accuracy.

6.1 Introduction

In July 2021, China successfully developed the first high-speed maglev train in the world, which can reach speeds of 600 km/h. At such high speeds, aerodynamic effects become a critical factor, significantly influencing the stability and riding comfort of maglev trains (Han et al., 2022; Che et al., 2023). Furthermore, railway transportation in coastal areas often encounters strong winds during typhoons, posing additional challenges. Ensuring the safety and stability of maglev systems under such extreme wind conditions remains one of the greatest challenges in the development of maglev transportation.

As mentioned in Chapter 3, the EMS-type maglev system faces inherent challenges due to its strong system nonlinearity, sensitivity to disturbances, and the characteristics of the magnetic circuit. As maglev train speeds increase, the influence of crosswinds becomes more pronounced, posing even greater challenges for control. Strong crosswinds can apply significant lateral and vertical forces, as well as rolling moments to the maglev train–guideway coupling system, potentially leading to a loss of control (Tian et al., 2023). Therefore, investigating the control performance of maglev trains under crosswind conditions and designing an advanced electromagnetic controller capable of mitigating the effects of crosswinds are essential to ensuring the safe and stable operation of maglev trains.

Previous research (Sun et al., 2023; Wang et al., 2023; Wang and Wang, 2024) has primarily focused on the control performance of the maglev train–guideway coupling system in windless environment and under external disturbances like track

irregularities (Yu et al., 2021) and random forces(Bu et al., 2024). Recently, some studies have begun to examine the aerodynamics of maglev transport systems. For instance, Tian et al. (2023) developed spatial analysis models of the maglev train–guideway coupling system and proposed a proportional–integral–derivative–acceleration (PIDA) control algorithm to control the maglev system. Using this model, the impact of crosswinds on the system's dynamic responses was analyzed, utilizing wind coefficients obtained from wind tunnel experiments. Wang et al. (Wang et al., 2023) constructed a full-scale numerical model of a high-temperature superconducting (HTS) high-speed maglev train controlled by a conventional feedback controller. They investigated the safety of maglev trains operating in open environments exposed to strong crosswinds, using aerodynamic force coefficients. Zhu et al. (Zhu et al., 2024) examined the unsteady aerodynamic characteristics of a high-speed maglev train (HSMT) equipped with a conventional controller during the opening process of braking plates and the stable braking stage. In addition, Huang et al. (Huang et al., 2024) analyzed the aerodynamics of maglev trains controlled by a conventional controller using the improved delayed detached eddy simulation (IDDES) method. These studies collectively highlight that wind forces play an important role in influencing the interaction behavior of maglev trains and guideway. Unlike traditional rail systems, maglev trains levitate above the guideway without mechanical contact or frictional resistance. Consequently, aerodynamic drag becomes the main source of resistance that should be surmounted. Therefore, the design of maglev control systems must account for both the fluctuation amplitude and average aerodynamic lift of each train car. This makes aerodynamic design a key yet challenging aspect of the levitation

control systems for high-speed maglev trains. However, most controllers considered in prior research rely on conventional linear control methods, such as PID controllers, which are highly sensitive to disturbances and may not provide robust performance under dynamic operating conditions.

Extensive research has been conducted to design nonlinear controllers for the maglev train–guideway coupling system to effectively handle external disturbances. Wang et al. (Wang et al., 2014) proposed a full-state feedback controller optimized using a particle swarm optimization (PSO) algorithm to determine the optimal control gains. The effectiveness of this controller was verified through simulations and test rigs under severe external disturbances. Zhou et al. (Zhou et al., 2017) introduced an innovative adaptive controller integrated with a pair of mirror FIR filters, which effectively suppressed vibration under various track irregularity scenarios, including random irregularities and sinusoidal track profiles. In another study (Sun et al., 2019) , Sun et al. developed a fuzzy adaptive tuning PID controller, capable of dynamically adjusting control gains in response to system changes. Teklu and Abdissa (2023) proposed a genetic algorithm-tuned super twisting sliding mode controller (GA–ST–SMC) and tested it under various conditions, such as different loads, external disturbances, and tracks with varying stiffness. Additionally, other advanced nonlinear controllers have been proposed to enhance the control performance of the maglev train–guideway system. These include a fuzzy adaptive controller assisted with historical database (Sun et al., 2020), robust control (Li and Shen, 2020), double loop PID integrated with control gain perturbation (Sun et al., 2023), feedback linearization control (Zhang et al., 2022), and sliding mode robust adaptive control (Chen et al., 2019). However, most of these controllers do not account for the

coupling effects of the maglev train–guideway system. Moreover, relatively limited research has addressed levitation control designs that incorporate the aerodynamic effects of maglev transport systems. This gap highlights the need for further studies to develop controllers that consider both coupling dynamics and aerodynamic influences to ensure the safe and stable operation of maglev systems under realistic conditions.

To address some of the challenging issues in traditional control fields, intelligent control methods (Wai and Lee, 2008; Fatemimoghadam et al., 2020), such as neural network (NN), convolutional neural network (CNN), and deep belief network (DBN) methods, have been applied to complex nonlinear systems. These intelligent control methods offer a degree of robustness against external disturbances and uncertainties, enhancing overall control performance. Recently, reinforcement learning (RL) algorithms have emerged as an automated framework for decision-making and control, capable of autonomously learning control policies. Due to their unique characteristics, RL algorithms have been widely applied in various domains, including video games, autonomous vehicles, and robotics.

In recent years, researchers have employed deep reinforcement learning (DRL) algorithms to tackle magnetic levitation control problems (Zhao et al., 2020; Zhu et al., 2024), demonstrating that these DRL controllers exhibit less overshoot and greater robustness compared to conventional PID and LQR controllers. In DRL, the agent uses a “trial and error” approach to explore potential actions based on the current state. Initially, a control problem is framed as a Markov Decision Process (MDP) within the DRL framework. A deep neural network (DNN) is then trained to maximize a predefined long-term reward without considering constraints (Zhou et al., 2022). Thus, the conventional DRL algorithms cannot guarantee the safety of the trained controller

under external disturbances, such as wind loads. Therefore, this chapter adopts a constrained MDP (CMDP) (Altman, 1999) framework, known as safe deep reinforcement learning (SDRL), to address the magnetic levitation control problem with safety constraints. This approach guarantees that the air gap between the maglev system and the guideway stays within a safe range when subjected to crosswinds. The main contributions of this work can be outlined as:

1) A numerical model is developed to determine the crosswind loads of maglev train–guideway coupling system in complex environments, and a wind tunnel test is carried out to validate the model.

2) A SDRL control method is proposed to manage the maglev train-guideway system under crosswinds. This method demonstrates greater robustness against external wind forces introduced to the system compared to the conventional PID controller and a genetic algorithm tuned super twisting sliding mode controller (GA–ST–SMC) .

The rest of the chapter is organized as follows. The wind tunnel test and validation of the numerical simulation is given in Section 6.2. The magnetic levitation system model and the SDRL controller design is described in Section 6.3. Section 6.4 presents numerical results and a discussion of the proposed controller by comparing the performance with the PID controller and the GA–ST–SMC under crosswinds with different speeds, respectively. Finally, the conclusion is given in Section 6.5.

6.2 Aerodynamic dynamic analysis model

6.2.1 Geometric modeling and boundary conditions

To investigate the variable aerodynamic characteristics and flow field evolution of a maglev train under crosswinds, a dynamic model of a three-car maglev train at a 1/10 scale is utilized. The geometric model is depicted in **Figure 6–1**. As shown, the maglev train model comprises a head car (HC), a middle car (MC), and a tail car (TC). To circumvent limitations related to mesh count and computational resources, the model simplifies certain train components, such as the windshield, windows, antenna box, and doors. Specifically, the train's height (H) is set at 0.4 m, and its width is 0.925H. The lengths of the HC, TC, and MC are 6.780H, 6.780H, and 6.375H, respectively. When the maglev train moves at a certain speed v_t with crosswinds impacting the left side of its direction of travel, the angle α between the crosswind speed v_w and the vehicle speed v_t in the opposite direction is referred to as the wind angle. The angle of divergence β is defined as the angle v_t between the crosswind speed and the synthesis speed v. **Figure 6–2** illustrates the schematic diagram of the different speeds and angles. To comprehensively consider various scenarios, six levels of crosswind speeds (from 5 m/s to 30 m/s) are included. Additionally, the wind angles considered are 15, 30, 45, 60, 75, and 90 degrees, and the maglev train speeds are set at 430 km/h and 600 km/h, respectively.

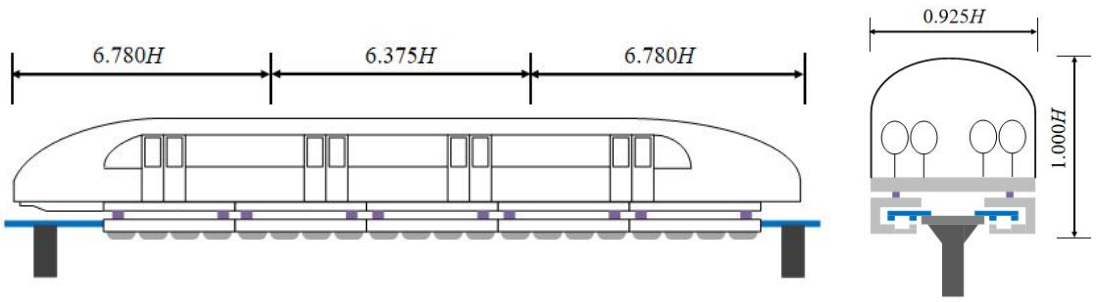


Figure 6–1 The geometric model of the maglev train

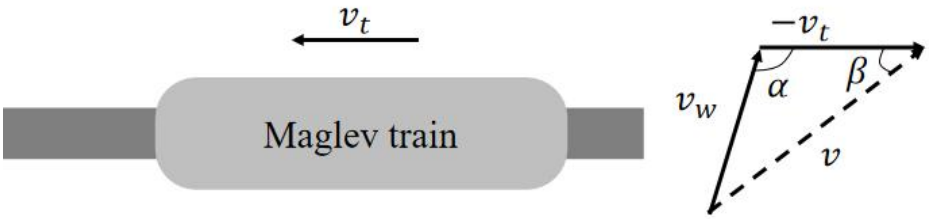


Figure 6–2 The schematic diagram of different speeds and angles

To avoid backflow during the computational process, which could compromise the accuracy of the results, the computational area is designed with dimensions of 80H in length, 40H in width, and 20H in height, respectively. Details of the computational area and related information of the boundary is illustrated in **Figure 6–3**. The surfaces BFGC and ABCD are assigned as pressure far fields, surfaces AEHD and EFGH are specified as pressure outlets, and surface ADFE is defined as a symmetry wall. Additionally, the track, viaduct, and ground are defined as moving no-slip wall surfaces to minimize their impact on the calculation results. To comply with the British Standard (BS) EN 14067 standard (B. S En, 2018), the upstream inlet is positioned 20H from the nose of the head car (HC), the downstream outlet is 50H

from the nose of the TC, and the longitudinal centerline of the viaduct is 15H from the crosswind inlet.

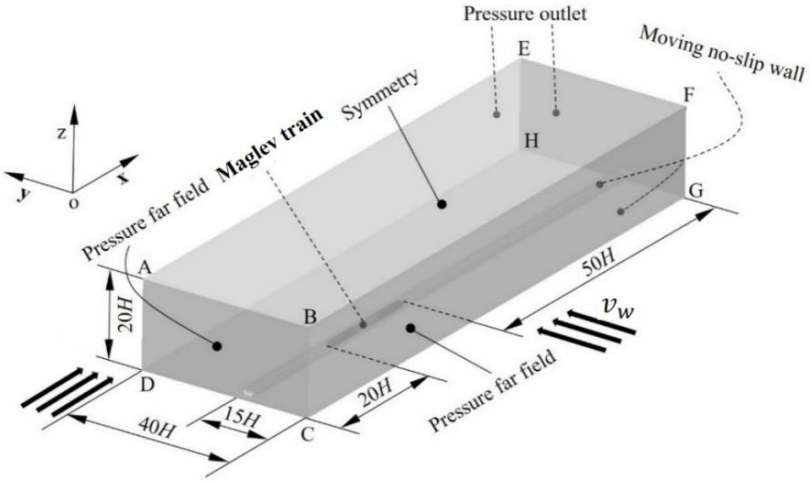


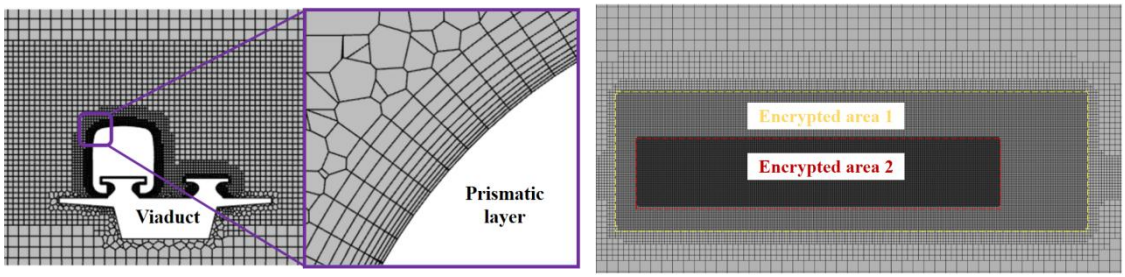
Figure 6–3 Details of the computational area and boundary conditions

6.2.2 Meshing strategy and numerical solution methods

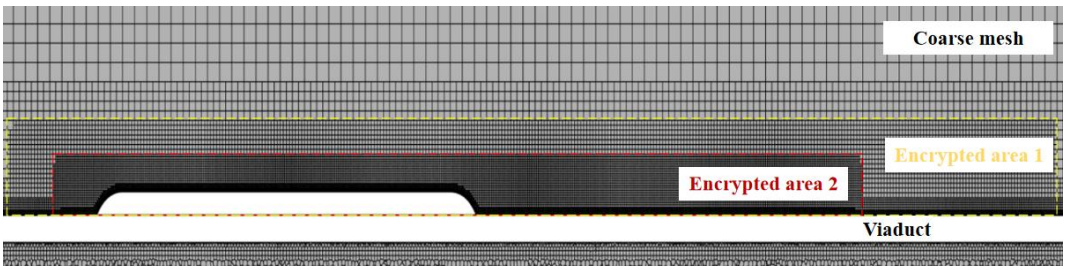
The Mosaic Poly-hexcore mesh is capable of creating a well prismatic layer mesh near the surface of the train, which gradually transitions to a coarser mesh in the flow field as it moves further away from the train. Therefore, the commercial software ANSYS Fluent Meshing is used to generate the Mosaic Poly-hexcore mesh around the train. The computational mesh of the maglev train model is depicted in **Figure 6–4**. To better capture the gradient changes in the boundary layer, the prismatic layer on the surface of the maglev train body in the computational model is divided into 10 layers, with each layer having a thickness of approximately 0.020H and a growth rate

of 1.2. To ensure a smooth transition between the prismatic layer grid and the distal outflow field grid, and to capture the vortex structures in the wake region and the leeward side of the train, the grid near the train is refined into two areas with sizes of $0.150H$ and $0.075H$, respectively.

In particular, a finite volume strategy is employed when using ANSYS Fluent Meshing for numerical simulation. Specifically, the Green-Gauss method and the Semi-Implicit Method were adopted to calculate scalar gradient and coupled pressure and velocity equations (Zhang et al., 2023; Chen et al., 2022; Chen et al., 2023). The energy equation is solved using the second-order upwind scheme, while the momentum equation is addressed with the bounded central differencing scheme (Zhang et al., 2022). For transient equations, the second-order implicit scheme is utilized (Chen et al., 2023). In the non-stationary numerical simulation, the physical time step is set as $1\mathrm{e}-4$ s, and the built-in time step is set to 30. The residual values for each solved equation are below $1\mathrm{e}-4$, meeting the convergence criteria requirements.



(a) (b)



(c)

Figure 6–4 Computational mesh of the maglev train model: (a) front view, (b) top view, (c) grid around the maglev train

6.2.3 Verification of the numerical model

To validate the numerical model, a wind tunnel test was carried out at the Hong Kong Polytechnic University. The wind tunnel laboratory features a closed-loop low-speed wind tunnel with a test section measuring 2.4 m × 0.6 m × 0.6 m. The maximum wind speed achievable is 50 m/s. The pressure on the train model was monitored under various crosswind speeds and yaw angles using pressure scanning valves to ensure reliable data collection. The schematic of the closed-loop low-speed wind tunnel is given in **Figure 6–5**.

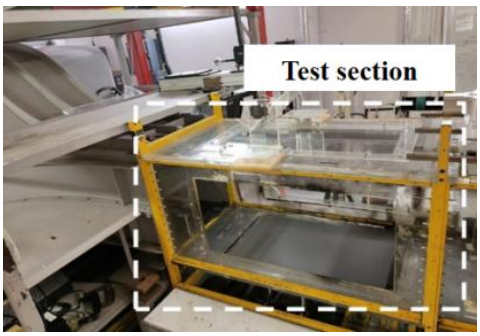


Figure 6–5 The schematic of the closed-loop low-speed wind tunnel: (a) Test section, (b) Inside of the test section, (c) Entrance of the test section, (d) Outlet of the test section

To ensure that the operation of the train model corresponds to the Reynolds number of an actual maglev train, while also considering the size constraints of the wind tunnel, the model shrinkage ratio was set to 1:40, as shown in **Figure 6–6**. The model measures 678 mm in length, 100 mm in height, and 92.5 mm in width. At yaw angles of 5, 10, 15, and 20 degrees, the obstruction ratios between the train and the wind tunnel are 1.67%, 2.51%, 3.27%, and 3.85%, respectively. **Figure 6–7** depicts the layout of the pressure holes on the surface of the train. These holes, each with a diameter of 1 mm to match the pressure scanning valves, are uniformly distributed and perpendicular to the model's surface. There are 63 holes on the train surface, measurement points 1-23 and 58-63 are located on the vertical plane of the train model, while the others are symmetrically distributed on both sides of the train. The pressure measuring equipment used is the ESP 64-channel pressure scanning valve of

TE Connectivity. Data collection and processing were carried out using the DTC INITIUM host computer and its associated PSI software, with a sampling frequency of 333 Hz. To minimize the length of the pressure measuring tube and reduce its impact on the flow field, the pressure scanning valve was placed at the bottom of the wind tunnel.

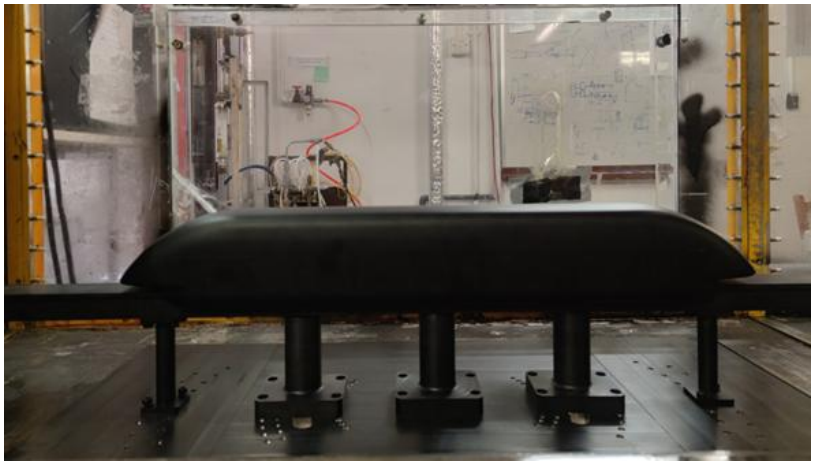


Figure 6–6 Model of the maglev train

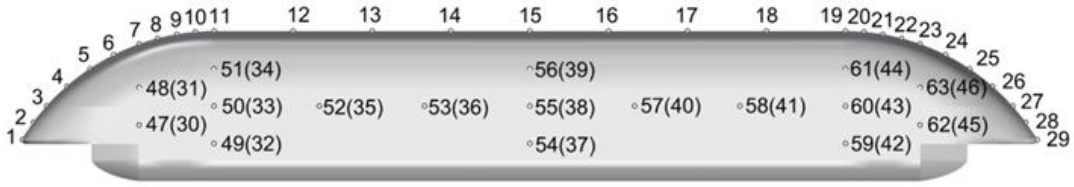
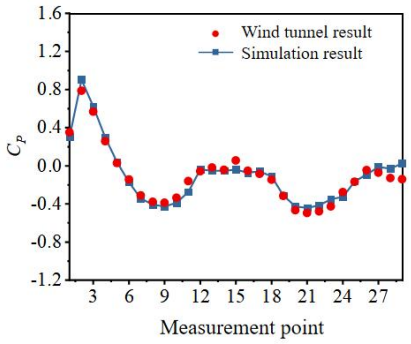


Figure 6–7 The arrangement of the pressure holes on the train surface

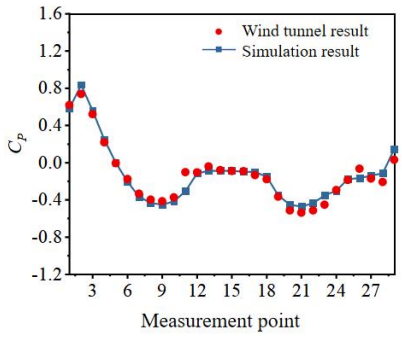
The pressure coefficients at the train surface measurement points from both the wind tunnel test and numerical simulation are shown in **Figure 6–8**. As observed, the longitudinal centerline pressure coefficient distribution curves at different yaw angles

(5, 10, 15, and 20 degrees) obtained from the numerical results are consistent with those from the wind tunnel test. Additionally, the pressure coefficients on the train's leeward side from the numerical solution at various yaw angles also align with the wind tunnel test results, with the pressure distribution characteristics curve mirroring that of the train's longitudinal centerline. The deviation between the numerical results and the wind tunnel test results is negligible. Therefore, the established numerical model is sufficiently accurate for subsequent analysis.

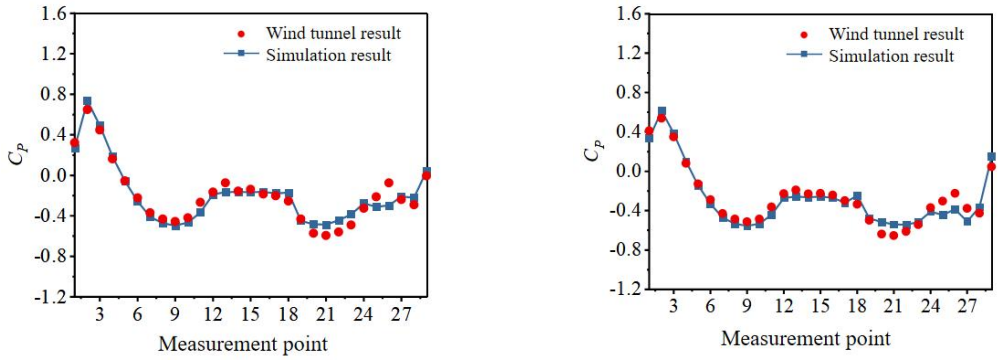
To assess the effect of crosswinds on the maglev train–guideway coupling system, which is controlled by the advanced SDRL controller, a fluid mechanics similarity criterion is employed to determine the crosswind forces using the validated numerical model.



(a)



(b)



(c) (d)

Figure 6–8 Pressure coefficient C_p of the upper surface along longitudinal centerline of the maglev train by simulation and wind tunnel test with wind speed as 15 m/s: (a) a yaw angle of 5°, (b) a yaw angle of 10°, (c) a yaw angle of 15°, (d) a yaw angle of 20°

6.3 Mathematical model of maglev system and SDRL controller design

6.3.1 Modeling of magnetic levitation system with flexible guideway

To implement the SDRL control algorithm, the original maglev train–guideway coupling system is simplified to a single-point model, where an electromagnet levitation system is coupled with a simply supported beam. This simplification helps avoid overly complex system analysis (Teklu and Abdissa, 2023). The configuration of the simplified maglev train–guideway coupling system is illustrated in **Figure 6–9**, which consists of maglev trains and an elevated guideway. In the appropriate control mode, the levitation electromagnets are instructed to generate an attractive force that adjusts the air gap between the electromagnets and the guideway to maintain it around a reference value, e.g., 9 mm.

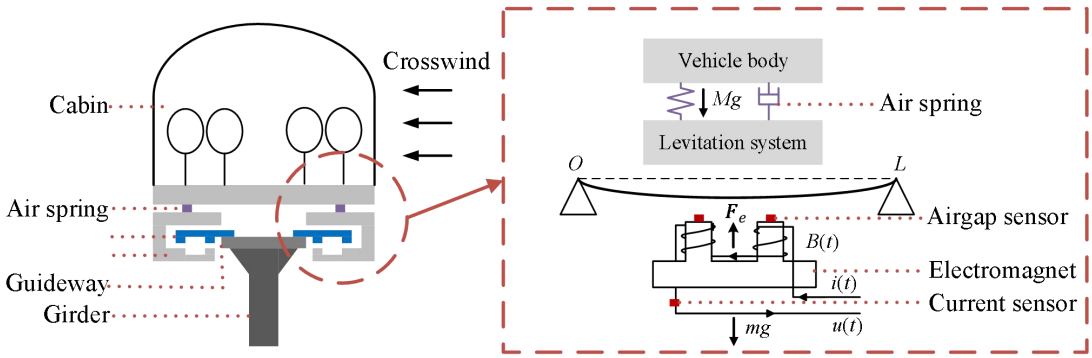


Figure 6–9 Cross-section of an EMS-type maglev system and a schematic of a single EMS module with flexible guideway

In the simplified model, the guideway is represented as a Bernoulli-Euler beam,

as its length is significantly greater than its other dimensions. The nonlinear behavior of the guideway is disregarded because the vibration amplitude is sufficiently small compared to the span of the guideway.

The motion of the guideway can be described by the following differential equation as

$$E_r I_r \frac{\partial^4 x_r}{\partial y^4} + \rho_r \frac{\partial^2 x_r}{\partial t^2} = f(y,t) \tag{6.1}$$

$$\text{where } f(y,t) = \begin{cases} F_e/l_e, y_0 \le y \le y_0 + l_e \\ 0, else \end{cases}$$

where x_r is the vertical displacement of the guideway along the y-axis; E_r and I_r denote the Young's modulus of elasticity and the cross-sectional moment of inertia of the guideway beam, respectively; T_r represents the tension generated when the beam is deformed; ρ_r is the linear mass density of the beam; c denotes the damping coefficient of the beam; k_r is the elasticity coefficient when the beam is elastically deformed; $f(y,t)$ is the distributed load density when the maglev train passes through; F_e denotes the electromagnetic levitation force; and l_e is the effective length of a single levitation module. For the simply supported concrete beam, the mathematical model of the flexible guideway's vertical displacement can be expressed as

$$x_r = \sum_n \varphi_n(y) q_n(t), n = 1, 2, \ldots, \infty \tag{6.2}$$

$$\varphi_n(y) = \sqrt{\frac{2}{M_r}} sin\left(\frac{n\pi y_0}{L}\right) \tag{6.3}$$

$$\ddot{q}_n(t) + 2\xi_n \left(\frac{n\pi}{L}\right)^2 \sqrt{\frac{E_r I_r}{\rho_r}} \dot{q}_n(t) + \left(\frac{n\pi}{L}\right)^4 \frac{E_r I_r}{\rho_r} q_n(t) = \sqrt{\frac{2}{M_r}} sin\left(\frac{n\pi y_0}{L}\right) F_e \tag{6.4}$$

where $\varphi_n(y)$ denotes the *n* th order mode shape function corresponding to the simply supported beam; $q_n(t)$ is the generalized coordinate corresponding to the mode shape function at time *t*; M_r represents the mass of the guideway, L is the span length of the guideway beam, and ξ_n is the damping ratio of the *n* th order mode.

As the first mode accounts for the majority of the response, it is often the dominant factor in the overall behavior of the system. Therefore, x_r can be regarded as the displacement of the first mode, expressed as

$$x_r = \sqrt{\frac{2}{M_r}} sin\left(\frac{\pi y_0}{L}\right) q_1(t) \tag{6.5}$$

Then the Equation (6.4) can be rewritten as follows

$$\ddot{q}_1(t) + 2\xi_1\left(\frac{\pi}{L}\right)^2 \sqrt{\frac{E_r I_r}{\rho_r}}\dot{q}_1(t) + \left(\frac{\pi}{L}\right)^4 \frac{E_r I_r}{\rho_r} q_1(t) = \sqrt{\frac{2}{M_r}} sin\left(\frac{\pi y_0}{L}\right) F_e \tag{6.6}$$

By replacing the q_1 in the Equation (6) using x_r in the Equation (6.5), the Equation (6.6) can be expressed as

$$\ddot{x}_r(t) + 2\xi_1\left(\frac{\pi}{L}\right)^2 \sqrt{\frac{E_r I_r}{\rho_r}}\dot{x}_r(t) + \left(\frac{\pi}{L}\right)^4 \frac{E_r I_r}{\rho_r} x_r(t) = \frac{2}{M_r} sin^2\left(\frac{\pi y_0}{L}\right) F_e \tag{6.7}$$

The vertical displacement of the guideway can be obtained by solving Equation (6.7).

It can be observed from **Figure 6–9** that the levitation air gap x_e with the flexible guideway can be obtained as

$$x_e = x_m - x_r \tag{6.8}$$

where x_m denotes the vertical displacement of the electromagnet.

The magnetic force $F_e\ (i, x_e)$ according to Maxwell's equation and Biot-Savart's theorem is as

$$F_e\,(i, x_e) = \frac{\int_0^t \psi_e(i,\ x_e)dt}{\partial x_e} \tag{6.9}$$

where i denotes electromagnet current. According to Kirchhoff magnetic-circuit law, $\psi_e(i, x_e)$ can be obtained as

$$\psi_e(i, x_e) = \frac{N^2 i(t)}{R(x_e)} \tag{6.10}$$

where reluctance $R(x_e) = 2x_e(t)/(\mu_0 A_e)$, N denotes the coil number of turns, μ_0 is air permeability, and A_e denotes the effective magnetic pole area. The air gap magnetic flux density of the levitation electromagnet is as follows

$$B(t) = \frac{\mu_0 N i(t)}{2x_e} \tag{6.11}$$

Then the expression of the magnetic force can be obtained as

$$F_e\,(i, x_e) = \frac{B(t)^2 A_e}{\mu_0} \tag{6.12}$$

Using Newton's second law, the dynamics equation of the levitation system considering the crosswind can be expressed as

$$m\ddot{x}_m(t) = (m + M)g - \frac{B(t)^2 A_e}{\mu_0} + f_{wf} \tag{6.13}$$

where M is mass of the cabin, m is mass of the electromagnet, and f_{wf} is the lift force and overturning moment of the crosswind.

The magnetic levitation system model with a flexible guideway considering the effect of the crosswind, based on the first-order vibration of the flexible guideway, is developed by integrating the vertical motion model of the guideway structure with the vertical motion of the electromagnet, as follows

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\dfrac{A_e}{m\mu_0}B(t)^2 + \dfrac{m+M}{m}g + f_{wf} \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = \dfrac{2}{M_r}sin^2\left(\dfrac{\pi y_0}{L}\right)\dfrac{A_e}{\mu_0}B(t)^2 - 2\xi_1\left(\dfrac{\pi}{L}\right)^2\sqrt{\dfrac{E_rI_r}{\rho_r}}x_4 - \left(\dfrac{\pi}{L}\right)^4\dfrac{E_rI_r}{\rho_r}x_3 \end{cases} \tag{6.14}$$

where $[x_1, x_2, x_3, x_4] = [x_m, \dot{x}_m, x_r, \dot{x}_r]$, and control signal is $B(t)$. Assuming that

$$k_1 = \frac{A_e}{\mu_0}, k_2 = \frac{2}{M_r}sin^2\left(\frac{\pi y_0}{L}\right), k_3 = \left(\frac{\pi}{L}\right)^2\sqrt{\frac{E_rI_r}{\rho_r}}$$

Then the nonlinear state space function can be expressed as

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\dfrac{k_1}{m}B(t)^2 + \dfrac{m+M}{m}g + f_{wf} \\ \dot{x}_3 = x_4 \\ \dot{x}_4 = k_1k_2B(t)^2 - 2\xi_1k_3x_4 - {k_3}^2x_3 \end{cases} \tag{6.15}$$

6.3.2 SDRL controller design

The primary control objectives of the maglev train – guideway coupling system are as follows: (1) to ensure that the air gap between the electromagnets and the guideway converges to and remains at a specified reference value, even in the presence of crosswinds; and (2) to minimize energy consumption. This chapter utilizes an SDRL algorithm to tackle this challenge. To formulate the problem, a Constrained Markov Decision Process (CMDP) is constructed. A CMDP comprises five key elements: a state space S; an action space A; an immediate or instantaneous reward R; transition dynamics P that maps a state-action pair at time t into a distribution of state at time $t + 1$; and a cost function C. The state of the CMDP is designed to fully capture the system's behavior and facilitate the calculation of new

states. We define the vertical displacement of the electromagnet d and velocity of the displacement v as the state, i.e., $s_t = [d(t), v(t)]$. The control signal, magnetic flux density of the levitation electromagnet $B(t)$, serves as the action a_t. In line with the objective, the reward r_t is defined as $r_t = -(d(t) - d_{eq})^2 - a_t{}^2$, where d_{eq} denotes the reference air gap. The constraints of the problem are represented as $d_{max} \geq d(t) \geq d_{min}$. accordingly, the cost functions are defined as $c_t^1 = d_{max} - d(t)$, and $c_t^2 = d(t) - d_{min}$.

The control system of the maglev train–guideway coupling system is treated as an agent that makes decision based on the current state of the system s_t. The action taken by the agent is defined as $a_t = \pi(s_t)$, where $\pi(\cdot)$ is the decision-making function At each time step, the agent observes the state s_t, performs action a_t according to the current policy, and receive scalar rewards r_t as along with costs c_t^1 and c_t^2 from the environment after the system transition occurs. The tuple $\left(s_t, a_t, r_t, c_t^{i\,(i=1,2)}, s_{t+1}\right)$ is then stored in a replay buffer $\mathcal{B}$. During training, the SDRL algorithm samples from $\mathcal{B}$ to update the policy. Ultimately, the controller aims to discover an optimal policy π^* that maximizes the the total amount of the received reward (return) $U(t) = \sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k}, 0 \leq \gamma \leq 1$ while adhering to the boundary constraints. Here, γ is the discount rate that determines the present value of future rewards.

To simplify the CMDP problem, a reciprocal control barrier function (RCBF) is integrated with the reward function, thereby eliminating the cost terms. Given a safe closed set $\mathcal{S} \subset \mathrm{R}^n$ as

$$\mathcal{S} = \{x(t) | h(x(t)) \geq 0\} \tag{6.16a}$$

$$\partial\mathcal{S} = \{x(t) | h(x(t)) = 0\} \tag{6.16b}$$

$$\text{Int}(\mathcal{S}) = \{x(t) | h(x(t)) > 0\} \tag{6.16c}$$

where $h(x(t))$ is a continuously differentiable function of $x(t)$.

In the RCBF, the value of the barrier function (BF) $B_\gamma(x) \to \infty$ as $x \to \partial\mathcal{S}$ can grow when it is far away from the boundary of $\mathcal{S}$. To solve this problem, the BF needs to fulfill a requirement that $\dot{B}_\gamma(x) \leq \alpha\big(1/B_\gamma(x)\big)$, where α is a class $\mathcal{K}$ function stated in **Definition 1**.

Definition 1. Class $\mathcal{K}$ function

A continuous function $\alpha: [0, a) \to [0, \infty)$ is a class $\mathcal{K}$ function if it strictly increasing and $\alpha(0) = 0$.

In conventional BFs, $\dot{B}_\gamma(x) \leq 0$ is enforced (Tee et al., 2009; Prajna et al., 2007), but this may not be desirable since it requires all sub-level sets of $\mathcal{S}$ to be invariant. Thus, the condition is relaxed to be $\dot{B}_\gamma(x) \leq \gamma/B_\gamma(x)$, where γ is positive. In a more general context, RCBF can be formulated as in **Definition 2**.

Definition 2. Reciprocal control barrier function (RCBF)

For a nonlinear dynamic system, a continuously differentiable function $B_\gamma(x)$: $\text{Int}(\mathcal{S}) \to \text{R}$ is considered a RCBF for the safe set $\mathcal{S}$ defined in Equation (1) for the continuously differentiable function $h(x)$, if there exist class $\mathcal{K}$ functions α_1, α_2, and α_3 such that for all $x \in \text{Int}(\mathcal{S})$:

$$\frac{1}{\alpha_1(h(x))} \leq B_\gamma(x) \leq \frac{1}{\alpha_2(h(x))} \tag{6.17a}$$

$$\dot{B_\gamma}(x) \leq \alpha_3(h(x)) \tag{6.17b}$$

A logarithmic barrier function candidate $B_\gamma(\boldsymbol{x}) = -\log(h(\boldsymbol{x})/(1+h(\boldsymbol{x})))$ is employed which satisfies the important properties as $\inf_{x\in \mathrm{Int}\mathcal{S}} B_\gamma(x) \geq 0$, $\lim_{x\to\partial\mathcal{S}} B_\gamma(x) = \infty$. To determine the relative dominance of the RCBF compared to the reward function r_t, $B_\gamma(x)$ is modified to be

$$B_\gamma(x) = -\log(\beta h(x)/(1+\beta h(x))) \tag{6.18}$$

where coefficient β balances safety and optimality by defining the extent to which safety takes precedence over other control objectives. The origin reward function r_t is modified to be

$$r_t = -(d(t)-d_{eq})^2 - {a_t}^2 - B_\gamma(d(t)) \tag{6.19}$$

In proposed formulation, safety is guaranteed while achieving the optimal control objective. The incorporated RCBF $B_\gamma(x)$ serves as a safety measure alongside the optimization of other control objectives. All these goals, including safety and the optimization of additional objectives, should be achieved iteratively as the value function can not be solved directly.

As mentioned earlier, the RCBF is the dominant term near the risky area. As a result, samples from the safe region and the safe boundary should both be collected for training. For the SDRL algorithm, an actor-critic DRL algorithm named twin delayed deep deterministic policy gradient (TD3) (Fujimoto et al., 2017) method is adopted, which enhances stability and performance while accounting for function approximation error. The detailed SDRL algorithm is given in **Algorithm 1**, and the complete schematic diagram of the algorithm is presented in **Figure 6–10**.

Algorithm 1 Pseudo code of proposed SDRL algorithm

Initialize critic networks $Q(s,a;\boldsymbol{w}_1), Q(s,a;\boldsymbol{w}_2)$, and actor network $\mu(s;\boldsymbol{\theta})$ with random parameters w_1, w_2, θ.

Initialize target critic networks $Q(s,a;\boldsymbol{w}_1^-), Q(s,a;\boldsymbol{w}_2^-)$, and actor network $\mu(s;\boldsymbol{\theta}^-)$ with parameters $\boldsymbol{w}_1^- \leftarrow \boldsymbol{w}_1, \boldsymbol{w}_2^- \leftarrow \boldsymbol{w}_2, \boldsymbol{\theta}^- \leftarrow \boldsymbol{\theta}$.

Initialize Replay buffer $\mathcal{B}$.

For iteration =1, 2, ..., T

Select action with exploration noise $a_t = \mu(s_t;\boldsymbol{\theta}) + \epsilon,\ \epsilon \sim \mathcal{N}(0,\sigma)$ and obtain RCBF incorporated reward r_t and next state s_{t+1}, then store the transition (s_t, a_t, r_t, s_{t+1}) in the replay buffer $\mathcal{B}$.

Sample mini-batches of N transitions from the replay buffer $\mathcal{B}$

$$a_{j+1}^- = \mu(s_{j+1};\boldsymbol{\theta}^-) + \epsilon, \epsilon \sim clip(\mathcal{N}(0,\sigma), -c, c)$$

$$Q_{1,j+1}^- = Q(s_{j+1}, a_{j+1}^-;\boldsymbol{w}_1^-), Q_{2,j+1}^- = Q(s_{j+1}, a_{j+1}^-;\boldsymbol{w}_2^-)$$

$$Q_{1,j} = Q(s_j, a_j;\boldsymbol{w}_1), Q_{2,j} = Q(s_j, a_j;\boldsymbol{w}_2)$$

TD target: $\eta_j = r_j + \gamma \cdot min\{Q_{1,j+1}^-, Q_{2,j+1}^-\}$

TD error: $\delta_{1,j} = Q_{1,j} - \eta_j, \delta_{2,j} = Q_{2,j} - \eta_j$

Update of critic networks:

$$\boldsymbol{w}_1 \leftarrow \boldsymbol{w}_1 - \alpha \cdot \delta_{1,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_1)$$

$$\boldsymbol{w}_2 \leftarrow \boldsymbol{w}_2 - \alpha \cdot \delta_{2,j} \cdot \nabla_{\boldsymbol{w}} Q(s_j, a_j; \boldsymbol{w}_2)$$

If iteration mod k, **then**

Update actor network:

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \beta \cdot \nabla_{\boldsymbol{\theta}} \mu(s_j; \boldsymbol{\theta}) \cdot \nabla_a Q(s_j, a_j; \boldsymbol{w}_1)$$

Update target networks:

$$\boldsymbol{\theta}^- \leftarrow \tau\boldsymbol{\theta} + (1-\tau)\boldsymbol{\theta}^-$$

$$\boldsymbol{w}_1^- \leftarrow \tau\boldsymbol{w}_1 + (1-\tau)\boldsymbol{w}_1^-$$

$$\boldsymbol{w}_2^- \leftarrow \tau\boldsymbol{w}_2 + (1-\tau)\boldsymbol{w}_2^-$$

End if

End for

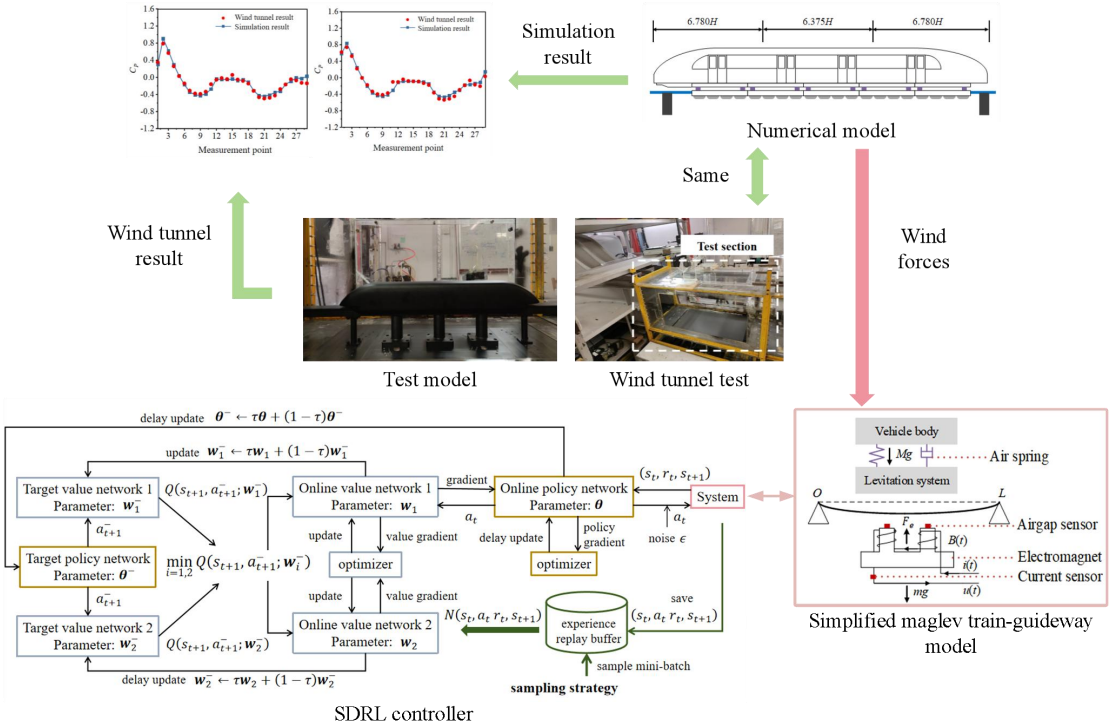


Figure 6–10 Schematic diagram of the SDRL algorithm

6.3.3 Safety and stability analysis of the controller

In this section, we will analyze the safety of the system, as well as the stability through Lyapunov function.

1) Safety analysis

To analyze the safety of the system, we first need to verify the existence of the value function and demonstrate the boundedness of the RCBF. Based on the proved results, safety is guaranteed. Before proceeding, admissible feedback control policy needs to be defined as in **Definition 3**.

Definition 3. Admissible feedback control policy

A control policy is considered admissible for a safe optimal control problem if it

stabilizes the system and ensures that the associated cost has boundness. It is defined as follow:

$$\boldsymbol{u} \in \mathcal{U}_a = \mathcal{U} \cap \mathcal{U}_{safe}$$

where $\mathcal{U}$ represents the admissible control policy and $\mathcal{U}_{safe}$ is a set of safe inputs as $\mathcal{U}_{safe} = \{\boldsymbol{u} \in \mathrm{R}^m | \boldsymbol{x^u} \in \mathrm{Int}(\mathcal{S})\}$, and $\boldsymbol{x^u}$ refers to the state of the system as influenced by the input $\boldsymbol{u}$.

Lemma 1.

Consider an admissible feedback control policy $\boldsymbol{u}_1 \in \mathcal{U}_a$. If a time invariant positive function $V \in \mathrm{C}^1$ exists such that

$$\frac{\partial V^T}{\partial \boldsymbol{x}}(f(\boldsymbol{x}) + g(\boldsymbol{x})\boldsymbol{u}_1) + Q(\boldsymbol{x}) + B_\gamma(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 = 0, V(\boldsymbol{x}_0, \boldsymbol{u}_1) = J(\boldsymbol{x}_0, \boldsymbol{u}_1) \quad (6.20)$$

Then V can be regarded as the value function of the safe optimal control system for all time $t \in [0, \infty)$, $V(\boldsymbol{x}, \boldsymbol{u}) = J(\boldsymbol{x}, \boldsymbol{u})$.

Proof for Lemma 1.

Assume $V(\boldsymbol{x}, \boldsymbol{u}_1) > 0$ exists, then we have

$$\mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}_0, \boldsymbol{u}_1) = \int_0^t \frac{\partial V}{\partial \boldsymbol{x}(\tau)}(f(\boldsymbol{x}(\tau)) + g(\boldsymbol{x}(\tau))\boldsymbol{u}_1)d\tau \quad (6.21)$$

Using $J(\boldsymbol{x}(t), \boldsymbol{u}) = \int_t^\infty Q(\boldsymbol{x}) + \boldsymbol{u}^T R \boldsymbol{U} + B_\gamma(\boldsymbol{x})d\tau$, we have

$$(\boldsymbol{x}(t), \boldsymbol{u}_1) - J(\boldsymbol{x}_0, \boldsymbol{u}_1) = -\int_0^t Q(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 + B_\gamma(\boldsymbol{x})d\tau \quad (6.22)$$

Based on (6.21) and (6.22),

$$\begin{aligned} J(\boldsymbol{x}(t), \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1) - (J(\boldsymbol{x}_0, \boldsymbol{u}_1) - \mathrm{V}(\boldsymbol{x}_0, \boldsymbol{u}_1)) \\ = -\int_0^t Q(\boldsymbol{x}) + \boldsymbol{u}_1^T R \boldsymbol{u}_1 + B_\gamma(\boldsymbol{x}) + \frac{\partial V}{\partial \boldsymbol{x}(\tau)}(f(\boldsymbol{x}(\tau)) \\ + g(\boldsymbol{x}(\tau))\boldsymbol{u}_1)d\tau \end{aligned} \quad (6.23)$$

Using (6.20) in **Lemma 1**, we have $J(\boldsymbol{x}(t), \boldsymbol{u}_1) = \mathrm{V}(\boldsymbol{x}(t), \boldsymbol{u}_1)$. The proof is completed.

Lemma 2.

Assume positive value functions $V(\boldsymbol{x}, \boldsymbol{u}_1)$, $V(\boldsymbol{x}, \boldsymbol{u}_2)$, ..., $V(\boldsymbol{x}, \boldsymbol{u}_n)$ are associated with admissible control policy sequence $\boldsymbol{u}_1$, $\boldsymbol{u}_2$, ..., $\boldsymbol{u}_n \in \mathcal{U}_a$. If corresponding minimized Hamiltonian values satisfy $H_{min1} \leq H_{min2} \leq \ldots \leq H_{minn}$, then the RCBF term at each time step is bounded. The Hamiltonian function is defined as

$$H_i(x, \boldsymbol{u}_i, \nabla V_i) = r(\boldsymbol{x}, \boldsymbol{u}_i) + (\nabla V_i)^T (f(\boldsymbol{x}) + g(\boldsymbol{x}) \cdot \boldsymbol{u}_i) \tag{6.24}$$

Then the minimized Hamiltonian function is given as $H_{mini} = H_i(x, \boldsymbol{u}_i^*, \nabla V_i)$.

Proof for Lemma 2.

For any *i* and *j* that fulfill $0 \leq i \leq j \leq n$, assume that $H_{mini} \leq H_{minj}$, given

$$V(\boldsymbol{x}, \boldsymbol{u}_j) = V(\boldsymbol{x}, \boldsymbol{u}_i) + V_d(\boldsymbol{x}, \boldsymbol{u}_i) \tag{6.25}$$

Then, optimal policy $\boldsymbol{u}_j = -0.5R^{-1}g^T \nabla V_j$ is adopted to replace the safe input term in the minimized Hamiltonian function

$$\begin{aligned} H_{minj} = Q(\boldsymbol{x}) + B_\gamma(\boldsymbol{x}) + 0.25\nabla V_j^T g R^{-1} g^T \nabla V_j + \left(\nabla V_j\right)^T (f(\boldsymbol{x}) + g(\boldsymbol{x}) \\ \cdot \left(-0.5R^{-1}g^T \nabla V_j\right)) = H_{mini} + \nabla V_d^T (f + g u_i^*) - (u_d^*)^T R u_d^* \end{aligned} \tag{6.26}$$

Since $H_{minj} - H_{mini} + (u_d^*)^T R u_d^* \geq 0$, then $\nabla V_d^T (f + g u_i^*) = \frac{d \nabla V_d^T}{dt} \geq 0$. In addition, $\lim_{t \to \infty} V_d(\boldsymbol{x}, \boldsymbol{u}_i) = 0$, thus $V_d(\boldsymbol{x}, \boldsymbol{u}_i)$ is verified to be less than 0. As a result, $V(\boldsymbol{x}, \boldsymbol{u}_j) \leq V(\boldsymbol{x}, \boldsymbol{u}_i)$, $0 \leq i \leq j \leq n$, $J(\boldsymbol{x}, \boldsymbol{u}_j) \leq J(\boldsymbol{x}, \boldsymbol{u}_i) \leq J(\boldsymbol{x}, \boldsymbol{u}_1)$. Since $J(\boldsymbol{x}(t), \boldsymbol{u})$ is bounded, then $r(\boldsymbol{x}, \boldsymbol{u})$ and $B_\gamma(\boldsymbol{x})$ are bounded.

Lemma 2 indicates that the $B_\gamma(\boldsymbol{x})$ remains within limits after every policy

enhancement step when using the optimal policy $\boldsymbol{u}_j = -0.5R^{-1}g^T\nabla V_j$, with the initial condition $\boldsymbol{x}_0 \in \mathrm{Int}(\mathcal{S})$, and admissible feedback control policy exists. As aforementioned, the RCBF value would approach infinity at the safe set boundary. Consequently, this ensures that the states of the system state will not reach the boundary.

2) Stability analysis

The proposed safe controller should also ensure stability within the safe region defined in **Definition 4**.

Definition 4. Safe region

The safe region for the optimal control problem is defined to be

$$D = \{\boldsymbol{x} | \boldsymbol{x} \in \mathrm{Int}(\mathcal{S}) - \beta(\boldsymbol{x}_h^*, r_0)\} \tag{6.27}$$

where $\boldsymbol{x}_h^* = \{\boldsymbol{x} | h(\boldsymbol{x}) = 0\}$, and β is the sphere surrounding the boundary with a radius of r_0 and $\boldsymbol{x}^*$ denotes the equilibrium point of the control system.

The coefficient γ in RCBF is chosen that $B_\gamma(\boldsymbol{x})/\big(B_\gamma(\boldsymbol{x}) + Q(\boldsymbol{x})\big) \leq 0.5,\ \boldsymbol{x} \in D$. Thus, $Q(\boldsymbol{x})$ is the primary component in the optimal control problem within the safe region.

Lemma 3.

Assume that $\boldsymbol{x} = 0$ is the equilibrium point, and safe region includes the origin. Besides, $W(t, \boldsymbol{x})$: $[0, \infty) \times \mathrm{D}$ is assumed as a continuously differentiable function, it is said to be a valid control Lyapunov function if there exists a control input $\boldsymbol{u}$, such that for any $\boldsymbol{x} \in D$, there is

$$N_1 \leq W(t, \boldsymbol{x}) \leq N_2 \tag{6.28a}$$

$$\frac{\partial W}{\partial t} + \frac{\partial W}{\partial \boldsymbol{x}}(f + g\boldsymbol{u}) \leq 0, \ \boldsymbol{x} \in D \tag{6.28b}$$

where N_1 and N_2 are continuous positive-definite functions in safe region D. The system is uniformly stable at origin point.

From **Lemma 1** and **Lemma 2**, it can be obtained that $V(\boldsymbol{x}, \boldsymbol{u}_j) \leq V(\boldsymbol{x}, \boldsymbol{u}_1)$, $1 \leq j$, and $V(\boldsymbol{x}, \boldsymbol{u}_1)$ is bounded. Thus, N can be defined as $N = \max_{\mathrm{t}} V(\boldsymbol{x}, \boldsymbol{u}_1)$. Besides, proof for **Lemma 2** indicates that $V(\boldsymbol{x}, \boldsymbol{u}_j)$ is decreasing. Thus, using Lemma 3, the safe optimal control system can be regarded as uniformly stable.

6.4 Numerical results and discussions

6.4.1 Effectiveness of the SDRL controller

The objective of the SDRL controller is to maintain a stable 9 mm air gap between electromagnets and the guideway. The environment of the designed controller is the simplified maglev system as in Section 6.3.1. The initial air gap is 16 mm and the value of the system model parameters are given in **Table 6-1**.

Table 6-1 Parameter values of the maglev train–guideway coupling system

Physical quantity	Value	Physical quantity	Value
Mass m / kg	700	Vacuum permeability μ_0 / (Hm^{-1})	$4\pi \cdot 10^{-7}$
Number of Turns of coil N	700	Area of coil A/m^2	0.024
Coil resistance R/Ω	1.2	Stable air gap x_{1eq} /m	0.009
Mass of track M / kg	8937.755	First mode natural frequency of track w	188.49564
First mode damping ratio of track ξ_1	0.005	-	-

For the training of the control algorithm, the SDRL algorithm is implemented by modifying the code of the TD3 algorithm. The structures of both the critic and actor networks in SDRL consist of three hidden layers. Both networks employ a rectified linear unit activation function (ReLU) , while the output layer of the actor network utilizes a tanh function. The learning rates are set to 1×10^{-3} for the critic networks and 1×10^{-4} for the actor networks. The training iteration is set to 20,000 with each

iteration comprising 500 the time step ($\Delta t = 0.001 s$). In addition, the discounted factor is set to 0.99 and the update parameter τ is 0.01. Exploration is facilitated by adding noise to the actions. The entire algorithm is implemented in Python 3.8 with PyTorch 1.5.1 with a mini-batch of 64 transitions sampled from a replay buffer $\mathcal{B}$ of a size of 1×10^5. The reward function is designed as follows

$$r_{sd}(\boldsymbol{x}) = -\zeta\left(d - d_{eq}\right)^2 - \dot{d}^2 - \log\left(\frac{\gamma_1(d - d_{min})}{1 + \gamma_1(d - d_{min})}\right) - \log\left(\frac{\gamma_2(d_{max} - d)}{1 + \gamma_2(d_{max} - d)}\right) \tag{6.29}$$

where ζ is a weight coefficient which is set to 10000, $d_{max} = 0.016$ and $d_{min} = 0$, γ_1 and γ_2 are designed coefficients which are set to 2.

After training the SDRL algorithm, the optimal neural network parameters obtained during the training process are set as the parameters of the SDRL controller. This ensures that the learned knowledge and policies are effectively utilized in the control process. To evaluate the effectiveness of the trained SDRL controller, a PID controller is employed for comparison. In **Figure 6–11**, the control curves of the PID, the GA–ST–SMC, and the proposed SDRL controllers are depicted. Notably, the SDRL and GA–ST–SMC exhibit smoother convergence towards the reference air gap compared to the PID controller. The SDRL and GA–ST–SMC achieves convergence in approximately 0.15 s, significantly faster than the PID controller, which takes about 0.8 s. Moreover, both the SDRL and PID controllers exhibit an overshoot value of 0 mm, while the GA–ST–SMC controller shows a maximum overshoot of around 0.8 mm.

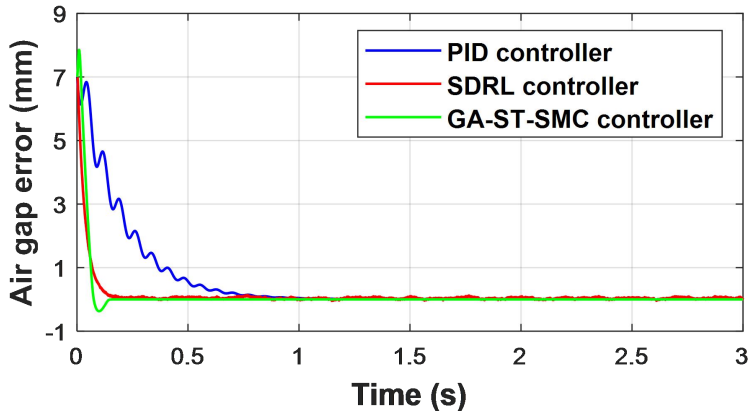


Figure 6–11 Control curves of the PID, the GA–ST–SMC, and the proposed SDRL controllers

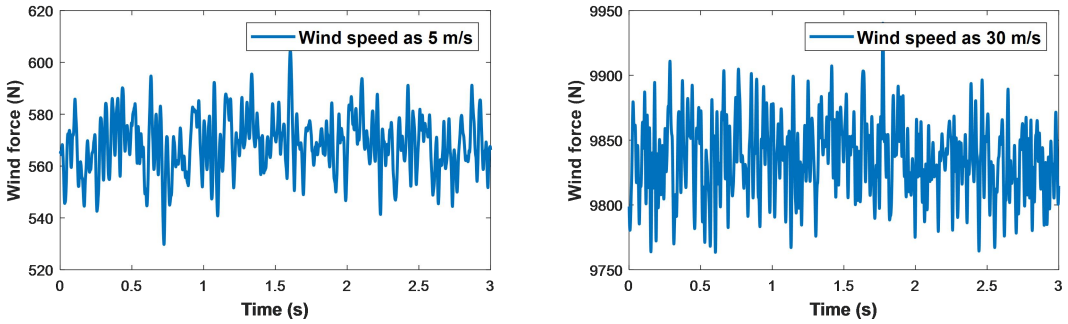
6.4.2 Impact of crosswinds on the control performance of SDRL controller

In this section, we investigate the impact of crosswinds on the control performance of the SDRL controller under varying train speeds and crosswind speeds. The wind forces for different scenarios are obtained using the validated numerical model.

6.4.2.1 Impact of crosswind speeds

To investigate the impact of varying crosswind speeds on the maglev train controlled by the SDRL controller, crosswind forces obtained in Section 6.2 are utilized. Especially, sets of crosswind forces contains six levels of wind speeds as 5 m/s, 10 m/s, 15 m/s, 20 m/s, 25 m/s, and 30 m/s, and two kinds of train speeds as 430 km/h and 600 km/h. As the maglev train–guideway coupling system is simplified to a single-point model featuring electromagnet levitation linked with a simply supported

beam of a specific span, the analysis focuses solely on the lift force and overturning moment for head car (HC). Examples for calculated vertical crosswind forces combing lift force and overturning moment are as in **Figure 6–12**. It can be observed that crosswind forces varies a lot with increase of wind speed.



(a) (b)

Figure 6–12 Crosswind forces with wind speed as 5 m/s and 30 m/s, and train speed as 430 km/h

Figure 6–13, **Figure 6–14** and **Figure 6–15** illustrate the control curves depicting air gap errors for the SDRL, PID, and GA–ST–SMC controllers under the influence of these crosswind forces. As shown in **Figure 6–13**, with increasing wind speeds, there is a marginal increase in the air gap error controlled using SDRL controller. Despite this slight increment, the entire maglev train–guideway coupling system demonstrates a smooth transition towards the equilibrium point. In contrast, the systems governed by the PID and GA–ST–SMC controllers, as shown in **Figure 6–14** and **Figure 6–15**, exhibits noticeable overshoots of approximately 3 mm and 1.4

mm in the control curves, which are significant. Additionally, the fluctuations in the system become more pronounced, potentially leading to passenger discomfort due to the less stable control response provided by the PID controller compared to the SDRL and GA–ST–SMC controllers.

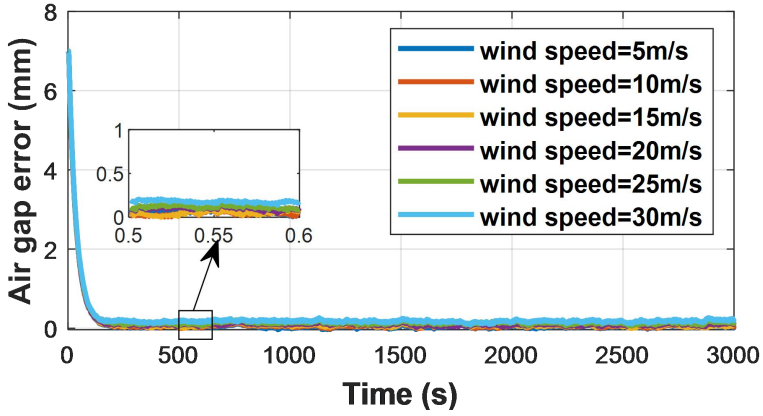


Figure 6–13 Air gap error under different crosswind speed using SDRL controller with train speed as 430 km/h

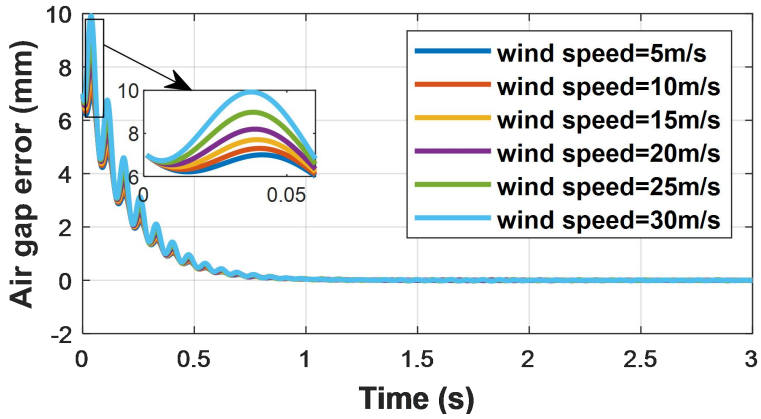


Figure 6–14 Air gap error under different crosswind speed using PID controller with

train speed as 430 km/h

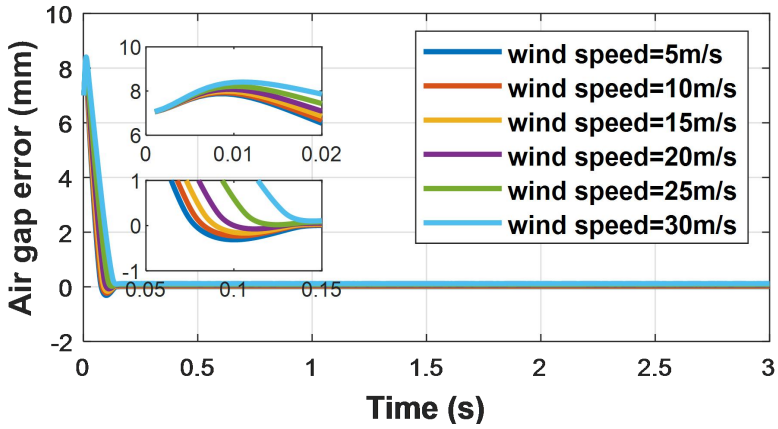
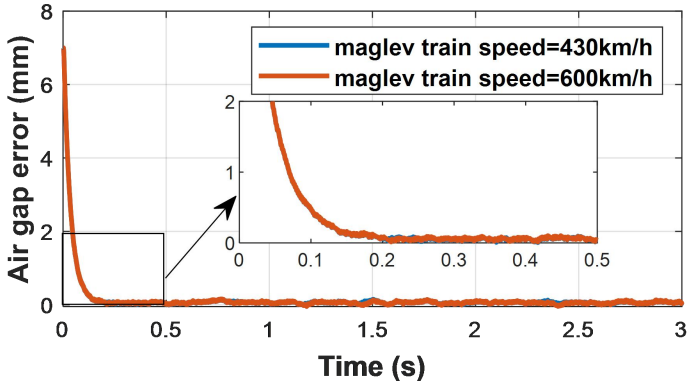


Figure 6–15 Air gap error under different crosswind speed using GA–ST–SMC controller with train speed as 430 km/h

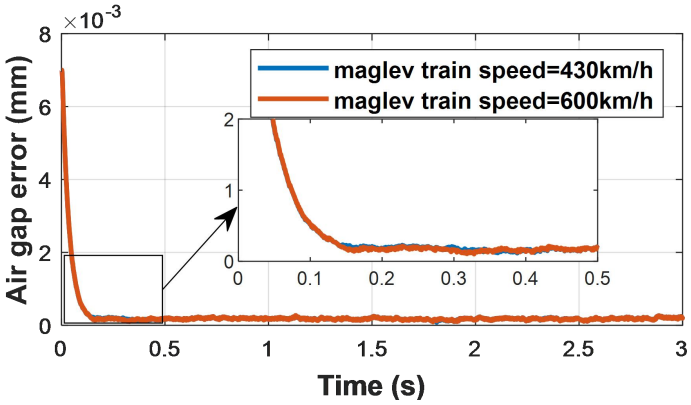
6.4.2.2 Impact of maglev train speeds

In a detailed exploration of the impact of crosswinds on maglev trains operating at varying speeds under SDRL, PID and GA–ST–SMC controllers, two specific levels of wind speeds (10m/s and 30m/s) and two levels of maglev train speeds (430 km/h and 600 km/h) are specifically chosen for analysis. The control curves representing air gap errors for the SDRL, GA–ST–SMC and PID controllers in response to these varying maglev train speeds are visually illustrated in **Figure 6–16**, **Figure 6–17**, and **Figure 6–18**. Notably, observations from these figures indicate that the control performance of all controllers remains relatively stable and are not significantly affected by the variations in maglev train speeds. This consistency in control

effectiveness across different operating speeds underscores the robustness and reliability of the SDRL controller in managing the maglev train system under diverse conditions, highlighting its potential for ensuring stable and efficient operations across a range of operational parameters.

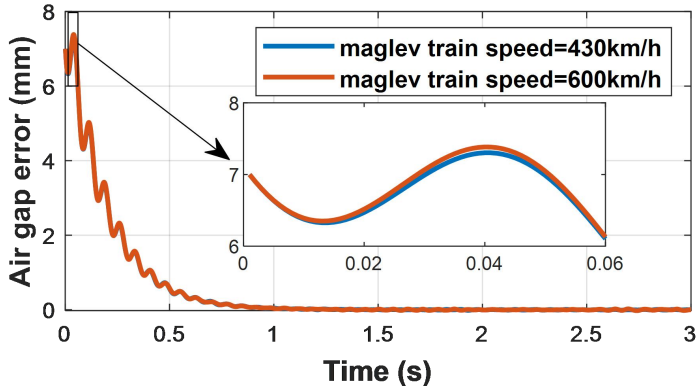


(a)

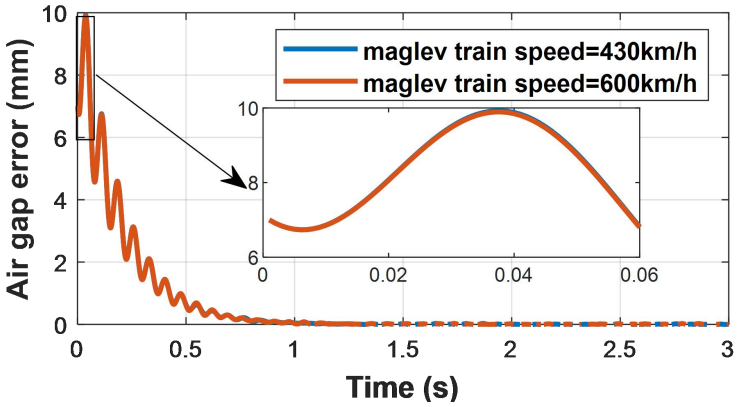


(b)

Figure 6–16 Air gap error under different train speeds using SDRL controller: (a) wind speed as 10 m/s, (a) wind speed as 30 m/s

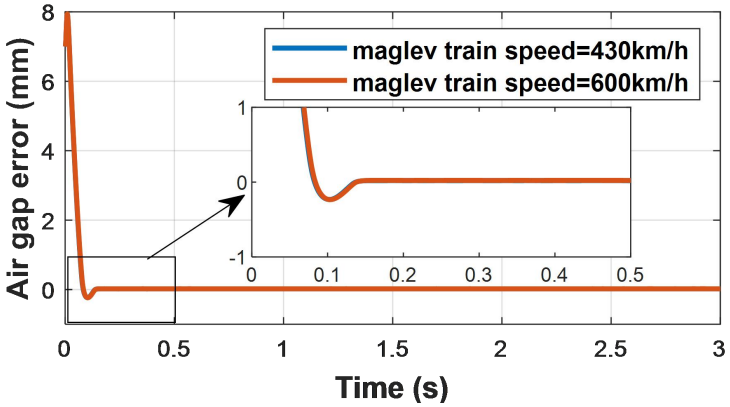


(a)

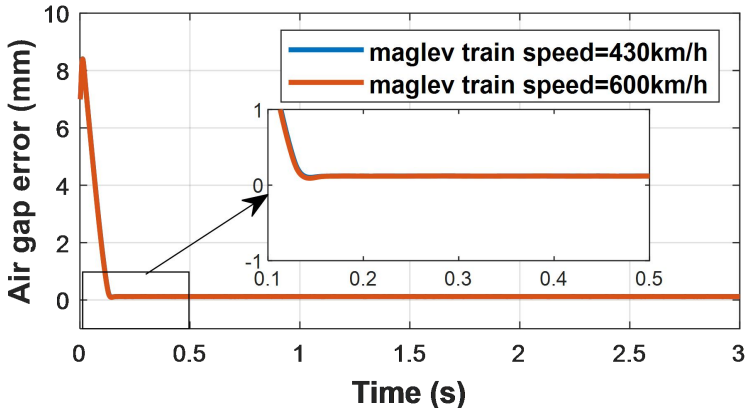


(b)

Figure 6–17 Air gap error under different train speeds using PID controller: (a) wind speed as 10 m/s, (a) wind speed as 30 m/s



(a)



(b)

Figure 6–18 Air gap error under different train speeds using GA–ST–SMC controller: (a) wind speed as 10 m/s, (a) wind speed as 30 m/s

6.4.3 Maximum system responses

It is evident that the crosswind has a significant impact on the vibration of maglev train-guideway coupling system, particularly for system controlled by conventional controller. To further analyze the effects of train speed and wind speed on system control performance in the presence of crosswinds, the peak value of system vibration response have been calculated and are presented in the **Figure 6–19** and **Figure 6–20**.

Figure 6–19 shows the maximum accelerations of the bogie controlled by both the conventional PID controller, a novel GA–ST–SMC controller, and a newly proposed SDRL controller. It can be observed that the acceleration increases with both train speed and wind speed when using the PID and GA–ST–SMC controllers. However, the maximum acceleration values for the SDRL controller remain relatively consistent across different train and wind speeds. Thus, it can be concluded that the

SDRL controller demonstrates significantly greater robustness than the PID and GA–ST–SMC controllers, making it suitable for practical applications that muse mitigate the effect of crosswinds.

The maximum overshoot values of the air gap controlled by the PID, GA–ST–SMC and SDRL controllers are depicted in **Figure 6–20**. Notably, the overshoot value for the SDRL controller remains at zero across various scenarios. in contrast, the overshoot values for the PID controller and GA–ST–SMC controller increase with the increase of train speed as well as wind speed. The maximum overshoot value approaches nearly 3 mm for the PID controller and 1.4 mm for the GA–ST–SMC controller when the wind speed reaches 30 m/s, which may lead to potential system's control failures.

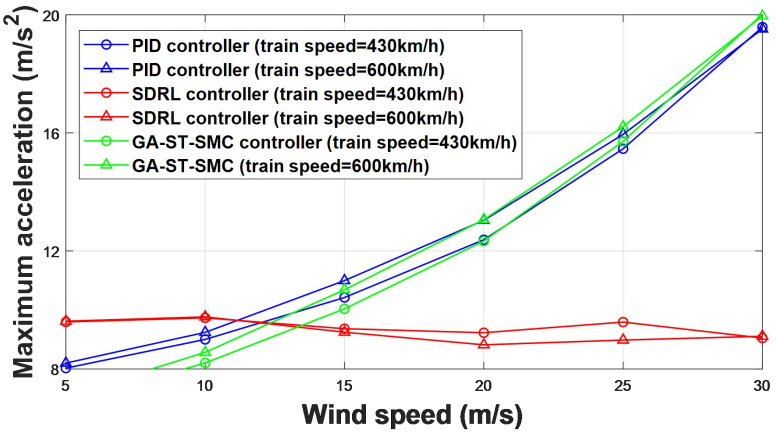


Figure 6–19 Maximum accelerations of the bogie controlled by PID, GA–ST–SMC and SDRL controllers

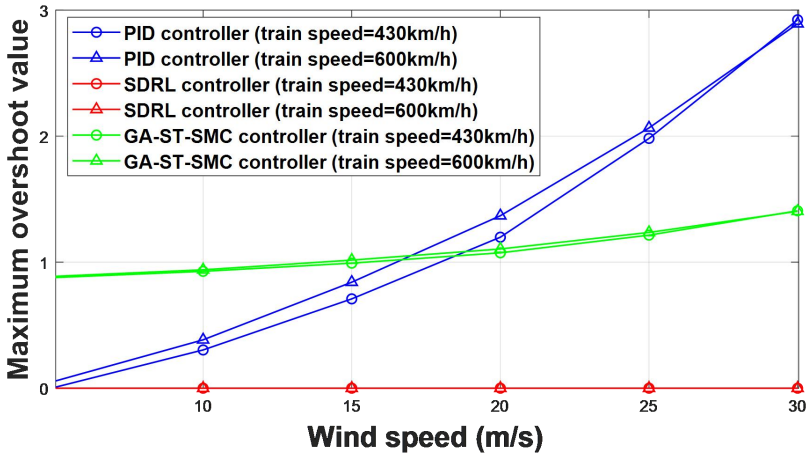


Figure 6–20 Maximum overshoot value of the air gap controlled by PID, GA–ST–SMC and SDRL controllers

6.5 Conclusion

In this chapter, a numerical model for the maglev train-guideway system is established, and a novel SDRL controller is introduced to regulate the control system. The accuracy of the model is validated through a wind tunnel test conducted at the Hong Kong Polytechnic University. Utilizing this validated numerical model, crosswind forces under varying wind speeds and train velocities are calculated. An analysis of the impact of crosswinds on the maglev train-guideway system under the SDRL controller is conducted, comparing it against the conventional PID controller and a novel GA–ST–SMC controller. The key findings regarding control performance and the influence of crosswinds are outlined as follows:

1) The SDRL controller and the GA–ST–SMC controller exhibits superior control efficiency compared to the PID controller, achieving convergence in approximately 0.15 seconds, in contrast to the conventional PID controller's convergence time of 0.8 second. However, both the SDRL and PID controllers exhibit an overshoot value of 0 mm, while the GA–ST–SMC controller shows a maximum overshoot of around 0.8 mm.

2) The SDRL controller demonstrates enhanced robustness when encountering crosswinds and increased train speed. As wind speed rises, the air gap controlled by the PID controller experiences significant fluctuations and overshoot, the overshoot value of the GA–ST–SMC also increases, whereas the SDRL controller maintains smooth control with zero overshoot values.

3) The vibration response of the maglev train-guideway system primarily depends on train speed and crosswind velocity. Beyond a wind speed threshold of 30 m/s, the PID

and GA–ST–SMC control systems may fail. Conversely, the proposed SDRL controller effectively manages the system within safe parameters, even when subjected to wind speeds of 30 m/s and train speeds of 600 km/h.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Conclusions

This thesis has significantly advanced the modeling, control, and performance optimization of electromagnetic suspension (EMS) maglev levitation systems. The research has addressed key challenges in robust and efficient control under real-world uncertainties, and the main contributions and insights are summarized below:

Task 1: TL–DRL adaptive levitation controller

A novel TL–DRL adaptive levitation controller was developed to address the nonlinear control challenges inherent in maglev systems. Comparative studies with conventional PID and ASMC demonstrated that the TL–DRL controller offers superior performance in terms of stability and adaptability. Notably, the controller maintained robust operation under a variety of uncertainties, including fluctuating train loads, passenger-induced mass changes, track irregularities, and environmental disturbances such as wind. This robustness is critical for ensuring safe and reliable maglev operation in dynamic, real-world environments.

Task 2: HJB－MADRL controller

For the first time, the HJB method was integrated with the MADRL framework to enhance controller performance. The HJB equation was leveraged to optimize the value function and improve the training efficiency of the critic network. The resulting controller effectively managed the MIMO maglev system without the need for

decoupling, and demonstrated strong robustness to parameter uncertainties and external disturbances. Experimental validation on a full-scale maglev bogie levitation system confirmed the controller's superior performance and rapid convergence compared to both PID and model-guided controllers.

Task 3: RCBF–SDRL controller

This work also pioneered the integration of RCBF into the SDRL framework to address optimal control under safety constraints. The RCBF–SDRL controller successfully transformed the constrained control problem into an unconstrained one, enabling safe and optimal operation of the maglev system even when coupled with a flexible guideway and subject to model uncertainties. Comparative analysis showed that this approach outperformed both PID and GA–ST–SMC controllers, particularly in the presence of external disturbances.

Task 4: Crosswind effect on the maglev train‑guideway coupling system

A comprehensive numerical model was developed to quantify crosswind loads on the maglev train‑guideway system, and its accuracy was validated through wind tunnel experiments. Building on this, an SDRL-based control strategy was proposed to manage the system under crosswind conditions. The controller demonstrated enhanced robustness and stability compared to conventional PID and GA‑ST‑SMC controllers, effectively mitigating the adverse effects of crosswinds and ensuring reliable levitation performance.

7.2 Recommendations and future works

While the research in this thesis has successfully addressed the stated objectives, several opportunities for further development and optimization exist. Future work in the area of magnetic levitation control could focus on the following:

DRL assisted Fault-tolerant control

During the operation of the maglev train, controllers of the train might fail to levitate the train. DRL assisted fault-tolerant control offers a promising approach for enhancing the safety and reliability of maglev train systems in the presence of faults or unexpected failures. By leveraging DRL, the control system can learn to detect, diagnose, and compensate for various types of faults—such as actuator malfunctions, sensor failures, or partial loss of levitation—through continuous interaction with the environment. The DRL agent can adapt its control policy in real time, enabling the system to maintain stable operation and minimize performance degradation even under fault conditions. Furthermore, DRL's ability to process high-dimensional data and extract complex patterns allows for more accurate fault identification and more effective recovery strategies compared to traditional fault-tolerant control methods. This approach not only improves the resilience of maglev train systems but also reduces the need for manual intervention, paving the way for safer and more autonomous railway operations.

Enhanced Passenger Comfort using DRL

Passenger comfort can be significantly improved by incorporating

comfort-related metrics into the reward function design of the DRL controller. By accurately reflecting factors such as vibrations, jerks, and noise within the reward structure, the DRL agent can learn control policies that actively minimize these undesirable effects during train operation. This approach enables the maglev train system to not only maintain stability and safety but also provide a smoother and quieter ride, thereby enhancing the overall travel experience for passengers. Integrating passenger comfort considerations into the control framework demonstrates the potential of DRL to address both technical performance and user satisfaction in advanced transportation systems.

Adaptive Speed Control in Varying Environments

To ensure safe and efficient operation under diverse weather and track conditions, it is essential for the maglev train to adapt its speed in response to environmental changes. Hybrid DRL models, which integrate model-based predictions with learning-based adjustments, offer a promising solution for this challenge. By leveraging predictive models to anticipate the effects of environmental factors and combining them with the adaptability of DRL, the control system can dynamically adjust train speed to optimize performance and safety. This approach enables the maglev train to respond effectively to real-time variations in weather, track conditions, and other external influences, thereby enhancing operational reliability and passenger safety.

Digital Twin assisted Adaptive DRL Control for Time Varying Dynamics

To address the time-varying dynamics of the maglev train—such as variations in load, guideway flexibility, and external disturbances—digital twin assisted adaptive DRL control presents a promising solution. A digital twin provides a real-time, high-fidelity virtual replica of the physical maglev system, enabling continuous monitoring, data collection, and simulation of system behavior under varying conditions such as changes in load, guideway flexibility, and external disturbances. By incorporating the digital twin into the adaptive DRL control loop, the controller can leverage accurate, up-to-date system information to enhance learning efficiency and policy adaptation. This synergy allows the adaptive DRL controller to anticipate and respond more effectively to dynamic changes, improving robustness and stability. Furthermore, the Digital Twin facilitates safe testing and validation of control strategies in a virtual environment before deployment, reducing risks and accelerating the development of reliable, data-driven control solutions for complex, nonlinear, and time-varying maglev train systems.

Crosswinds measurement assisted controller design

The measured crosswind information can be effectively utilized in a maglev train DRL controller by incorporating it as part of the state input to the neural network. Specifically, the real-time crosswind data obtained from onboard sensors can be included in the observation vector that the DRL agent receives at each time step. By doing so, the DRL controller is able to perceive the current crosswind conditions and learn control policies that explicitly account for these external disturbances.

During the training phase, the DRL agent interacts with the environment where crosswind forces are either simulated or provided as measured data. The agent learns to take actions (such as adjusting levitation and guidance forces) that maximize a reward function, which can be designed to penalize deviations from the desired trajectory or instability caused by crosswinds. By continuously receiving crosswind information as part of its observations, the DRL controller can develop robust strategies to compensate for varying and unpredictable wind conditions.

REFERENCES

Adil, H. M. M., Ahmed, S., & Ahmad, I. (2020). Control of MagLev System Using Supertwisting and Integral Backstepping Sliding Mode Algorithm. *IEEE Access*, *8*, 51352–51362. https://doi.org/10.1109/ACCESS.2020.2980687.

Ahmad, I., Shahzad, M., & Palensky, P. (2014, May). Optimal PID control of magnetic levitation system using genetic algorithm. In *2014 IEEE International Energy Conference (ENERGYCON)* (pp. 1429-1433). IEEE.

Altman, E. (1999). *Constrained markov decision processes*. Boca Raton: Chapman & Hall.

Ames, A. D., Xiangru Xu, Grizzle, J. W., & Tabuada, P. (2017). Control Barrier Function Based Quadratic Programs for Safety Critical Systems. *IEEE Transactions on Automatic Control*, 62(8), 3861 – 3876. https://doi.org/10.1109/TAC.2016.2638961

Arefin, M. R., & Asadujjaman, M. (2016). Minimizing Average of Loss Functions Using Gradient Descent and Stochastic Gradient Descent. *The Dhaka University Journal of Science*, *64*(2), 141-145.

Baxter, J., & Bartlett, P. L. (2001). Infinite-horizon policy-gradient estimation. *Journal of Artificial Intelligence Research*, 15, 319 – 350.

Bellemare, M. G., Naddaf, Y., Veness, J., & Bowling, M. (2013). The arcade learning environment: An evaluation platform for general agents. *Journal of Artificial Intelligence Research*, *47*, 253-279.

Bidikli, B. (2020). An observer-based adaptive control design for the maglev system. *Transactions of the Institute of Measurement and Control*, *42*(14), 2771-2786.

Bu, X., Wang, L., Han, Y., Liu, H., Hu, P., and Cai, C. (2024). Dynamic model of high-speed maglev train-guideway bridge system with a nonlinear suspension controller. *Advances in structural engineering*, vol. 27, no. 8, pp. 1328 – 1348, 2024, doi: 10.1177/13694332241247921.

B. S En, *Railway Applications-Aerodynamics Part 6: Requirements and Test Procedures for Crosswind Assessment*. CEN European Standard, 2018.

Che, Z., Chen, Z., Ni, Y., Huang, S., & Li, Z. (2023). Research on the impact of air-blowing on aerodynamic drag reduction and wake characteristics of a high-speed maglev train. *AIP Publishing LLC*.

Chen, C., Xu, J., Ji, W., Rong, L., & Lin, G. (2019). Sliding mode robust adaptive control of maglev vehicle's nonlinear suspension system based on flexible track: Design and experiment. *IEEE Access*, *7*, 41874-41884.

Chen, C., Xu, J., Lin, G., Sun, Y., & Gao, D. (2020). Fuzzy adaptive control particle swarm optimization based on TS fuzzy model of maglev vehicle suspension system. *Journal of Mechanical Science and Technology*, *34*, 43-54.

Chen, C., Xu, J., Lin, G., Sun, Y., & Ni, F. (2022). Model identification and nonlinear adaptive control of suspension system of high-speed maglev train. *Vehicle System Dynamics*, *60*(3), 884-905.

Chen, Q., Tan, Y., Li, J., Oetomo, D., & Mareels, I. (2018). Model-guided data-driven decentralized control for magnetic levitation systems. *IEEE Access*, *6*, 43546-43562.

Chen, Z., Guo, Z., Ni, Y., Liu, T., & Zhang, J. (2023). A suction method to mitigate pressure waves induced by high-speed maglev trains passing through tunnels. *Sustainable Cities and Society*, 96: 104682.

Chen, Z. W., Rui, E. Z., Liu, T. H., Ni, Y. Q., Huo, X. S., Xia, Y. T., Li, W. H., Guo, Z. J., & Zhou, L. (2022). Unsteady Aerodynamic Characteristics of a High-Speed Train Induced by the Sudden Change of Windbreak Wall Structure: A Case Study of the Xinjiang Railway. *Applied Sciences*, 12(14), 7217-.

Chen, Z., Zeng, G., Hashmi, S., Liu, T., Zhou, L., Zhang, J., & Hemida, H. (2023). Impact of the windbreak transition on flow structures of the high-speed railway and mitigation using oblique structure and circular curve structure transition. *International Journal of Numerical Methods for Heat & Fluid Flow*, 33(4), 1354-1378.

Chi, Z., & Li, J. (2017, July). Simulation analysis of the vehicle-guideway coupling vibration of EMS maglev train. In *2017 36th Chinese Control Conference (CCC)* (pp. 10376-10380). IEEE.

Cirillo, M. D., Mirdell, R., Sjöberg, F., & Pham, T. D. (2019). Tensor decomposition for colour image segmentation of burn wounds. *Scientific reports*, *9*(1), 3291.

Dalwadi, N., Deb, D., & Muyeen, S. M. (2021). A reference model assisted adaptive

control structure for maglev transportation system. *Electronics*, *10*(3), 332.

DBJ50T-347-2020 General Technical Standard for Urban Rail Express Vehicles, 2020.

Deng, Z., Zhang, X., & Zhao, Y. (2020). Transfer learning based method for frequency response model updating with insufficient data. *Sensors*, *20*(19), 5615.

de Jesús Rubio, J. (2018). Robust feedback linearization for nonlinear processes control. *ISA transactions*, *74*, 155-164.

Duka, A. V., Dulău, M., & Oltean, S. E. (2016). IMC based PID control of a magnetic levitation system. *Procedia Technology*, *22*, 592-599.

Fatemimoghadam, A., Toshani, H., & Manthouri, M. (2020). Control of magnetic levitation system using recurrent neural network-based adaptive optimal backstepping strategy. *Transactions of the Institute of Measurement and Control*, *42*(13), 2382-2395.

Foerster, J., Assael, I. A., De Freitas, N., & Whiteson, S. (2016). Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, *29*..

Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018, April). Counterfactual multi-agent policy gradients. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 32, No. 1).

Fujimoto, S., Hoof, H., & Meger, D. (2018, July). Addressing function approximation error in actor-critic methods. In *International conference on machine learning* (pp. 1587-1596). PMLR.

Ghosh, A., Krishnan, T. R., Tejaswy, P., Mandal, A., Pradhan, J. K., & Ranasingh, S. (2014). Design and implementation of a 2-DOF PID compensation for magnetic levitation systems. *ISA transactions*, *53*(4), 1216-1222.

Gottzein, E., Brock, K. H., Schneider, E., & Pfefferl, J. (1977). Control aspects of a tracked magnetic levitation high speed test vehicle. *Automatica*, *13*(3), 205-223.

Gu, S., Lillicrap, T., Ghahramani, Z., Turner, R. E., & Levine, S. (2016). Q-prop: Sample-efficient policy gradient with an off-policy critic. *arXiv preprint arXiv:1611.02247*.

Guang, H., Yun, L., Zhiqiang, L., & Zhide, J. I. (2010, October). Research on fault tolerant control technology based on networked control system of maglev train. In *2010 International Conference on Intelligent System Design and Engineering Application* (Vol. 2, pp. 210-214). IEEE.

Gupta, J., Pathak, S., & Kumar, G. (2022, May). Deep learning (CNN) and transfer learning: A review. In *Journal of Physics: Conference Series* (Vol. 2273, No. 1, p. 012029). IOP Publishing.

Han, H. S., & Kim, D. S. (2016). Magnetic levitation. *Springer Tracts on Transportation and Traffic. Springer Netherlands*, *247*.

Han, H. S., Yim, B. H., Lee, N. J., Hur, Y. C., and Kim, S. S. (2019). Effects of the guideway's vibrational characteristics on the dynamics of a maglev vehicle. *Vehicle System Dynamics*, 47, 309 – 324.

Han, S., Zhang, J., Xiong, X., Ji, P., Zhang, L., Sheridan, J., & Gao, G. (2022). Influence of high-speed maglev train speed on tunnel aerodynamic effects. *Building and Environment,* 223, 109460-. https://doi.org/10.1016/j.buildenv.2022.109460.

Hao, J., Huang, D., Cai, Y., & Leung, H. F. (2017). The dynamics of reinforcement social learning in networked cooperative multiagent systems. *Engineering Applications of Artificial Intelligence*, *58*, 111-122.

Hasselt, H. Van, Guez, A., & Silver, D. (2016). Deep reinforcement learning with double Q-Learning. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence, (AAAI-16)* (pp. 2094–2100).

He, G., Li, J., & Cui, P. (2015). Decoupling control design for the module suspension control system in maglev train. *Mathematical Problems in Engineering*, *2015*.

He, G., Li, J., Cui, P., & Li, Y. (2015). TS fuzzy model based control strategy for the networked suspension control system of maglev train. *Mathematical Problems in Engineering*, *2015*.

He, G., Li, J., Li, Y., & Cui, P. (2013, June). Interactions analysis in the maglev bogie with decentralized controllers using an effective relative gain array measure. In *2013 10th IEEE International Conference on Control and Automation (ICCA)* (pp. 1070-1075). IEEE.

Hessel, M., Modayil, J., Van Hasselt, H., Schaul, T., Ostrovski, G., Dabney, W., &

Silver, D. (2018). Rainbow: Combining improvements in deep reinforcement learning. In *32nd AAAI Conference on Artificial Intelligence (AAAI)* (pp. 3215 – 3222).

Huang, C. M., Yen, J. Y., & Chen, M. S. (2000). Adaptive nonlinear control of repulsive maglev suspension systems. *Control Engineering Practice*, *8*(12), 1357-1367.

Huang, Z., Zhou, Z., Chang, N., Chen, Z., & Wang, S. (2024). *Aerodynamic features of high-speed maglev trains with different marshaling lengths running on a viaduct under crosswinds*. Tech Science Press.

Hypiusová, M., & Osuský, J. (2010, February). PID controller design for magnetic levitation model. In *International Conference February* (Vol. 10, p. 13).

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *32nd International Conference on Machine Learning (ICML)* (pp. 448 – 456).

Kaloust, J., Ham, C., Siehling, J., Jongekryg, E., & Han, Q. (2004). Nonlinear robust control design for levitation and propulsion of a maglev system. *IEE Proceedings-Control Theory and Applications*, *151*(4), 460-464.

Kelley, C. T. (2003). *Solving nonlinear equations with Newton's method*. Society for Industrial and Applied Mathematics.

Kim, K. J., Han, J. B., Han, H. S., & Yang, S. J. (2015). Coupled vibration analysis of maglev vehicle-guideway while standing still or moving at low speeds. *Vehicle System Dynamics*, *53*(4), 587-601.

Konda, V., & Tsitsiklis, J. (1999). Actor-critic algorithms. *Advances in neural information processing systems*, *12*.

Konda, V. R., & Tsitsiklis, J. N. (2003). On actor-critic algorithms. *SIAM Journal on Control and Optimization*, 42(4), 1143 – 1166. https://doi.org/10.1137/S0363012901385691

Kong, E., Song, J. S., Kang, B. B., & Na, S. (2011). Dynamic response and robust control of coupled maglev vehicle and guideway system. *Journal of Sound and Vibration*, *330*(25), 6237-6253.

Koutník, J., Cuccu, G., Schmidhuber, J., & Gomez, F. (2013, July). Evolving

large-scale neural networks for vision-based reinforcement learning. In *Proceedings of the 15th annual conference on Genetic and evolutionary computation* (pp. 1061-1068).

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Lee, H. W., Kim, K. C., & Lee, J. (2006). Review of maglev train technologies. *IEEE transactions on magnetics*, *42*(7), 1917-1925.

Lee, J., Hyun Kim, D., & Edgar, T. F. (2005). Static decouplers for control of multivariable processes. *AIChE journal*, *51*(10), 2712-2720.

Leng, P., Li, Y., Zhou, D., Li, J., & Zhou, S. (2019). Decoupling control of maglev train based on feedback linearization. *IEEE Access*, *7*, 130352-130362.

Lengyel, G., & Kocsis, A. (2014). Vehicle-guideway interaction in maglev systems using a continuously coupled, deformable model. *Journal of Engineering Mechanics*, *140*(1), 182-192.

Levine, S., Finn, C., Darrell, T., & Abbeel, P. (2016). End-to-end training of deep visuomotor policies. *Journal of Machine Learning Research*, *17*(39), 1-40.

Li, C., Sanchez, R. V., Zurita, G., Cerrada, M., Cabrera, D., & Vásquez, R. E. (2015). Multimodal deep support vector classification with homologous features and its application to gearbox fault diagnosis. *Neurocomputing*, *168*, 119-127.

Li, G., Ren, Z., Li, B., Fu, T., & Duan, P. (2022). Global fast terminal integral sliding mode control based on magnetic field measurement for magnetic levitation system. *Asian Journal of Control*, *24*(5), 2363–2377. https://doi.org/10.1002/asjc.2660.

Li, J., Kuang, K., Wang, B., Liu, F., Chen, L., Wu, F., & Xiao, J. (2021, August). Shapley counterfactual credits for multi-agent reinforcement learning. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining* (pp. 934-942).

Li, J., Zhou, D., Li, J., Zhang, G., & Yu, P. (2015). Modeling and simulation of CMS04 maglev train with active controller. *Journal of Central South University*, 22(4), 1366 – 1377. https://doi.org/10.1007/s11771-015-2654-z.

Li, J. H., Li, J., & Zhang, G. (2013). A practical robust nonlinear controller for maglev levitation system. *Journal of Central South University*, *20*(11), 2991-3001.

Li, Q., Leng, P., Yu, P., Zhou, D., Li, J., & Qu, M. (2023, August). Decoupling Control for Module Suspension System of Maglev Train Based on Feedback Linearization and Extended State Observer. In *Actuators* (Vol. 12, No. 9, p. 342). MDPI.

Li, Q., & Shen, G. (2020). Study on Control Method of Maglev Vehicle-Guideway Coupling System Based on Robust Control Theory. In *Advances in Dynamics of Vehicles on Roads and Tracks: Proceedings of the 26th Symposium of the International Association of Vehicle System Dynamics, IAVSD 2019, August 12-16, 2019, Gothenburg, Sweden* (pp. 882-889). Springer International Publishing.

Li, Y. (2017). Deep reinforcement learning: An overview. *arXiv preprint arXiv:1701.07274*.

Lillicrap, T. P., Hunt, J. J., Pritzel, A., Heess, N., Erez, T., Tassa, Y., Silver, D., & Wierstra, D. (2015). Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*.

Lin, F. J., Wai, R. J., & Chen, H. P. (1998). A PM synchronous servo motor drive with an on-line trained fuzzy neural network controller. *IEEE Transactions on Energy Conversion*, *13*(4), 319-325.

Liu, H., Zhang, X., & Chang, W. (2009, April). PID control to maglev train system. In *2009 International Conference on Industrial and Information Systems* (pp. 341-343). IEEE.

Liu, J. (2018). *Intelligent control design and Matlab simulation* (pp. 113-233). Singapore: Springer.

Liu, K. Z., & Yao, Y. (2016). *Robust control: theory and applications*. John Wiley & Sons.

Liu, Z., Stichel, S., & Berg, M. (2022). Overview of technology and development of maglev and hyperloop systems.

Long, Z., Xue, S., Zhang, Z., & Xie, Y. (2007, August). A new strategy of active fault-tolerant control for suspension system of maglev train. In *2007 IEEE International Conference on Automation and Logistics* (pp. 88-94). IEEE.

Lowe, R., Wu, Y. I., Tamar, A., Harb, J., Pieter Abbeel, O., & Mordatch, I. (2017). Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, *30*.

Luo, B., Liu, D., & Wu, H. N. (2017). Adaptive constrained optimal control design for data-based nonlinear discrete-time systems with critic-only structure. *IEEE Transactions on Neural Networks and Learning Systems*, *29*(6), 2099-2111.

MacLeod, C., & Goodall, R. M. (1996). Frequency-shaping LQ control of Maglev suspension systems for optimal performance with deterministic and stochastic inputs. *IEE Proceedings-Control Theory and Applications*, *143*(1), 25-30.

Mahadevan, S., & Connell, J. (1992). Automatic programming of behavior-based robots using reinforcement learning. *Artificial intelligence*, *55*(2-3), 311-365.

Malisani, P., Chaplais, F., & Petit, N. (2016). An interior penalty method for optimal control problems with state and input constraints of nonlinear systems. *Optimal Control Applications & Methods*, 37(1), 3 – 33. https://doi.org/10.1002/oca.2134

Min, D. J., Jung, M. R., Kim, M. Y., & Kwark, J. W. (2017). Dynamic interaction analysis of maglev-guideway system based on a 3D full vehicle model. *International Journal of Structural Stability and Dynamics*, *17*(01), 1750006.

Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv*: 1312.5602.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, L., King, H., Kumaran, D., Wierstra, D., Legg S., & Hassabis D. (2015). Human-level control through deep reinforcement learning. *Nature*, *518*(7540), 529-533.

Morari, M., Skogestad, S., & Rivera, D. F. (1984, June). Implications of internal model control for PID controllers. In *1984 American Control Conference* (pp. 661-666). IEEE.

Nguyen, N. T. (2018). *Model-reference adaptive control* (pp. 83-123). Springer International Publishing..

Ni, F., Mu, S., Kang, J., & Xu, J. (2021). Robust controller design for maglev

suspension systems based on improved suspension force model. *IEEE Transactions on Transportation Electrification*, *7*(3), 1765-1779.

Ni, Q., Cai, Y., Zhang, D., Fang, D., Li, J., Li, J., & Sadarangani, K. (2016). A Practical Control Strategy for the Maglev Self-Excited Resonance Suppression. *Mathematical Problems in Engineering*, 2016(2016), 1 – 9. https://doi.org/10.1155/2016/8071938

O'Donoghue, B., Munos, R., Kavukcuoglu, K., & Mnih, V. (2016). Combining policy gradient and q-learning. *arXiv preprint arXiv:1611.01626*.

Peters, J., & Schaal, S. (2008). Reinforcement learning of motor skills with policy gradients. *Neural networks*, *21*(4), 682-697.

Phuah, J., Lu, J., Yasser, M., & Yahagi, T. (2005, May). Neuro-sliding mode control for magnetic levitation systems. In *2005 IEEE International Symposium on Circuits and Systems (ISCAS)* (pp. 5130-5133). IEEE.

Polydoros, A. S., & Nalpantidis, L. (2017). Survey of model-based reinforcement learning: Applications on Robotics. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 86(2), 153 – 173.

Prajna, S., Jadbabaie, A., & Pappas, G. J. (2007). A Framework for Worst-Case and Stochastic Safety Verification Using Barrier Certificates. *IEEE Transactions on Automatic Control*, 52(8), 1415 – 1428. https://doi.org/10.1109/TAC.2007.902736.

Qi, Y., Shen, C., Wang, D., Shi, J., Jiang, X., & Zhu, Z. (2017). Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery. *Ieee Access*, *5*, 15066-15079.

Schaul, T., Quan, J., Antonoglou, I., & Silver, D. (2016). Prioritized experience replay. In *Proceedings of the 4th International Conference on Learning Representations (ICLR)* (pp. 1 – 21).

Schmidt, L. M., Brosig, J., Plinge, A., Eskofier, B. M., & Mutschler, C. (2022, October). An introduction to multi-agent reinforcement learning and review of its application to autonomous mobility. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)* (pp. 1342-1349). IEEE.

Schulman, J., Chen, X., & Abbeel, P. (2017). Equivalence between policy gradients and soft q-learning. *arXiv preprint arXiv:1704.06440*.

Schulman, J., Moritz, P., Levine, S., Jordan, M., & Abbeel, P. (2015). High-dimensional continuous control using generalized advantage estimation. *arXiv preprint arXiv:1506.02438*.

Shiakolas, P. S., Van Schenck, S. R., Piyabongkarn, D., & Frangeskou, I. (2004). Magnetic levitation hardware-in-the-loop and MATLAB-based experiments for reinforcement of neural network control concepts. *IEEE Transactions on Education*, *47*(1), 33-41.

Shinskey, F. Greg. (1996). *Process control systems : application, design, and tuning* (4th ed.). McGraw-Hill.

Silver, D., Lever, G., Heess, N., Degris, T., Wierstra, D., & Riedmiller, M. (2014, January). Deterministic policy gradient algorithms. In *International conference on machine learning* (pp. 387-395). Pmlr.

Sina Tayarani-Bathaie, S., Sadough Vanini, Z. N., & Khorasani, K. (2014). Dynamic neural network-based fault diagnosis of gas turbine engines. *Neurocomputing (Amsterdam)*, *125*, 153–165. https://doi.org/10.1016/j.neucom.2012.06.050.

Sinha, P. K., & Pechev, A. N. (1999). Model reference adaptive control of a maglev system with stable maximum descent criterion. *Automatica (Oxford)*, *35*(8), 1457–1465. https://doi.org/10.1016/S0005-1098(99)00040-0.

Srinivasan, M., Coogan, S., & Egerstedt, M. (2018). *Control of Multi-Agent Systems with Finite Time Control Barrier Certificates and Temporal Logic*. https://doi.org/10.48550/arxiv.1808.02393

Su, J., Vasconcellos, D. V., Prasad, S., Sgandurra, D., Feng, Y., & Sakurai, K. (2018). Lightweight Classification of IoT Malware Based on Image Recognition. *2018 IEEE 42ND ANNUAL COMPUTER SOFTWARE AND APPLICATIONS CONFERENCE (COMPSAC 2018), VOL 2*, *2*, 664–669. https://doi.org/10.1109/COMPSAC.2018.10315.

Su, X., Yang, X., Shi, P., & Wu, L. (2014). Fuzzy control of nonlinear electromagnetic suspension systems. *Mechatronics (Oxford)*, *24*(4), 328–335. https://doi.org/10.1016/j.mechatronics.2013.08.002.

Sukhbaatar, S., Szlam, A., & Fergus, R. (2016). Learning Multiagent Communication with Backpropagation. *arXiv.Org*. https://doi.org/10.48550/arxiv.1605.07736.

Sun, Y., He, Z., Xu, J., Sun, W., & Lin, G. (2023). Dynamic analysis and vibration control for a maglev vehicle-guideway coupling system with experimental verification. *Mechanical Systems and Signal Processing*, *188*, 109954-. https://doi.org/10.1016/j.ymssp.2022.109954.

Sun, Y., Li, W., & Qiang, H. (2016, December). The design and realization of magnetic suspension controller of low-speed maglev train. In *2016 IEEE/SICE International Symposium on System Integration (SII)* (pp. 1-6). IEEE.

Sun, Y., Qiang, H., Xu, J., & Lin, G. (2019). Internet of Things-based online condition monitor and improved adaptive fuzzy control for a medium-low-speed maglev train system. *IEEE Transactions on Industrial Informatics*, *16*(4), 2629-2639.

Sun, Y., Xu, J., Qiang, H., & Lin, G. (2019). Adaptive neural-fuzzy robust position control scheme for maglev train systems with experimental verification. *IEEE Transactions on Industrial Electronics*, *66*(11), 8589-8599.

Sun, Y., Xu, J., Qiang, H., Wang, W., & Lin, G. (2019). Hopf bifurcation analysis of maglev vehicle–guideway interaction vibration system and stability control based on fuzzy adaptive theory. *Computers in Industry*, *108*, 197-209.

Sun, Y., Xu, J., Wu, H., Lin, G., & Mumtaz, S. (2021). Deep learning based semi-supervised control for vertical security of maglev vehicle with guaranteed bounded airgap. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), 4431-4442.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine learning*, *3*, 9-44.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Taghirad, H. D., Abrishamchian, M., Ghabcheloo, R., & Toosi, K. N. (1998, May). Electromagnetic levitation system: An experimental approach. In *Proceedings of the 7th international Conference on Electrical Engineering, Power System Vol* (pp. 19-26).

Tan, T., Bao, F., Deng, Y., Jin, A., Dai, Q., & Wang, J. (2019). Cooperative deep reinforcement learning for large-scale traffic grid signal control. *IEEE transactions on cybernetics*, 50(6), 2687-2700.

Tee, K. P., Ge, S. S., & Tay, E. H. (2009). Barrier Lyapunov Functions for the control of output-constrained nonlinear systems. *Automatica (Oxford)*, 45(4), 918 – 927. https://doi.org/10.1016/j.automatica.2008.11.017.

Teklu, E. A., & Abdissa, C. M. (2023). Genetic Algorithm Tuned Super Twisting Sliding Mode Controller for Suspension of Maglev Train with Flexible Track. *IEEE Access*, *11*, 30955-30969.

Thornton, R. D. (1991). Why the U.S. Needs a Maglev System. In *Technology review (1998)* (pp. 31-42-).

Tian, X., Xiang, H., Chen, X., and Li, Y. (2023). Dynamic response analysis of high-speed maglev train-guideway system under crosswinds. *Journal of Central South University*, 30(8), pp. 2757–2771. doi: 10.1007/s11771-023-5403-8.

Thrun, S., & Schwartz, A. (2014, March). Issues in using function approximation for reinforcement learning. In *Proceedings of the 1993 connectionist models summer school* (pp. 255-263). Psychology Press..

Unni, A. C., Junghare, A. S., Mohan, V., & Ongsakul, W. (2016, September). PID, fuzzy and LQR controllers for magnetic levitation system. In *2016 International Conference on Cogeneration, Small Power Plants and District Energy (ICUE)* (pp. 1-5). IEEE.

Van Hasselt, H., Guez, A., & Silver, D. (2016, March). Deep reinforcement learning with double q-learning. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 30, No. 1).

Wade, H. L. (1997). Inverted decoupling: a neglected technique. *ISA transactions*, *36*(1), 3-10.

Wai, R. J., Chen, M. W., & Yao, J. X. (2016). Observer-based adaptive fuzzy-neural-network control for hybrid maglev transportation system. *Neurocomputing*, *175*, 10-24.

Wai, R. J., & Lee, J. D. (2005). Performance comparisons of model-free control strategies for hybrid magnetic levitation system. *IEE Proceedings-Electric Power Applications*, *152*(6), 1556-1564.

Wai, R. J., & Lee, J. D. (2008). Adaptive fuzzy-neural-network control for maglev transportation system. *IEEE Transactions on Neural Networks*, *19*(1), 54-70.

Wai, R. J., & Lee, J. D. (2008). Robust levitation control for linear maglev rail system using fuzzy neural network. *IEEE transactions on control systems technology*, *17*(1), 4-14.

Wai, R. J., Yao, J. X., & Lee, J. D. (2014). Backstepping fuzzy-neural-network control design for hybrid maglev transportation system. *IEEE Transactions on Neural networks and learning systems*, *26*(2), 302-317.

Wan, Z., Jiang, C., Fahad, M., Ni, Z., Guo, Y., & He, H. (2020). Robot-Assisted Pedestrian Regulation Based on Deep Reinforcement Learning. *IEEE Transactions on Cybernetics*, *50*(4), 1669–1682. https://doi.org/10.1109/TCYB.2018.2878977.

Wang, D., Li, X., Liang, L., & Qiu, X. (2020). Dynamic interaction analysis of bridges induced by a low-to-medium–speed maglev train. *Journal of Vibration and Control*, *26*(21–22), 2013–2025. https://doi.org/10.1177/1077546320910006.

Wang, D., Li, X., and Wang, C. (2023). Measurement and numerical analysis on dynamic performance of the LMS maglev train-track-continuous girder coupled system with running speed-up state. *Measurement: journal of the International Measurement Confederation*, vol. 217, pp. 113052-.

Wang, D. and Wang, C. (2024). Lateral vibration characteristics of low- to medium-speed maglev train–track–bridge coupled system in an accelerated state: Experimental and theoretical investigation. *Structures (Oxford)*, vol. 63, pp. 106373-, 2024, doi: 10.1016/j.istruc.2024.106373.

Wang, H., Shen, G., & Zhou, J. (2014). Control strategy of maglev vehicles based on particle swarm algorithm. *Journal of Modern Transportation*, *22*(1), 30–36. https://doi.org/10.1007/s40534-013-0031-x.

Wang, J., & Han, B. (2023). *Theory and technology for improving high-speed railway transportation capacity*. Elsevier.

Wang, K., Ma, W., Luo, S., Zou, R., & Liang, X. (2018). Coupling vibration analysis of full-vehicle vehicle-guideway for maglev train. *Australian Journal of Mechanical Engineering*, *16*(2), 109–117. https://doi.org/10.1080/14484846.2018.1486794.

Wang, L., Zhou, Y., & Shi, W. (2024). Seismic multi-objective stochastic parameters optimization of multiple tuned mass damper system for a large podium twin towers structure. *Soil Dynamics and Earthquake Engineering (1984)*, *177*, 108428-. https://doi.org/10.1016/j.soildyn.2023.108428.

Wang, S. M., Zhu, Q., Ni, Y. Q., Junqi Xu, J. Q., & Chen, F. (2023). Dynamic performance of low-and medium-speed maglev train running on the turnout. *Mechatron. Intell Transp. Syst*, *2*(1), 32-41.

Wang, X., Hu, X., Wang, J., Wang, L., Li, H., Deng, Z., & Zhang, W. (2023). Safety analysis of high temperature superconducting maglev train considering the aerodynamic loads under crosswinds. Proceedings of the Institution of Mechanical Engineers. *Part C, Journal of Mechanical Engineering Science*, 237(10), 2279 – 2290. https://doi.org/10.1177/09544062221140033.

Wang, Y. J., Yau, J. D., Shi, J., & Urushadze, S. (2020). Vibration reduction for interaction response of a maglev vehicle running on guideway girders. *Structural Engineering and Mechanics, An Int'l Journal*, *76*(2), 163-173.

Wang, Z., Schaul, T., Hessel, M., Van Hasselt, H., Lanctot, M., & De Freitas, N. (2016). Dueling network architectures for deep reinforcement learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)* (pp. 1995–2003).

Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, *8*, 279-292.

Williams, R. J. (1992). Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Machine learning*, *8*, 229-256.

Xia, W., Long, Z., & Dou, F. (2020). Disturbance rejection control using a novel velocity fusion estimation method for levitation control systems. *IEEE Access*, *8*, 173092-173102.

Xu, J., Chen, C., Gao, D., Luo, S., & Qian, Q. (2017). Nonlinear dynamic analysis on maglev train system with flexible guideway and double time-delay feedback control. *Journal of Vibroengineering*, *19*(8), 6346-6362.

Xu, J., Chen, C., Sun, Y., Ji, W., & Lin, G. (2019). Nonlinear dynamic analysis of Maglev Vehicle Based on flexible guideway and random irregularity. *International Journal of Applied Electromagnetics and Mechanics*, *60*(2), 263-280.

Xu, J., Chen, C., Sun, Y., Rong, L., & Lin, G. (2019). Nonlinear dynamic characteristic modeling and adaptive control of low speed maglev train. *International Journal of Applied Electromagnetics and Mechanics*, *62*(1), 73-92.

Xu, J., Chen, Y. H., & Guo, H. (2015). Robust levitation control for maglev systems

with guaranteed bounded airgap. *ISA transactions*, *59*, 205-214.

Xu, J., Du, Y., Chen, Y. H., & Guo, H. (2018). Adaptive robust constrained state control for non-linear maglev vehicle with guaranteed bounded airgap. *IET Control Theory & Applications*, *12*(11), 1573-1583.

Xu, J., Sun, Y., Gao, D., Ma, W., Luo, S., & Qian, Q. (2018). Dynamic modeling and adaptive sliding mode control for a maglev train system based on a magnetic flux observer. *Ieee Access*, *6*, 31571-31579.

Yang, J., Sun, R., Cui, J., & Ding, X. (2004, November). Application of composite fuzzy-PID algorithm to suspension system of Maglev train. In *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004* (Vol. 3, pp. 2502-2505). IEEE.

Yang, J., Zolotas, A., Chen, W. H., Michail, K., & Li, S. (2011). Robust control of nonlinear MAGLEV suspension system with mismatched uncertainties via DOBC approach. *ISA transactions*, *50*(3), 389-396.

Yang, J., Sun, R., Cui, J., & Ding, X. (2004, November). Application of composite fuzzy-PID algorithm to suspension system of Maglev train. In *30th Annual Conference of IEEE Industrial Electronics Society, 2004. IECON 2004* (Vol. 3, pp. 2502-2505). IEEE.

Yang, Y. B., & Lin, C. W. (2005). Vehicle–bridge interaction dynamics and potential applications. *Journal of sound and vibration*, *284*(1-2), 205-226.

Yang, Y. B., & Yau, J. D. (2015). Vertical and pitching resonance of train cars moving over a series of simple beams. *Journal of Sound and Vibration*, 337, 135 – 149. https://doi.org/10.1016/j.jsv.2014.10.024.

Yang, Y. B., Yau, J. D., Yao, Z., & Wu, Y. S. (2004). *Vehicle-bridge interaction dynamics: with applications to high-speed railways*. World Scientific.

Yang, Z. J., Miyazaki, K., Kanae, S., & Wada, K. (2004). Robust position control of a magnetic levitation system via dynamic surface control technique. *IEEE Transactions on Industrial Electronics*, *51*(1), 26-34.

Yaseen, H. M. S., Siffat, S. A., Ahmad, I., & Malik, A. S. (2022). Nonlinear adaptive control of magnetic levitation system using terminal sliding mode and integral backstepping sliding mode controllers. *ISA transactions*, *126*, 121-133.

Yau, J. D. (2009). Vibration control of maglev vehicles traveling over a flexible guideway. *Journal of sound and vibration*, *321*(1-2), 184-200.

Yau, J. D. (2010). Aerodynamic vibrations of a maglev vehicle running on flexible guideways under oncoming wind actions. *Journal of Sound and Vibration*, *329*(10), 1743-1759.

Yau, J. D. (2010). Interaction response of maglev masses moving on a suspended beam shaken by horizontal ground motion. *Journal of Sound and Vibration*, *329*(2), 171-188.

Yu, C., Velu, A., Vinitsky, E., Gao, J., Wang, Y., Bayen, A., & Wu, Y. (2022). The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems*, *35*, 24611-24624.

Yu, P., Song, M., Chen, Q., Wang, L., He, G., and Leng, P. (2021). Dynamic Response Analysis of Medium-Speed Maglev Train with Track Random Irregularities. *Journal of advanced transportation*, vol. 2021, pp. 1 – 16, 2021, doi: 10.1155/2021/1668496.

Zhou, D., Li, J., & Zhang, K. (2014). Amplitude control of the track-induced self-excited vibration for a maglev system. *ISA transactions*, *53*(5), 1463-1469.

Zhou, D., Yu, P., Wang, L., & Li, J. (2017). An adaptive vibration control method to suppress the vibration of the maglev train caused by track irregularities. *Journal of Sound and Vibration*, *408*, 331-350.

Zhang, D., Guo, Z., Ni, Y., Chen, Z., Ao, W., Bordbar, A., & Zhou, F. (2023). Correlation between cargo properties and train overturning safety for a high-speed freight train under strong winds. *Engineering Applications of Computational Fluid Mechanics*, 17(1), 2221308.

Zhang, J., Guo, Z., Han, S., Krajnović, S., Sheridan, J., & Gao, G. (2022). An IDDES study of the near-wake flow topology of a simplified heavy vehicle. *Transportation Safety and Environment Online*, 4(2). https://doi.org/10.1093/tse/tdac015.

Zhang, L. (2022). Vibration analysis and multi-state feedback control of maglev vehicle-guideway coupling system. *Electronic Research Archive*, *30*(10), 3887-3901.

Zhang, L., Zhang, Y., Zhang, C., & Zhao, H. (2019). Research on the improvement of feedback linearization control in suspension system countering inductance variation. *Mathematical Problems in Engineering*, *2019*.

Zhang, T., Zhou, D., Li, J., Wang, L., & Chen, Q. (2022). Research on Magnetic Suspension Control Scheme Based on Feedback Linearization under Low Track Stiffness. *Machines*, *10*(8), 692.

Zhang, Z., & Li, X. (2018). Real-time adaptive control of a magnetic levitation system with a large range of load disturbance. *Sensors*, *18*(5), 1512.

Zhao, F., You, K., Song, S., Zhang, W., & Tong, L. (2021). Suspension regulation of medium-low-speed maglev trains via deep reinforcement learning. *IEEE Transactions on Artificial Intelligence*, *2*(4), 341-351.

Zheng, L., Wang, Y., Yang, R., Wu, S., Guo, R., & Dong, E. (2023). An efficiently convergent deep reinforcement learning-based trajectory planning method for manipulators in dynamic environments. *Journal of Intelligent & Robotic Systems*, *107*(4), 50.

Zhou, D., Yu, P., Wang, L., & Li, J. (2017). An adaptive vibration control method to suppress the vibration of the maglev train caused by track irregularities. *Journal of Sound and Vibration*, 408, 331 – 350. https://doi.org/10.1016/j.jsv.2017.07.037

Zhou, X., Zhang, X., Zhao, H., Xiong, J., & Wei, J. (2022). Constrained Soft Actor-Critic for Energy-Aware Trajectory Design in UAV-Aided IoT Networks. *IEEE Wireless Communications Letters*, 11(7), 1 – 1.

Zhu, F., Xie, J., Lv, D., Xu, G., Yao, H., & Niu, J. (2024). Transient aerodynamic behavior of a high-speed Maglev train in plate braking under crosswind. *Physics of Fluids (1994)*, 36(3). https://doi.org/10.1063/5.0189686.

Zhu, Q., Wang, S. M., & Ni, Y. Q. (2024). Cooperative Control of Maglev Levitation System via Hamilton-Jacobi-Bellman Multi-Agent Deep Reinforcement Learning. *IEEE Transactions on Vehicular Technology*, 1 – 13.

Zhu, Y., Yang, Q., Li, J., & Wang, L. (2022). Research on Sliding Mode Control Method of Medium and Low Speed Maglev Train Based on Linear Extended State Observer. *Machines*, *10*(8), 644.