5-2024

Reinforcement Learning-Based Energy Management for Electric Vehicle Application

Yiming Ye
Clemson University, yimingy@g.clemson.edu

REINFORCEMENT LEARNING-BASED ENERGY MANAGEMENT FOR ELECTRIC VEHICLE APPLICATION

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Automotive Engineering

---

by
Yiming Ye
May 2024

Accepted by:
Dr. Jiangfeng Zhang, Committee Chair
Dr. Benjamin Lawler
Dr. Srikanth Pilla
Dr. Rajendra Singh
Dr. Zheyu Zhang

ABSTRACT

The increasing concerns about transportation pollution and fossil fuel depletion motivate many studies on vehicle electrification and advanced energy-saving propulsion systems. When comparing with traditional internal combustion engine vehicles, electrified vehicles, such as battery and supercapacitor electric vehicles, are equipped with more than one power source in the hybrid propulsion system, which can save more energy through efficient power combinations. Lithium-ion batteries are the preferred choice for energy storage in electric vehicles due to their superior energy density and cost-effectiveness. Nevertheless, matching the required power input and output leads to in an unwanted growth in the size of the battery, and the frequent charge and discharge operations adversely affect battery life. To circumvent the challenges mentioned above, researchers have suggested the development of combined energy storage solutions that merge the capabilities of both batteries and supercapacitors. The supercapacitor aims to increase the range that electric vehicles can travel, improve their dynamic performance, prolong the lifespan of batteries, and mitigate the strain on batteries during rapid energy spikes by exploiting the high instantaneous power capability of the supercapacitor. Energy management strategies are formulated based on the capabilities of the hybrid energy storage system comprising batteries and supercapacitors, aiming to allocate the optimal power output from the battery and supercapacitor for improved vehicle performance, energy efficiency, and battery lifecycle. There are several energy management strategies in battery and supercapacitor hybrid electric vehicles, which include Dynamic Programming, Equivalent Consumption Minimization Strategy, and Model Predictive Control. The state-of-the-art method is the

reinforcement learning (RL) based energy management strategy. This includes Q-learning, Double Q-learning, Deep Q-networks, and Deep deterministic policy gradient methods, which have been studied in the problem of energy management for the electric vehicles equipped with hybrid energy storage system.

However, there are still challenges in RL-based electric vehicle energy management strategies that need further study. First, the RL methods need many iterations, from 15000 to 150000, to converge. The RL methods mimic human brain activity to use experiences to update the agent, which causes a long training time and computational burden. Thus, reducing the number of iterations becomes a critical challenge for RL-based energy management strategies. Furthermore, there is a lack of study on the real-time implementation of RL algorithms for energy management using existing hardware from vehicles. Although the real-time hardware implementation of conventional energy management strategies is abundant, the existing hardware in vehicles is challenging to meet the high-performance computing requirement of RL-based energy management strategies. Advanced hardware products, like GPU from NVIDIA and TPU from Google, are always used in computer vision, natural language processing, and AI supercomputing, and it is not economically viable to apply these expensive products for electric vehicle energy management. Developing the relevant feasible performance learning techniques is very important to reduce hardware implementation costs for electric vehicle energy management.

This dissertation addresses these challenges and seeks to contribute to the field in several ways. Firstly, a dedicated driving cycle for EVs is developed, providing a realistic

representation of driving conditions. Based on this EV driving cycle, an energy management strategy is developed by Q-learning to increase energy efficiency and minimize battery aging. Then, an advanced energy management strategy is designed by imitation learning to decrease the learning time and computational cost associated with Q-learning. Furthermore, a novel Lithium-Sulfur battery with bilateral solid electrolyte interphase is studied and adopted to lower the operating cost of EVs. Lastly, to solve the continuous control problem, a deep reinforcement learning-based energy management strategy is introduced, which incorporates the digital twin technology for real-time implementation. Through these contributions, this dissertation seeks to contribute to the comprehension and practical implementation of energy management methods within the hybrid energy storage systems utilized in electric vehicles. The research findings hold the potential to drive more sustainable and efficient electric vehicle technology while considering practical implementation and cost-effectiveness.

DEDICATION

To my mother.

爱你，老妈~

ACKNOWLEDGEMENT

I would like to convey my sincere appreciation to Dr. Jeff Zhang, my advisor, for being an outstanding mentor during my time at Clemson. His insightful vision, timely guidance, and motivating support played a crucial role in successfully completing my Ph.D. research. Dr. Zhang imparted invaluable lessons on research methodology and the art of presenting research effectively. Working and studying under his guidance has been a tremendous honor and pleasure, and I am truly grateful for the invaluable contributions he has made to my academic journey. Also, thank you to Dr. Bin Xu as a co-supervisor. His scientific and research knowledge in the domain of experimentation and contributions to reinforcement learning have profoundly inspired my study. Gratitude to Dr. Xuan Zhao for his unreserved support throughout my master's and Ph.D. studies.

I extend my heartfelt gratitude to my committee members, Dr. Benjamin Lawler, Dr. Srikanth Pilla, Dr. Rajendra Singh and Dr. Zheyu Zhang, for their invaluable and timely suggestions throughout my research. Their support and willingness to share their professional expertise were instrumental whenever I faced challenges in my projects. I am sincerely thankful for their guidance and assistance.

I would also like to thank my elder brother and sister, Mr. Chang Liu and Mrs. Yaqi Wu. Your support has been invaluable. Thank you for being an important part of my life.

Finally, I want to express my profound gratitude to my parents for their unwavering support. Your love, patience, and encouragement have significantly enhanced this journey, and I am deeply thankful for your presence in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Research motivation

As global emission standards for vehicle exhaust tighten, aiming for approximately 100 g CO_2 per kilometer by 2020–2025, the imperative to curb emissions intensifies. Europe sets an ambitious goal of 95 g CO_2 per kilometer by 2020–2021, with additional debasement of 15% by 2025 and 37.5% by 2030, highlighting the urgency of the matter [1]. These emission reduction objectives are pushing automobile manufacturers towards electrification, a pathway that could satisfy these targets. However, Realizing the full benefits of vehicle electrification requires the application of an energy management strategy (EMS) for the electric drive systems of these vehicles. Traditionally, the creation and verification of EMSs utilize internationally recognized driving cycles, such as the New European Driving Cycle (NEDC), Highway Fuel Economy Test Cycle (HWFET), Urban Dynamometer Driving Schedule (UDDS), Federal Test Procedure (FTP), European Economic Commission 15-mode test cycle (ECE 15), Japanese Industrial Standards Committee 08 test cycle (JC08), and Japan 10/15 mode test cycle (J10/15) [2], [3], [4], [5], [6]. However, the real-world performance of electric vehicles often diverges significantly from estimations based on these ISDCs. Factors such as electric motor drivetrain, regenerative braking, transmission, torque delivery, energy conversion, lead to disparities in energy consumption estimations [6], [7], [8], [9]. Consequently, energy consumption

estimates for EVs under ISDCs are inherently inaccurate. Hence, there is a need for the creation of a specialized driving cycle tailored specifically for EVs.

In the electrified propulsion system, the energy storage system (ESS) is a vital part. In the automotive industry, the lithium-ion battery (LIB) stands out as the prevailing choice for ESS owing to its remarkable energy density. Nonetheless, concerns persist regarding the cycling life and power density of LIBs, which impede the widespread adoption of battery EVs [10]. Conversely, supercapacitors (SC) offer substantially longer lifespan and higher power density, albeit at the expense of lower energy density [11]. Another well-researched ESS is the fuel cell. Although a fuel cell has a moderate power density and high energy density, it has the highest capital cost among the three ESSs described above [11]. Problems with hydrogen storage also restrict the application of fuel cells. Hence, a perfect energy storage system (ESS) capable of meeting all application requirements remains elusive. A growingly popular solution for striking a balance between energy density, cycling life, power density, and cost is the adoption of hybrid energy storage systems (HESS). The associated EMS is formulated to leverage the strengths of different ESSs while mitigating their limitations. It manages the power from diverse energy sources to satisfy the driver's requirements across different traffic scenarios. Enhancing energy utilization and reducing battery wear are among the primary goals of EMS. Theoretically, ensuring the fulfillment of the driver's power requirements leads to a nonlinearly constrained optimization problem when optimizing the control of multiple energy sources simultaneously at each time-step [12]. Various control strategies and mathematical tools facilitate the attainment of both global and local optimal solutions for such optimization

problems. However, the computational load of such optimization exceeds the capabilities of on-board microprocessors. Besides, the designed EMSs may be suitable for some specific driving conditions but cannot handle the complexity in the real-world driving environment.

1.2 Research problem statement

This dissertation aims to develop an energy management system for battery & SC EV using deep reinforcement learning (DRL). The proposed approach is capable of real-time implementation in control systems to achieve optimal control outcomes. The proposed EMS will take advantage of different ESSs to accomplish optimal management on a system-wide level. Aiming at comprehensively develop the inherent characteristics of EVs and design a more accurate EMS for EVs, a dedicated EV driving cycle is needed. according to the proposed EV driving cycle, a Q-learning based EMS is designed to allocate the power distribution. To reduce the training time, the imitation learning method is integrated with the Q-learning-based EMS. By imitating the existing expert experiences and heuristic rules, imitation learning can boost the training process to reduce iterations and time costs. Although the imitation learning algorithm makes the Q-learning-based EMS capable of handling real-time control, it is still trained through a fixed EV driving cycle that cannot deal with the complexity and randomness in real-world traffic conditions. Also, the Q-learning-based EMS is designed to deal with the discrete optimization questions. Therefore, the digital twin technology and DRL algorithms will be introduced in the dissertation. The DRL-based EMS can solve continuous optimal energy management problems. However, training neural networks in the DRL algorithms causes a heavy

computational burden, which exceeds the computing capacity of vehicle on-board microprocessor. The digital twin is adopted to handle the computing workload and adjust the parameters of the proposed EMS to adapt to different driving conditions. The digital twin model operates within a virtual environment, mirroring real-world data, and it absorbs real-time information to improve the adaptiveness of the proposed EMS to achieve better control performance under different traffic conditions.

This comprehensive research approach integrates cutting-edge technologies, blending DRL algorithms, imitation learning, and digital twin technology. By doing so, it strives to bridge the gap between theoretical advancements and practical implementation, pushing the boundaries of what is achievable in real-time energy management for EVs.

1.3 Research challenges

There are several distinctive challenges in designing and implementing EV driving cycle construction, table-based reinforcement learning (RL) energy management strategy (EMS), and neural network-based RL EMS. An essential challenge in constructing driving cycles is ensuring the representativeness of the developed driving cycle, which could be compromised by inaccurate results from driving segment classification. Various classification algorithms, such as K-means and fuzzy C-means, have been employed in prior research. However, in cases where driving segments must be classified into different categories or where the separation between cluster centers is minimal, the clustering outcomes may not be optimal, which can lead to local optimality.

Challenges also exist in RL-based EMS. Although RL-based EMS has been gaining significant attention recently, it possesses substantial obstacles during real-time implementation. Several methods have been adopted to accelerate the training process including variable learning rate, which may lead to an unstable control effect after training and heavy reliance on past experiences that lowers the control performance. Warm-start initialization helps to decrease training intricacies, but it might not achieve globally optimal solutions.

Another challenge lies in the consideration of battery degradation costs, which are a significant part of EV's total operating costs. The EMS should be devised in a manner to minimize battery aging. Additionally, state-of-the-art battery technologies, such as the bilateral solid electrolyte interphase (SEI) Lithium-Sulfur (Li-S) battery, should be incorporated to reduce battery investment and prolong battery life [13].

Moreover, any designed EMS should be fine-tuned according to the delay and model discrepancy when being implemented to solve real-time control. Importantly, existing RL-based EMSs, though having been designed through simulation by high-performance computation platforms, have seldom been applied in industry-level hardware due to computational burdens.

1.4 Contributions

There are five contributions from this dissertation, which are listed below:

1. Development and Verification of EV Driving Cycle:

A specific EV driving cycle is developed and verified systematically. The proposed EV driving cycle forms the foundation for precise EMS for EVs, ensuring the realism and applicability of the proposed solutions in real-world driving scenarios.

2. Construction of RL-based EMS Training and Li-S Battery Application:

   Proposing a Q-learning EMS for EVs, the study delves into EV dynamics, emphasizing advanced Li-S battery with new SEI film and diverse HESS configurations. The Q-learning EMS with adaptive learning, dynamically optimizes real-time energy distribution between the battery and SC, eliminating preset rules. Simulations on the established EV driving cycle demonstrate its effectiveness, notably in reducing energy consumption and battery degradation. Results showcase Q-learning's practical relevance for enhancing EV energy systems.

3. Acceleration of Q-learning Based EMS with imitation learning technology:

   The integration of the imitation learning method significantly accelerates the training phase of Q-learning based EMS. This approach effectively reduces EV operating costs. By leveraging existing expert experiences and heuristic rules, the training cost of Q-learning-based EMS is reduced, making it more efficient in real-time control scenarios.

4. Integration of DRL-based EMS with digital twin Technology:

   DRL-based EMSs are studied and integrated with digital twin technology. This integration creates a robust framework capable of addressing continuous optimal energy management challenges. By blending cutting-edge DRL algorithms with

digital twin capabilities, the proposed EMS achieves a higher level of adaptability, accuracy, and efficiency in varying driving conditions.

5. Establishment of Hardware-in-the-Loop (HIL) Test Platform:

   A HIL test platform dedicated to EVs equipped with battery and SCs is established. This platform serves as a vital tool for the validation and verification of the proposed energy management methods. Through real-time simulations and hardware interaction, the efficacy and reliability of the proposed methods are tested, ensuring their practical viability and paving the way for future advancements in EV technology.

These contributions collectively enhance the understanding and application of EMSs in the realm of EVs, facilitating the transition toward more efficient, adaptive, and sustainable transportation solutions.

CHAPTER 2

LITERATURE REVIEW

The following chapter unfolds as a critical synthesis of the extensive body of literature that surrounds the EV's EMS, shedding light on key theories, methodologies, and findings to contextualize and guide the present study.

2.1 EV driving cycle

A driving cycle denotes a profile of speed versus time that characterizes typical driving patterns observed in a particular city or region [14], [15].

The driving pattern within each city or region is distinct owing to variations in factors such as location, vehicle types, traffic density, ownership rates, urban road configurations, and road network structures, all of which profoundly influence the driving cycle [8], [16]. Hence, numerous driving cycles have been devised across various nations, as shown in Table 2.1. The creation of a driving cycle to represent real-world traffic scenarios involves three primary stages: selecting a test route, collecting data, and constructing the driving cycle. Table 2.1 outlines commonly utilized approaches for the three phases involved in the creation of driving cycles.

During route selection, it's essential to choose a road segment that accurately represents the general road conditions and traffic patterns. The chosen test routes could encompass a variety of road types, including expressways, arterial roads, sub-arterial roads, and branch roads. Key factors in route selection encompass road gradients, traffic volume,

origin-destination (O-D) pattern, travel time, population density, and location of business districts. In some studies, a particular vehicle's trajectory is chosen as the test route. For instance, the Hamburg driving cycle study selected 12 routes from various buses [17], while the Dublin driving cycle study utilized 1485 journeys from seven Mitsubishi EVs [7]. In the Florence driving cycle research, a total of 12 different electric vehicles were chosen to cover the 2500km route [18]. The Tehran driving cycle study utilized a passenger car circuit for its test route [19]. In the Fuzhou driving cycle study, 18 bus routes were chosen for analysis [20]. Some studies opt for test routes based on prevailing traffic scenarios. For instance, the Edinburgh study focused on six primary city center arteries with the peak daily traffic volumes [21], while the Celje study chose a main artery representing average daily commuter traffic [22], Similarly, the Mashhad study selected two important roads that has the highest traffic volumes [23]. In Aleppo, test routes were chosen based on traffic congestion levels and distance between the downtown and the University of Aleppo [15]. Colombo adopted test routes by daily traffic volumes [24], and Tianjin chose routes with two different traffic scenarios [25]. Several studies recognize home-to-work and work-to-home journeys as significant components of daily travel. Consequently, test routes are often chosen by O-D patterns or specifically home-to-work commute. For instance, in Winnipeg, test routes were selected based on a fleet of 76 vehicles' specific O-D options [26], and a similar approach was taken in Sri Lanka where test routes were designed by O-D pairs [24]. Additionally, some studies employ random selection methods based on experiential knowledge. For instance, in Pune, five main roads covering round 55 km were randomly chosen as test routes [27]. In Chennai, a corridor

spanning 14 km near the Indian Institute of Technology campus was designated to make the test route [28]. Furthermore, certain driving cycles are formulated by taking into account factors such as route utilization, traffic conditions, road types, origin-destination pairs, and travel durations, as demonstrated in Singapore [29], Hong Kong [8], Hefei [30], Eleven Chinese cities [31], and California [32].

When examining average real-world driving patterns, three primary approaches are commonly employed to gather test data: on-board measurement, chase car, and a hybrid approach. The dependability of the on-board measurement method is contingent upon factors such as mileage covered and the number of test vehicles utilized. Greater amounts of collected test data lead to test results that more closely align with real-world driving cycles. Therefore, a large size of data collection is vital, as shown in the research of Dublin [7], Hamburg [17], Winnipeg [26], Florence [18], Colombo [24], Tehran [19], Chennai [28], Aleppo [15], Sri Lanka [24], Fuzhou [20], and Tianjin [25]. The chase car method is a commonly adopted technique for collecting speed-time data of real-world driving cycles. In this approaches, a test vehicle follows a target vehicle, following its driving pattern within the traffic flow. If the target vehicle stops or turns to inaccessible, the test vehicle switches to following other target vehicles randomly. A skilled driver is required to prevent data discrepancies arising from potential speed variations between the chasing vehicles and target or potential interference between them. Numerous driving cycle studies have utilized this chase method, such as Celje [22], Mashhad [23], Edinburgh [21], Pune [27], Hefei [30], Beijing, Shanghai, Chongqing, Tianjin, Chengdu, Changchun, Ningbo, Mianyang, Jilin, Jiutai, and Zitong [31]. The hybrid approach combines on-board measurement and car

chase methods, and it has been applied in the driving cycles obtained for Singapore [29] and Hong Kong [8].

The phase of constructing a driving cycle includes segmenting original data into driving segments, constructing criteria for selecting segments, synthesizing different alternative driving cycles by connecting selected driving segments, and establishing a specific driving cycle to meet the assessment criteria. The driving segments are segmented based on predefined speed and distance intervals to form the driving cycle research in Chennai [28], Tehran [19], California [32], Winnipeg [26], or two continuous idling periods in Sri Lanka [24] and eleven China cities [31]. The selection criteria for driving segments are primarily established through simulation and matching. The matching method can involve employing random selection or clustering algorithms to select driving segments. This methodology was utilized in developing driving cycles in Edinburgh [21], California [32], Celje [22], Winnipeg [26], Florence [18], Tehran [19], Mashhad [23], Pune [27], Chennai [28], Aleppo [15], Sri Lanka [24], Singapore [29], Hong Kong [8], Fuzhou [20], Tianjin [25], and eleven Chinese cities [31]. The simulation method operates under the assumption that a driving cycle adheres to the characteristics of a Markov chain, with events occurring in a dedicated sequence. This approach was employed in driving cycle studies conducted in Colombo [33], Changchun [34], and Hefei [30].

Within the driving cycle development phase, the K-means clustering algorithm stands out as the preferred method to classify the driving segments because of its simplicity and high efficiency [35]. However, if driving segments have to be clustered to multiple categories or the distance between cluster centroid is minimized, the results obtained from

the K-means algorithm may exhibit instability and tend to converge to a local optimum. Then some studies used the Support Vector Machine to solve this issue [36], but still experienced a similar problem of being trapped in a local optimum.

Table 2.1 Development methodologies of existing driving cycles

Region	Duration	Vehicle type	Test route selection	Data collection	Driving cycle construction method	Reference
Dublin	1800s	EVs	Select 1485 journals	OBM	Learning vector quantization neural networks	[7]
Edinburgh	826s	ICEPVs	Identify six principal central urban roads with the most significant daily vehicle flow"	CC	Chosen arbitrarily according to the index of traffic volume	[21]
California	600s	ICEPVs	Choose city streets, outskirts avenues, and expressways	Floating cars data collected by CDTPMS	Arbitrarily chosen from the database of speed intervals	[32]
Celje	2400s	ICEPVs	Spanning 12.9km, this major route serves as the primary entry and exit point for the city, beginning and concluding at a designated street	CC	Randomly select	[22]
Hamburg	1200s	ICEBs	Choose 12 urban bus lines to encompass the entire city's region.	HM.	Choose short journeys based on the quality of Fit metric	[17]
Winnipeg	3309s	PLEVs	Select routes according to origins and destination of 76 volunteers	OBM	Arbitrarily choose the quantity of categorized short trips	[26]
Florence	3000s	EVs	No predefined itinerary	OBM	Choose by chance in line with matching criteria	[18]
Tehran	1533s	ICEPVs	Select private car track as test route	OBM	Choose closest short journey toward group hubs	[19]

Mashhad	1000s	ICEPVs	Identify two primary pathways sharing a common start and end point, based on their maximum traffic flow	CC	Creating a varied assortment of short trips by the error of 10 fundamental criteria.	[23]
Pune	1533s	ICEPVs	Select 5 major roads about 55km	CC	Choose short tours within predefined flexibility margins	[27]
Chennai	1488s	ICEPVs & motorcycles	Choose a stretch roughly 14 km surrounding the Indian Institute of Technology campus	OBM	Arbitrarily choose short journeys based on evaluation standards	[28]
Aleppo	Urban 2900s Motorway 900s	ICEPVs	The route encompass the entire University of Aleppo campus, alongside specific paths connecting the City Centre with the University of Aleppo, prioritizing routes based on traffic density and overall distance	OBM	Choose based on specific attribute criteria	[15]
Colombo	1200s	ICELVs	Two major trips: intercity and intracity	OBM	Markov chain	[33]
Sri Lanka-expressway	1200s	ICELVs	Choose routes by O-D pairs	OBM	Choose sections based on the median durations of travel	[24]
Singapore	2344s	ICEPVs	Develop 12 principal pathways factoring in usage patterns, orientation, central business district, and both inner and outer circular roads	HM	Partially randomize the selection of short journeys based on categorized groups	[29]
Hong Kong	1200s	ICEPVs	Choose four city paths, one route in the outskirts, and four major roadways according to the highest yearly average daily traffic counts	HM.	Select micro-trips on a random basis following specific evaluation guidelines	[8]

Hefei	651	ICEPVs	Choose five exemplary routes based on factors such as traffic patterns and driving duration	CC	Markov chain	[30]
Fuzhou	1200s	ICEBs	Select 18 buses routes	OBM	Choose short journeys based on the proximity of each journey to the central point of the cluster	[20]
Tianjin	1100s	EVs	Choose trial paths in actual traffic scenarios	OBM	Select micro-trips randomly based on the outcomes of linear discriminant analysis categorization	[25]
Eleven Chinese cities	1200s	ICEPVs	Choose testing paths based on criteria such as traffic density, urban scale, types of roadways, commercial zones, and dwelling districts	CC	Pick randomly based on feature parameters	[31]

Note: ICEPVs-Internal Combustion Engine Passenger vehicles; ICELVs-Internal Combustion Engine Light vehicles; ICEBs-Internal Combustion Engine buses; PLEVs-Plug-in light duty EVs; OBM-On-board measurement; CC-Chase car; HM- Hybrid method of On-board measurement and Chase car; CDTPMS-California Department of Transportation Performance Measure System

2.2 RL-based EMS

On the basis of established EV driving cycle, the design of the EMS for EV can be conducted. EMSs are categorized into rule-based EMS, optimization-based EMS, and learning-based EMS based on the adopted methods [37]. Rule-based EMSs are created using heuristic rules or expert knowledge [12]. offer readily and intuitive implementable solutions for energy management challenges [38]. But, the effectiveness of Rule-based EMSs relies on expertise heavily and may struggle to address complex scenarios [39]. Reference [40] introduces a rule-based EMS capable of controlling both increased thrust capabilities and energy recapture through braking across various scenarios. Utilizing different modes, the EMS in [40] enhances energy efficiency by 8-25%. In contrast to rule-based EMSs, optimization-based EMSs are formulated using various optimal methods, enabling superior control outcomes and enhanced energy efficiency. Reference [41] proposes a method that integrates fuzzy logic control, wavelet transform, and neural networks. This approach leads to a 44.22% improvement in regenerative braking energy and reduces battery aging by 18% in a testing system comprising a 72 V battery and 96 V SC HESS.

Reference [42] suggests that an optimal power distribution strategy based on convex optimization yields improved control effect, resulting in a 5.7% reduction in energy consumption. Dynamic programming (DP), a prominent optimization technique, has found extensive application in optimizing EMS, often serving as a benchmark for evaluating alternative methods [43]. Nevertheless, its high computational complexity poses challenges for real-time implementation [44], [45].

Study of RL-based EMSs for electric vehicles is emerging. It utilizes a temporal difference learning approach to create optimal control for hybrid electric vehicles (HEV) [46]. This method exhibits a more faster convergence rate and delivers superior performance even in environments lacking Markovian properties. In [47], a RL-based EMS for the PHEV is introduced by data-driven method, taking into account the charging status throughout the trip to achieve near-optimal outcomes. Similarly, in [48], an RL-based EMS integrated with a terrain knowledge is developed for the PHEV, aiming to reduce battery usage and fuel consumption of internal combustion engines. Additionally, reference [48] incorporates trip distance as a system state to build the RL environment, the data illustrates a clear relationship between the miles yet to be traveled and the total amount of energy required. In reference [49], a Q-learning based method is devised to decrease energy usage and battery wear. Compared to following a predetermined set of rules, the technique proposed in source [49] results in a 1.5-2% increase in the distance the vehicle can travel on a single charge, while also decreasing battery wear by 13-20%. In [50], a hybrid model predictive control and reinforcement learning approach are integrated to develop an EMS for HEVs. The study utilizes Q-learning to tackle the energy management challenge. It builds a speed forecasting model by combining fuzzy logic encoding with a nearest neighbor prediction approach based on past driving data. Meanwhile, reference [51] proposes a reinforcement learning-based energy management system for hybrid energy storage systems. This system incorporates a transition probability matrix that adapts over time using the Kullback-Leibler (KL) divergence rate and a forgetting index. A novel EMS is devised for HEV employing Q-learning, with cloud computing integration to mitigate

the computational load during real-time training [52]. Another approach, referenced as [53] adopts fuzzy logic control to enhance the Q-learning technique to formulate the EMS for HEV. This method adjusts the parameters of fuzzy logic using Q values and integrates a neural network for estimating the action-value function. Furthermore, a hybrid electric tracked vehicle benefits from an online predictive EMS achieved through the fusion of a fuzzy logic controller and RL method [54]. This integration aims to mitigate the impact of prediction inaccuracies. In another study [55] an online RL-driven EMS for HEV is introduced, employing Q-learning. The reward function in this approach is shaped by the weighted values derived from fuel usage and battery energy.

2.3 Imitation Q-learning based EMS

RL method has emerged as a powerful technique for training agents to perform tasks in complex and dynamic environments. Although the RL-based EMS has achieved progresses in this field, there still are obstacles to compensate the application of RL-based methods. Traditionally, RL algorithms require substantial exploration and trial-and-error to discover optimal policies. However, this process can be time-consuming, costly, and inefficient, especially in scenarios where an expert already possesses the desired behavior. Imitation learning, also known as learning from demonstrations or apprenticeship learning, offers a promising alternative by allowing agents to directly imitate the behavior demonstrated by human experts [56] . this approach aims to bridge the gap between human expertise and the learning process of autonomous agents. By observing and imitating expert demonstrations, agents can learn to perform tasks in a manner similar to the demonstrated behavior, reducing the need for extensive exploration. Imitation learning techniques can be

broadly classified into two primary categories: behavioral cloning and policy learning. Behavioral cloning algorithms learn to mimic the expert's actions exactly, while policy learning algorithms learn to predict the expert's policy, which is a mapping from states to actions. The techniques of imitation learning have been effectively utilized across numerous domains, including robotics, autonomous driving, game playing, and natural language processing. For example, imitation learning has been used to train robots to perform tasks such as walking, picking up objects, and manipulating tools. It has also been used to train self-driving cars to navigate roads and avoid obstacles.

2.4 Digital twin integration and DRL-based EMS

The proposed imitation Q-learning based EMS can obtain optimal control performance and can be able to apply for real-time control problem. However, the proposed method is a great solution for discrete problem but cannot handle the continuous problem. To solve the continuous optimal problem, this dissertation integrates the digital twin technology and DRL method. The NASA Apollo space program pioneered the concept of the 'digital twin,' utilizing two identical space vehicles. One vehicle, stationed on Earth, mirrored, simulated, and forecasted the conditions of its counterpart in space. This Earth-bound vehicle served as the twin to the spacecraft executing the mission in space [57]. In reference to the work cited in [58], a simulation digital twin model for EVs was proposed. This model aims to forecast and analyze the impacts of various parameters on the performance attributes of EVs. Additionally, reference [59] presented a digital twin model focused on the temperature-energy consumption dimension, derived from the conventional model. This model is utilized to forecast the energy consumption of EVs and validate the method's

feasibility. In contrast to table-based RL methods, approximate-based RL methods driven by neural networks are gaining popularity in the domain of EMS. A deep Q-network (DQN) based EMS for PHEVs is proposed that the performances of both standard DQN and dueling DQN are evaluated and compared [44]. The comparison results indicate that the dueling DQN can converge faster than the normal DQN. The deep deterministic policy gradient (DDPG) method are incorporated to achieve continuous control solutions in EMS [45]. The DDPG-based EMS for PHEV does not ask for the discretization of states and actions [45]. The proximal policy optimization (PPO) is known as a DRL algorithm to solve the continuous action space, which illustrates that the agent can update the policy robustly through the local controller in the training process [60]. A DRL-based EMS is suggested for electric buses based on soft actor-critic method, aimed at effectively managing the energy distribution among various power sources. Comparative analysis between DQN-based and soft actor-critic-based EMS demonstrates that the soft actor-critic EMS exhibits superior optimization performance and quicker convergence [61]. The DDPG algorithm is utilized to develop a DRL based EMS. However, issues of overestimation have affected its performance. In response, the Twin Delayed DDPG (TD3) algorithm is integrated to derive the EMS [62]. Comparative analysis with DQN and DDPG indicates that the TD3 algorithm demonstrates superior manage effectiveness.

CHAPTER 3

ELECTRIC VEHICLE DRIVING CYCLE CONSTRUCTION

This chapter introduces a methodical and pragmatic approach to develop a driving cycle that accurately captures the typical driving patterns experienced EVs. The methodology addresses four key aspects: identifying an appropriate testing route, gathering real-world vehicle operational data, processing and analyzing the collected data, and finally synthesizing a representative driving cycle profile. In the step of processing data, the dimensions of motion characteristic parameters are diminished by the principal component analysis (PCA) method. Furthermore, a hybrid algorithm combining the Self-Organizing Map (SOM) and Support Vector Machine (SVM) is utilized for the classification of driving segments. The process of generating the EV driving cycle leverages probabilistic techniques, specifically employing Markov models and Monte Carlo simulations. Crucial factors in identifying the most realistic driving cycle include the relative deviation from actual data, specific performance metrics, and the probabilistic distributions of vehicle speeds and accelerations. Following cycle construction, the characteristic parameters, driving range capabilities, and energy usage are evaluated across the different synthesized driving cycles.

3.1 Research gaps and proposed methods

A driving cycle is a chronological sequence of vehicle speeds that represents typical driving behaviors observed in specific regions or cities over time [14], [15], plays a pivotal role in monitoring various aspects of vehicle performance, including fuel usage and tailpipe

discharges for ICEVs [63], [64]. Additionally, it serves as a fundamental metric for assessing energy consumption, driving range, and EMS of EVs. Furthermore, driving cycles play a crucial role in conducting simulation experiments on a laboratory chassis dynamometer and developing car model simulation techniques. In the realm of vehicle design and the development of next generation vehicles, driving cycles function as uniform measurement processes for certification and evaluation [7], [19].

However, there exists a significant research gap concerning the design and utilization of dedicated EV driving cycles [65]. Most existing study on EV design, control strategies, and EMSs relies on driving cycles established for ICEVs. These ISDCs are often based on data obtained through ICEVs, those are powered by internal combustion engines, but EVs draw energy from batteries and employ electric motors for propulsion. The unique torque and power characteristics of electric motors, distinct from internal combustion engines, contribute to distinctive starting, acceleration, and driving features in EVs compared to ICEVs. Furthermore, EVs showcase distinct braking performance and sensations owing to electric motor regenerative braking systems. These embedded variations in driving and deceleration mechanisms give rise to differences in driving cycles between EVs and ICEVs. Earlier endeavors to formulate driving cycles specific to EVs include the study by Brady et al., where they gathered 1485 driving logs from the Dublin region. They then employed learning vector quantization neural networks to synthesize an 1800-second EV driving cycle based on this data [7]. In another effort, Smith et al. collected real-world driving data from 76 plug-in hybrid electric vehicles in Winnipeg. Their approach involved randomly selecting and combining shorter micro-trip segments to construct a representative driving

cycle [26]. The dynamic characteristics and temporal aspects of the driving pattern have been overlooked by these two methods. To bridge the gap, this dissertation integrate the Markov chain and Monte Carlo techniques to establish the EV driving cycle. This technique properly reflects the dynamic aspects of real-world driving patterns and the temporal correlations before and after certain events.. The detailed methodology for this development process is illustrated in Fig 3.1.

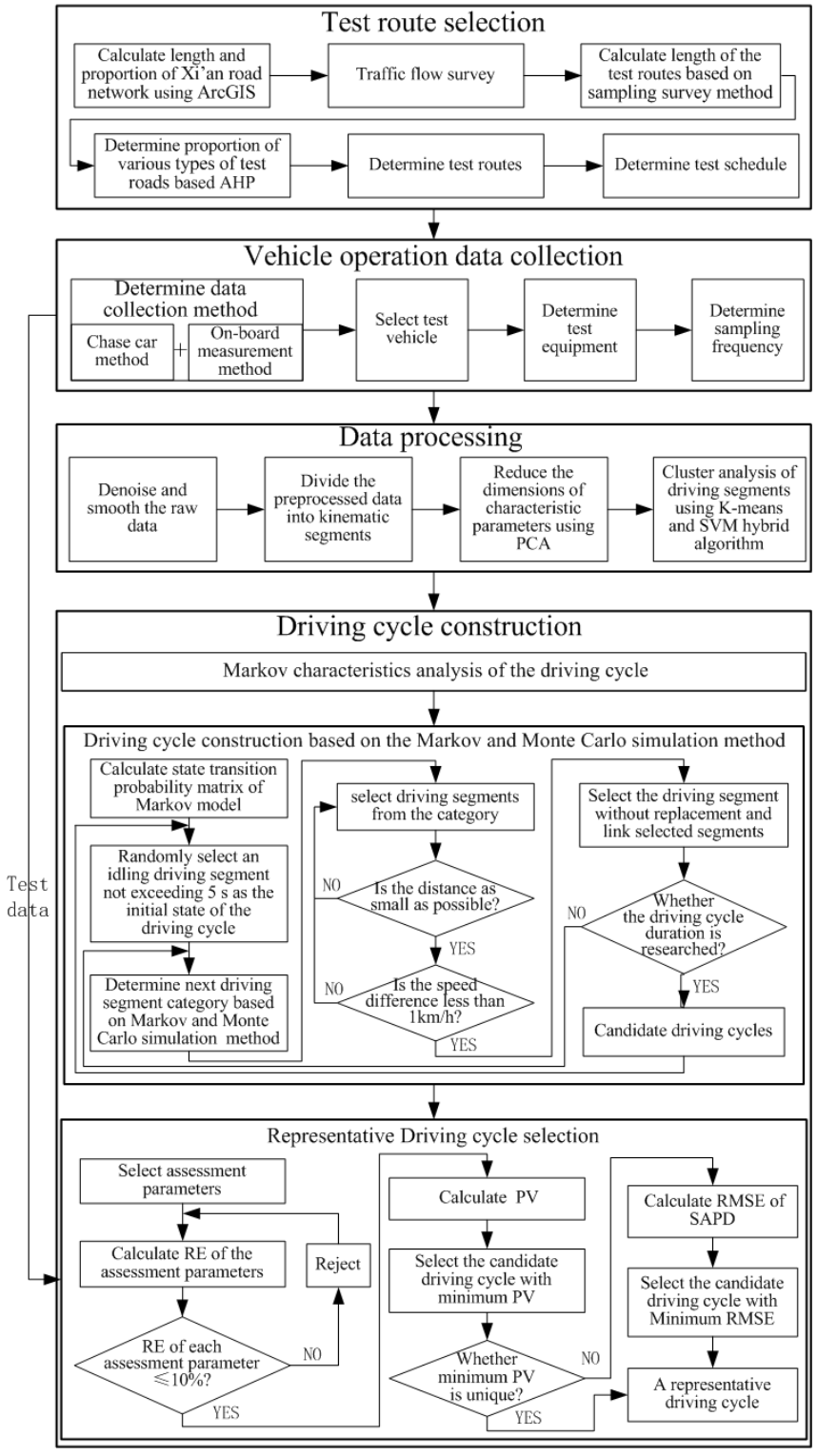

Fig 3.1 The Methodology for EV driving cycle construction

3.2 Test route selection

The selection of an appropriate test route is critical, as it should accurately reflect the driving patterns in a city or area, ensuring the collected representativeness of the data [16], [23]. In the selecting test route, it is normally essential to evaluate parameters such as road type, urban layout, traffic volume, driving speeds, distribution of population density, and origin-destination (O-D) patterns [16], [23], [24], [66]. In prior research efforts, the process of identifying influencing factors and subsequently selecting test routes has largely relied on subjective and qualitative assessments by the researchers themselves. The typical approach has involved leveraging the researchers' personal familiarity and comprehension of the local traffic conditions to inform their choices regarding suitable test routes, relying on experience rather than quantitative analysis and scientific methods [8], [16]. This dissertation introduces a novel approach that combines qualitative and quantitative analyses to design test routes. Firstly, an analysis and calculation of the road network distribution and composition of road types were conducted. Subsequently, an investigation into the traffic flow across different types of roads during various time periods was carried out. Utilizing data on the road network layout and traffic flow patterns, a sampling survey methodology is employed to determine the appropriate lengths for the test routes. Additionally, the Analytic Hierarchy Process (AHP) technique is applied to establish the proportional composition of various road types to be included within the selected test routes. With the analysis and criteria established, the specific test route is devised, taking into consideration a range of factors and actual road conditions. This comprehensive approach integrates quantitative data with qualitative factors. By synthesizing road network

distribution, traffic flow data, and various road types, the test routes were meticulously designed. This methodological fusion of quantitative precision and qualitative understanding ensures that the selected routes comprehensively represent actual driving conditions, providing a solid foundation for meaningful data collection and analysis.

The process begins with the formulation of a multi-level hierarchical model, where the overall objective is divided into three tiers: the overarching goal, the set of evaluation criteria, and the alternative scheme options, as illustrated in Fig 3.2.

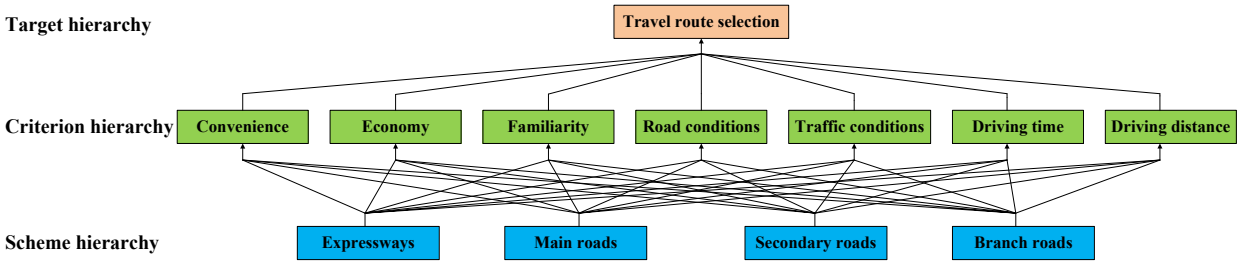

Fig 3.2 The hierarchical structure model

The next step involves creating a pairwise comparison matrix. This matrix captures the relative importance or priority of each element within a given tier, when evaluated against the higher-level tier it belongs to. The values in the judgment matrix are quantified based on the Eq (3.1) provided.

$$A = (a_{ij})_{n*n}, a_{ij} > 0, a_{ij} = \frac{1}{a_{ij}} (i, j = 1,2, \dots, n) \tag{3.1}$$

In the equation provided, a_{ij} represents the ratio of the importance of factor i to factor j relative to the upper hierarchy, and a_{ji} represents the ratio of the importance of factor

j to factor i to the upper hierarchy. Following Saaty's recommendation [67], the value of a_{ij} typically ranges from 1 to 9, or its reciprocal, depending on the scale chosen.

The third phase involves assessing the logical consistency of the pairwise comparison matrices constructed for each criterion in the hierarchy. The consistency of each matrix is evaluated by calculating a consistency index, as expressed through the Eq (3.2).

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{3.2}$$

In the equation provided, $\lambda_{\max}$ represents the biggest eigenvalue of the judgment matrix, and n represents the dimension of the judgment matrix.

The random consistency index RI is used to assess consistency, as depicted in Eq. (3.3). The smaller the CI, the larger the consistency.

$$RI = \frac{CI_1 + CI_1 + \cdots + CI_n}{n} \tag{3.3}$$

The random consistency index RI is associated with the judgment matrix. Typically, as the order of the matrix increases, there's a higher likelihood of random deviation affecting consistency. The corresponding relationship is detailed in Table 3.1.

Table 3.1 The value of the random consistency index

Order of matrix	1	2	3	4	5	6	7	8	9	10
RI	0	0	0.58	0.90	1.12	1.24	1.32	1.41	1.45	1.49

Considering that consistency deviation might arise due to random factors, it is essential to compare the consistency index CI with a threshold to determine whether the judgment matrix can satisfy the consistency. The coefficient is expressed in Eq. (3.4).

$$CR = \frac{CI}{RI} \tag{3.4}$$

In general, if CR is less than 0.1, the judgment matrix is deemed to pass the consistency test, indicating satisfactory consistency. Otherwise, if CR exceeds 0.1, the judgment matrix lacks satisfactory consistency.

The CR of each judgment matrix was determined as follows: {0.0320, 0.0171, 0.0039, 0.0077, 0.0079, 0.0265, 0.0265, 0.0039}. Since all of the CR are less than 0.1;, the consistency of judgment matrices is regarded satisfactory.

The fourth step involves calculating the weight of each option in relation to the overall decision goal, as given in Eq. (3.5).

$$\omega = \begin{bmatrix} \omega_1^T \\ \omega_2^T \\ \omega_3^T \\ \omega_4^T \\ \omega_5^T \\ \omega_6^T \\ \omega_7^T \end{bmatrix}^{\mathrm{T}} \times \omega_0 = \begin{bmatrix} 0.1112 & 0.2290 & 0.3014 & 0.3584 \\ 0.1089 & 0.3512 & 0.3512 & 0.1887 \\ 0.2761 & 0.3056 & 0.3056 & 0.1127 \\ 0.4765 & 0.2879 & 0.1547 & 0.0810 \\ 0.4203 & 0.2685 & 0.1899 & 0.1213 \\ 0.4531 & 0.2616 & 0.1671 & 0.1182 \\ 0.1089 & 0.1887 & 0.3512 & 0.3512 \end{bmatrix}^T \times \begin{bmatrix} 0.1009 \\ 0.1181 \\ 0.0641 \\ 0.1181 \\ 0.2810 \\ 0.1589 \\ 0.1589 \end{bmatrix} = \begin{bmatrix} 0.3054 \\ 0.2652 \\ 0.2455 \\ 0.1839 \end{bmatrix} \tag{3.5}$$

In Eq. (3.5), ω_0, ω_1, ω_2, ω_3, ω_4, ω_5, ω_6, ω_7 represent the weights of each option in relation to the overall decision objective.

Finally, the designed test route is 38.46 km in total. The percentage and length of different kinds of roads are indicated in Table 3.2.

Table 3.2 Proportion and length of various types of test roads

Road type	Expressways	Main roads	Secondary roads	Branch roads
Proportion (%)	30.54	26.52	24.55	18.39
Length (km)	11.75	10.20	9.44	7.07

During the determination of test routes, comprehensive factors were considered, including the urban road network structure, central business district, O-D pattern, population density, regional disparities, traffic volume, test sample size, percentage of various test road, as well as other factors including the placement of test equipment and EV charging stations. Consequently, a ring test route, 38.26 km in length, was designed for this study. The finalized test routes are depicted in Fig 3.3.

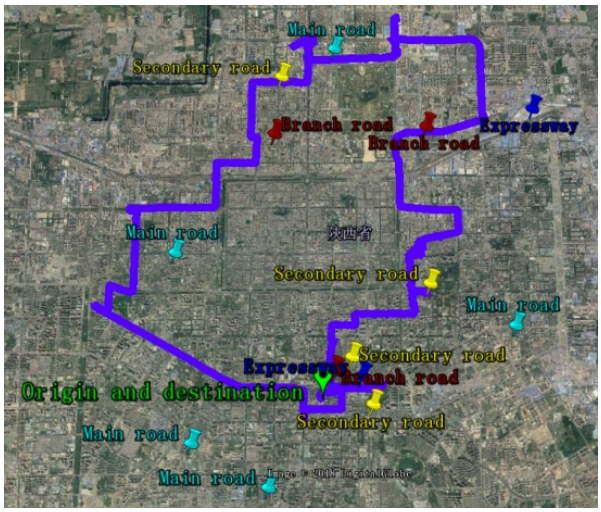

Fig 3.3 Test routes

3.3 Data collection and processing

The data collection is based on my previous work [35], [68]. Within the test, a pure electric taxi vehicle was chosen as the test vehicle to minimize vigilance of the driver of the chased vehicle [16]. The type of collected data are outlined in Table 3.3.

Table 3.3 Driving data type

Parameters	Sign	Unit
Driving time	T	s
Velocity	v	km/h
Driving distance	s	km
Longitude	long	°
Latitude	lat	°
Altitude	H	m
Acceleration in X-axis	ax	g
Acceleration in Y-axis	ay	g
Acceleration in Z-axis	az	g
Electric motor torque	eT	N/m
Electric motor speed	eN	r/min
Electric motor power	eP	kw
Engine torque	T	N/m
Engine speed	N	r/min
Engine power	P	kw
Power bus voltage	U	v
Power bus current	I	A
Break pedal status	strain/loosen	/

The sampling frequency is a critical determinant of data authenticity and accuracy, which impacts the establish of the driving cycle. An excessively high frequency not only escalates the workload in processing data, but also generates spikes in the acceleration phase, leading to an overly aggressive driving cycle representation. Conversely, an overly low sampling frequency might filter out crucial data characterizing vehicle acceleration and deceleration. In this dissertation, the selection of the appropriate sampling frequency was meticulously assessed. The focus was on the acceleration characteristic derived from speed-time differentials. To establish the ideal frequency, the correlation coefficient between estimated acceleration at sampling intervals of 10 Hz, 1 Hz, and 0.1 Hz, and the observed acceleration collected by the inertial navigation system was evaluated. The results revealed a correlation coefficient of 99.51% at 10 Hz, 98.24% at 1 Hz, and 47.47% at 0.1 Hz. While a 10 Hz sampling frequency displayed the highest correlation, the associated data processing workload was tenfold higher than that of 1 Hz. Consequently, considering both data accuracy and practical burden, a sampling frequency of 1 Hz was deemed optimal for the vehicle operation data collection process. This choice strikes a balance between data accuracy and the feasibility of data processing, ensuring a robust foundation for the construction of the driving cycle. The collection of driving data is conducted through test vehicles traversing predefined routes for a duration of seven days under authentic traffic conditions. The acquired real-world driving data are subsequently categorized into peak traffic and off-peak traffic datasets based on the results of a traffic flow survey. Comparative analyses of the distinctive characteristics exhibited by the EV and ICEV in diverse traffic scenarios are depicted and illustrated in on Fig 3.4 and Fig 3.5.

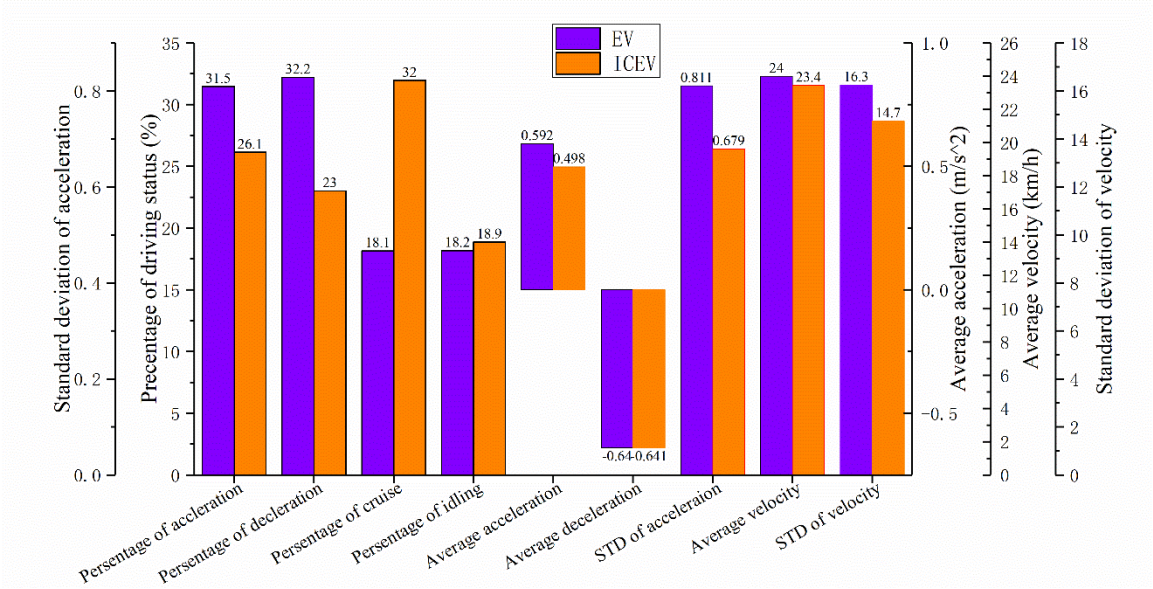

Fig 3.4 EV and ICEV driving data comparison during peak traffic

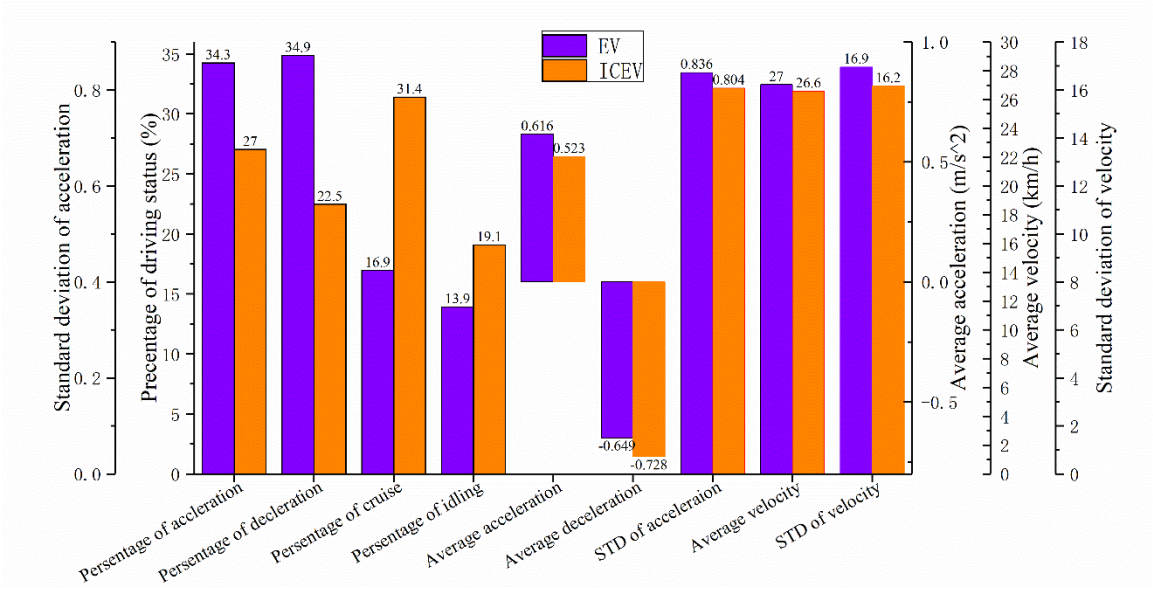

Fig 3.5 EV and ICEV driving data comparison during off-peak traffic

The comparative analyses reveal that, irrespective of the traffic status being peak or off-peak, acceleration and deceleration consistently constitute a significant proportion of the gathered data. In the peak traffic condition, the EV driving acceleration part and deceleration part are 20.69% and 40.00% larger than the acceleration and deceleration of ICEV driving data. Notably, the EV exhibits larger average acceleration by 18.87% and higher standard deviation of acceleration 17.78%, compared to the ICEV in both traffic scenarios. However, the average velocities of the EV surpass those of the ICEV to a lesser extent at 2.54% and 1.50% in both traffic conditions, as the similar traffic flow mitigates the advantage of acceleration in the EV, resulting in a marginally higher average velocity. This observation suggests that despite the similarity in average speed between EVs and ICEVs, substantial disparities in the dynamic characteristics such as acceleration and deceleration underscore the necessity for a distinct driving cycle to effectively assess the performance of EVs. Consequently, this dissertation introduces a dedicated EV driving cycle tailored to capture and evaluate the specific characteristics of EVs' dynamics.

within data processing phase, the gathering of original driving data is initially sliced into driving segments. These segments are then grouped together based on similarities using SOM and SVM classification methods. The cluster group comprises driving segments with similar pattern, such as comparable traffic scenarios. The detailed steps involved in this process are as follows:

The collected data is separated into different driving segments by the threshold value of acceleration and velocity by Eq (3.6). driving data received from the EV and ICEV are separated to 16045 and 13479 driving segments.

$$\begin{cases} \text{acceleration} & a \geq 0.15\text{m/s}^2 \\ \text{deceleration} & a \leq -0.15\text{m/s}^2 \\ \text{uniform} & v \geq 2\text{m/s} \Lambda -0.15\text{m/s}^2 < a < 0.15\text{m/s}^2 \\ \text{idling} & v < 2\text{m/s} \Lambda -0.15\text{m/s}^2 < a < 0.15\text{m/s}^2 \end{cases} \tag{3.6}$$

In the combined approach using SOM and SVM for classifying kinematic segments, the SOM results are utilized as the training set for the SVM. Typically, in previous literature, the training set is generated using algorithms like K-Means clustering. For instance, in driving cycle studies for Dublin [7], Tehran [19], Fuzhou [20], and Florence [18]. Although the K-Means is efficient in clustering data, its results are highly dependent on the choice of clustering kernel. Variations in clustering kernels can lead to significant discrepancies in classified results, resulting in instability. within this dissertation, SOM is employed to mitigate the vibration of classification results. The SVM is then updated using the clustering outcomes of SOM to enhance clustering performance and achieve a near-optimal result. This approach aims to improve the stability and accuracy of the clustering process by leveraging SOM's ability to capture underlying data structures and patterns.

SOM is an unsupervised competitive machine learning algorithm that has been applied in clustering, dimensionality reduction, and high dimensional visualization. The algorithm assumes that there are some topological structures in the input object, The process of transforming data from a high-dimensional input space to a lower-dimensional output space serves to reduce the dimensions while preserving the underlying topological structure intact. SOM consists of two neural networks, the input layer and the competitive layer. The input layer can be a vector of any dimension, which is connected with the

competitive layer fully. The competitive layer is also the output layer, which is composed of neurons with vectors of weights. These neurons have topological relationships with each other, and every neuron in the input layer is connected to each competitive neuron .

In the training process, a competitive strategy is adopted. That is, each input sample finds a neuron in the output layer that best matches its pattern, The neurons in the output layer engage in competition to determine activation, with only a single output neuron being activated. Then, the neuron is regarded as the activating neuron, also known as the winning neuron. When the status of the other neurons is deactivated, only the winning neuron has the authority to adjust the weights. The learning strategy of SOM differs from competitive learning in that not only does the winning neuron need to adjust its weight, but other neurons will also undergo weights adjustments in the winning neighborhood by the influence of the winning neuron. The training approach is to calculate the distance between the vector of weight for each neuron and each sample. The neuron with the smallest distance wins, and the winning neuron's weight vectors and its neighboring neurons will be adjusted according to the distance. This process iterates until it converges, with each winning neuron representing a clustered class.

The flow of SOM is as follows:

Let X_{n*k} denote the input sample where $X_{n*k} = (X_1, X_2, \dots, X_{k)}$, with n representing the number of samples and k representing the number of features. The competitive layer consists of m neurons, equipped with a weight vector $w_j = \left(w_{j1}, w_{j2}, \dots, w_{jk}\right), j = 1,2, \dots, m.$

1) Calculate the quantity of neurons within the competitive layer, and initialize the weight vector w_j with random values for every neuron. Define the starting neighborhood $N_j(0)$ with a broader setting that will progressively reduce as the training iterations advance, ultimately converging the neighborhood radius to zero.

2) Normalize both the sample data and the weight vectors to convert them into unit length while maintaining their original orientation. The normalization process is demonstrated as follows:

$$\hat{X}_P = \frac{X_p}{\|X_p\|}, p = 1,2, \dots, n \tag{3.7}$$

$$\hat{w}_j = \frac{w_j}{\|w_j\|}, j = 1,2, \dots, m \tag{3.8}$$

where $X_p = (x_{p1}, x_{p2}, \dots, x_{pk})$ is the p-th sample data, and w_j is the weight vector.

3) Compute the cosine similarity of the normalized input vector $\hat{X}_P$ and the normalized weight vector $\hat{w}_j$, and identify the neuron j^* with the highest cosine similarity as the winning neuron. Subsequently, determine the winning neighborhood and update the weights vectors of all neurons within that neighborhood. The updated weights can be calculated as follows:

$$w_{ji}(t+1) = w_{ji}(t) + \eta(t,s)\left[x_{pi} - w_{ji}(t)\right], i = 1,2, \dots, k, j \in N_{j^*}(t) \tag{3.9}$$

where x_{pi} is the i-th characteristic of the p-th sample data, $w_{ji}(t)$ denotes the weight vector component i of neuron j at time t, and $\eta(t,s)$ represents the learning rate, that is determined

by the Euclidean distance of neuron j and the winning neuron j^* within the winning neighborhood. The learning rate can be expressed as follows:

$$\eta(t,s) = \alpha(t)e^{-s} = \frac{e^{-s}}{t+2} \tag{3.10}$$

4) Proceed with the iteration process until reaching either the highest quantity of iterations or when the learning rate drops below the designated threshold. At this point, conclude the training process. For clustering data that may not be easily separated by straight lines, SVM clustering offers an advantage by finding clusters in higher dimensional spaces. This method involves mapping the input data, which may not be separable in the input space, into a high-dimensional kernel space using a kernel function. Through this method, the algorithm can efficiently cluster data sets featuring non-uniform cluster boundaries.

However, it is important to note that SVM is an algorithm used in supervised learning, which necessitates the selection of suitable samples for training the model. Additionally, , the choice of kernel function often falls on the Gaussian radial basis function since it can manage the complexity of the model and accurately define the high-dimensional spatial structure.

To quantitatively analyze the classification results of the driving segments, the Davies-Bouldin (DB) and Silhouette indices are utilized to assess the clustering results. The DB index calculates the average maximum similarity between a cluster and all other clusters in a dataset, and it can be computed as follows:

$$\bar{R} = \frac{1}{n}\sum_{i=1}^{n} R_i = \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j}(R_{ij}) = \frac{1}{n}\sum_{i=1}^{n} \max_{i \neq j}\left[\frac{\sigma_i + \sigma_j}{l(c_i, c_j)}\right] \tag{3.11}$$

$$\sigma_i = \sqrt[1/p]{\frac{1}{t_i}\sum_{k=1}^{t_i} |x_{ik} - c_i|^p} \tag{3.12}$$

where R_i represents the maximum similarity between the i-th cluster and all other clusters; R_{ij} denotes the similarity between the i-th cluster and the j-th cluster; σ_i signifies the dispersion degree index of all data in the i-th set, with σ_i being the standard deviation of the i-th cluster when p is equal to 2; $l(c_i, c_j)$ corresponds to the Euclidean distance between the centroids of the i-th and j-th clusters; t_i indicates the volume of the i-th cluster; x_{ik} represents the k-th element in the i-th cluster; and c_i denotes the centroid of the i-th cluster.

The Silhouette index is employed to assess the clustering effectiveness of various cluster kernels within the same clustering approch and to gauge the effectiveness of various classification techniques. The Silhouette index value is calculated using Eq. (3.13), with its value falling within the range of -1 to 1."

$$S(i) = \frac{b(i) - a(i)}{max[a(i), b(i)]} = \begin{cases} 1 - \dfrac{a(i)}{b(i)}, a(i) < b(i) \\ 0 \quad , \ a(i) = b(i) \\ \dfrac{b(i)}{a(i)} - 1, a(i) > b(i) \end{cases} \tag{3.13}$$

In one cluster, $a(i)$ signifies the mean distance between the i-th entity and all other entities within the same cluster. Conversely, $b(i)$ represents the shortest distance between the i-th

entity in a chosen group and all entities in the remaining clusters. These parameters are crucial in calculating the DB index and Silhouette index, which are further detailed in Table 3.4.

Table 3.4 Classification index of different method

Index	SOM & SVM	K-Means & SVM
DB index	0.9578	1.1077
Silhouette index	0.5045	0.4529

The information conveyed by Eq. (3.11) suggests that a reduced R_{ij} value results in less pronounced similarity between two clusters. This implies that while homogeneity within one cluster is not as evident, there is a considerable disparity between the two clusters. The DB index decreases and achieves a more accurate classification result, when there is a large inter-cluster distance and a small intra-cluster diameter. When examining Table 3.4, one can observe that the DB indexes for SOM & SVM under various traffic scenarios and vehicle categories are lower than those for K-means & SVM. In a contrasting manner, a larger $S(i)$ value suggests that the i-th element aligns well with the existing cluster, but is inconsistent with the other clusters, indicating a more rational classification. If the $S(i)$ value drops below 0, the corresponding element is not suitably clustered and should be relocated to a different cluster. Upon comparing the Silhouette indexes in Table 3.4, it is seen that those corresponding to SOM & SVM are higher than those related to K-means & SVM. And the indexes show the proposed method achieved better classification results.

3.4 Driving cycle construction

By applying cluster analysis, the driving segments are classified into six distinctive categories, corresponding to the six states in the Markov model. Utilizing the available test data and referring to Eq. (3.14), the transition probabilities between these states are obtained. This constructs the matrix of Markov transition probability, as represented by Eq. (3.15).

$$P_{ij} = \frac{N_{ij}}{\sum_{j=1}^{l} N_{ij}} \tag{3.14}$$

where N_{ij} denotes the count of state transitions from state i to state j, while $\sum_{j=1}^{l} N_{ij}$ refers to the count of transitions from state i to all possible states. Additionally, l represents the total quantity of states.

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & p_{56} \\ p_{61} & p_{62} & p_{63} & p_{64} & p_{65} & p_{66} \end{bmatrix} = \begin{bmatrix} 0.6574 & 0 & 0.0713 & 0.2557 & 0.0111 & 0.0045 \\ 0 & 0.6984 & 0.0049 & 0.0034 & 0.1599 & 0.1334 \\ 0.0533 & 0.2739 & 0.0217 & 0.0231 & 0.4147 & 0.2133 \\ 0.2485 & 0.0681 & 0.0171 & 0.0524 & 0.1921 & 0.4218 \\ 0.3203 & 0.0312 & 0.2090 & 0.2873 & 0.0912 & 0.0610 \\ 0.0665 & 0.1168 & 0.0254 & 0.2189 & 0.0907 & 0.4817 \end{bmatrix} \tag{3.15}$$

Among existing research, the construction of driving cycles strictly adhered to the selection of driving segments utilizing the probabilistic transition matrix derived from the Markov model [33], [69], [70], [71]. Nonetheless, the approach that relies on maximum likelihood estimation to categorize driving segments tends to disregard the occurrence of secondary maximum probability events or low probability events. The limitations of this

method become more evident when the second highest probability is very close in magnitude to the maximum probability. Adhering to the Markov property, for any given current state, the sum of transition probabilities from that state to all possible next states, including a transition to itself, must equate to 1. In equation form, this is represented as $\sum_i^j P_{ij} = 1$. To improve the representation of low-probability occurrences, a series of independent trials can be conducted, each generating a uniformly distributed random value between 0 and 1. The resulting distribution corresponds to the probabilistic transition behavior captured by the Markov model's transition matrix. The succeeding event of importance can be selected in reference to this random number. This concept aligns with the Monte Carlo simulation methodology, a random simulation technique that models real-world physical processes by examining geometrical quantities and characteristics of movements.

When the present state is denoted as i, the probabilities of transitioning to other states are respectively represented as $P_{i1}, P_{i2}, P_{i3}, P_{i4}, P_{i5}, P_{i6}$. Following this, the range of [0, 1] is subdivided in accordance with the magnitude of the likelihood for transition between states, meaning the segment length equates to the likelihood for transition between states, as depicted in Fig 3.6. As per a sequence of independent random tests, a random number r falls into an interval determined by Eq. (3.16), which governs the consequent driving segment.

$$\sum_{j=0}^{k-1} P_{ij} < \omega < \sum_{j=0}^{k} P_{ij}\,, k = 1,2,\dots,6 \tag{3.16}$$

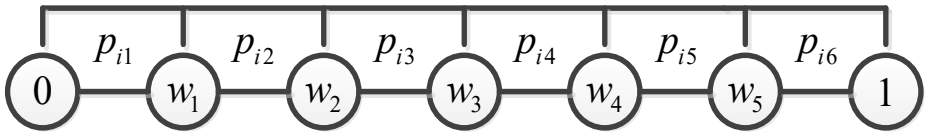

Fig 3.6 State segmentation

The approach generates potential driving cycle profiles by piecing together the driving segments synthesized through the Markov and Monte Carlo simulation techniques. However, the overall length of these constructed driving cycles remained undetermined. An important consideration is that the duration of a driving cycle should avoid being excessively brief, as that would fail to adequately capture the diversity of driving conditions. Conversely, an overly extended duration could lead to redundancy and an inefficient testing process, as this could compromise the representation of real-world driving conditions, nor be excessively long, which could render it inapplicable in actual tests. Most driving cycles fall within a range of 600 to 1800 seconds, while a reasonable duration cited for urban driving cycles in other studies is approximately 1200 seconds [8], [33]. Taking these factors into account, the researchers selected a duration of 1200 seconds as the target length for the synthesized driving cycle in this dissertation. To initiate the cycle construction process, an idling segment with a maximum length of 5 seconds was designated as the starting state for the cycle profile. Following this, Markov chain and Monte Carlo simulation method were utilized the to determine the appropriate category for the subsequent driving segment. Once the category was ascertained, a driving segment from the determined category was chosen without repetition. This successive selection process continued, with each new driving segment being absorbed into the current cycle, until meet the predetermined length

of 1200 seconds . The process of choosing driving segments ought to abide by three central principles: The primary objective is to minimize the separation between each driving segment and the centroid of the cluster it belongs to. Secondly, the disparity of the starting velocity of the succeeding driving segment and the concluding velocity of the selected segment should not exceed 1 km per hour. Lastly, in cases where multiple driving segments fulfill the first two principles, preference should be given to the segment that is in closest proximity to the cluster centroid. The proposed EV driving cycle is represented in Fig 3.7. The Speed-Acceleration Probability Distribution (SAPD) of both the EV driving cycle and the authentic real-world traffic date can be observed in Fig 3.8 and Fig 3.9. The findings indicate that the EV driving cycle predominantly includes low-speed profiles, characterized by pronounced fluctuations. The frequency and intensity of speed changes are largely influenced by road conditions and traffic flow. Table 3.5 indicates the evaluation parameters of the EV driving cycle and the real-world driving data. A noticeably lower Relative Error (RE) and minimal Root Mean Square Error (RMSE) is indicative of a high degree of similarity between the formulated EV driving cycle and the real-world conditions. Thus, the methodology proposed in this dissertation enabled the creation of an EV driving cycle that manifests speed-acceleration profiles and the evolution of vehicle speeds over time effectively.

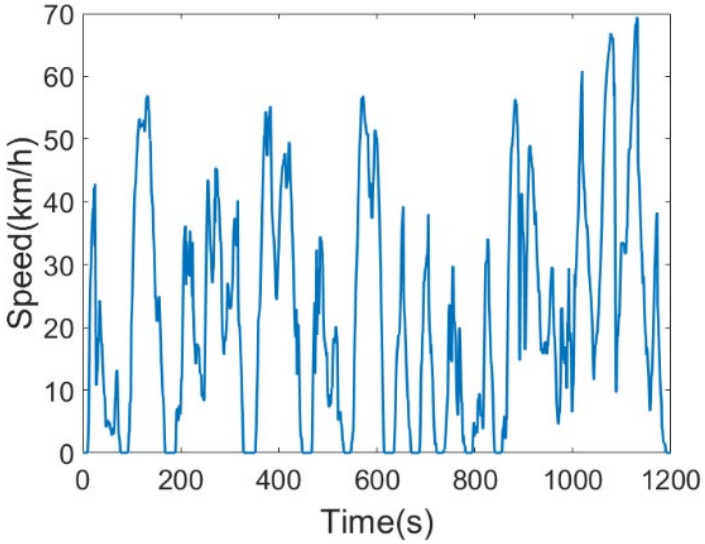

Fig 3.7 EV driving cycle

Table 3.5 Comparison of assessment parameters between the Xi'an urban driving cycle and the real-world driving cycle

Assessment parameter	Real-world	Driving cycle	RE	RMSE
Average speed (km/h)	20.00	21.18	5.90%	1.4%
Standard deviation of speed (km/h)	16.48	18.09	9.77%	
Average positive acceleration (m/s^2)	0.64	0.68	6.25%	
Average deceleration (m/s^2)	0.64	0.69	7.81%	
Standard deviation of acceleration (m/s^2)	0.87	0.88	1.15%	
Percentage of acceleration (%)	34.5	34.3	0.58%	
Percentage of deceleration (%)	32.0	33.3	4.06%	

Percentage of constant speed (%)	15.9	15.6	1.89%
Percentage of idling (%)	17.6	16.8	4.55%

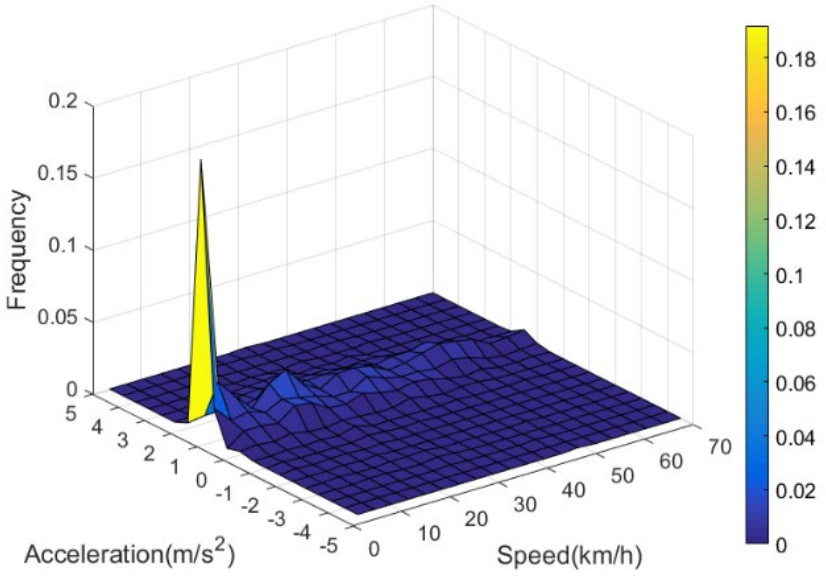

Fig 3.8 SAPD of the Xi'an urban driving cycle

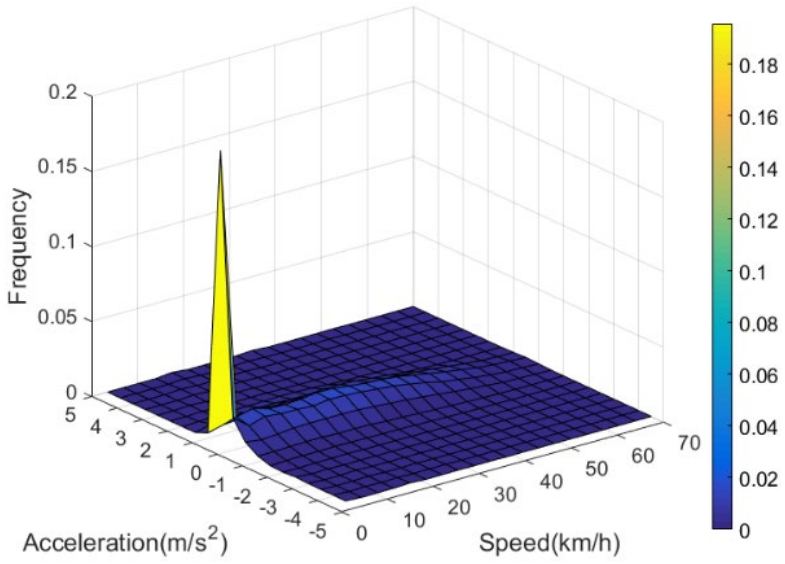

Fig 3.9 SAPD of the real-world driving data

To validate the accuracy and representative nature of the synthesized EV driving cycle, comparative evaluations were performed by analyzing the driving range under this cycle in both simulated conditions and real-world driving environments. For the real-world testing, seven driving range trials were executed. Each trial commenced with the battery at a 100% SOC and concluded once the SOC diminished to approximately 20%.

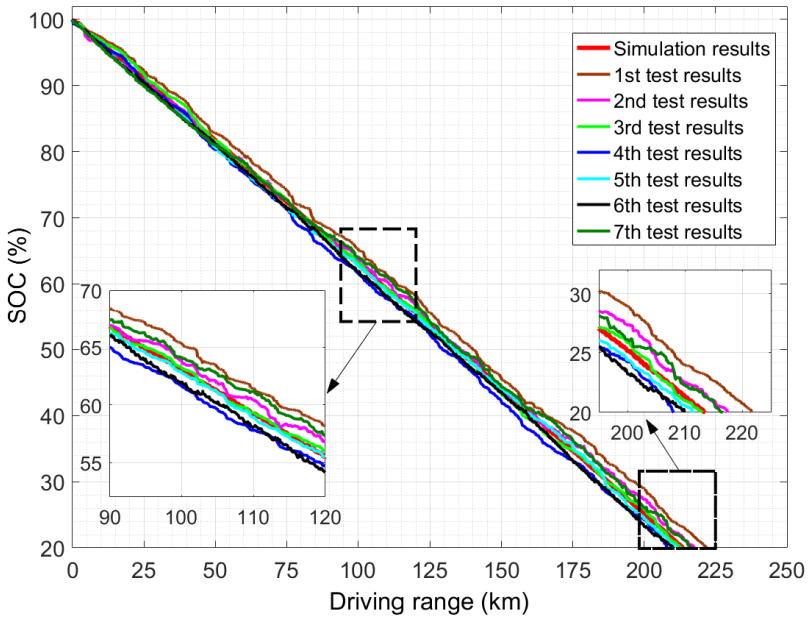

Fig 3.10 Driving range evaluation

Table 3.6 Comparison results of driving range

Test type	Driving range (km)	Relative error (%)
Simulation results	213.4	-
1st test results	221.7	3.89
2nd test results	217.4	1.87

3rd test results	213.1	-0.14
4th test results	208.2	-2.44
5th test results	211.5	-0.89
6th test results	210.3	-1.45
7th test results	216.6	1.50

The range and SOC were logged using OBD. The comparative findings between the simulation and the tests are displayed in Fig 3.10 and Table 3.6. These findings reveal that the driving range spans 208.2 to 221.7 km during the seven real-world road tests. Under the simulation of the developed EV driving cycle, the driving range stands around 213.4 km. By analyzing the comparative results, it indicates that the discrepancy in the distance traveled on a single charge when comparing the constructed driving cycle profile against the real-world vehicle testing data lies within the range of -2.44% to 3.89%. This close alignment demonstrates that the synthesized EV driving cycle successfully captures the realistic driving patterns experienced by EV in actual operating conditions.

3.5 Conclusion

In this chapter, a dedicated urban driving cycle is developed to evaluate EV performance and energy management systems effectively. The approach began by selecting diverse urban routes and collecting comprehensive operational data. Employing principal component analysis, I reduced data complexity while retaining essential characteristics, which was refined with a hybrid SOM-SVM classification to categorize driving conditions. I then synthesized the driving cycle using Markov chains and Monte

Carlo simulations to replicate the stochastic nature of urban driving, confirmed by meticulous validation through statistical accuracy metrics. The final driving cycle bridges the gap between theoretical EV research and practical energy management, offering a novel methodology and a validated tool for optimizing EV efficiency and advancing the field.

CHAPTER 4

Q-LEARNING BASED EMS FOR ELECTRIC VEHICLES

In this chapter, a Q-learning based EMS is devised based on developed EV driving cycle to achieve better energy efficiency and alleviate the battery degradation. Also, a new Lithium-Sulfur battery with bilateral solid electrolyte interphase is studied and implemented to lower the EV operating cost.

4.1 Research gaps and proposed methods

For EVs equipped with HESS, numerous EMSs have been developed, encompassing rule-based approach [72], global optimization strategy [73], instantaneous optimization method [74], and artificial intelligence approach [75].]. Typically, the rule-based method, which is built on predefined threshold values and can be enhanced by incorporating fuzzy logic principles or leveraging optimization techniques [76]. Dynamic programming (DP), a prominent optimization approach, has seen widespread adoption in the design of EMS. Its solutions provide a valuable reference point for assessing the performance of alternative methodologies [77]. DP distinguishes itself as a globally optimal supervisory control strategy, particularly renowned for its capability to deliver unparalleled energy efficiency over a predetermined driving cycle profile. Nonetheless, the computing capability requirement makes it incredibly difficult to implement in real-time [63, 71]. Conventional method a few drawbacks when compared to RL-based EMS. Firstly, conventional EMS approaches rely on pre-defined control algorithms that are designed based on specific driving conditions or scenarios. This lack of adaptability makes them less efficient in

dynamically changing driving conditions, such as varying traffic patterns or road gradients. And conventional EMS approaches often aim to achieve a fixed objective, such as maximizing fuel efficiency or minimizing energy consumption. However, they may not consider the real-time variation of factors like traffic conditions and driving behavior. As a result, they may not achieve optimal performance consistently. Besides, EVs dynamics and energy consumption patterns are inherently complex and nonlinear. Conventional EMS methods often rely on simplified models that might not capture the intricacies of these systems accurately. As a result, their performance may be suboptimal in real-world scenarios. Also, Conventional EMS approaches usually do not leverage historical data effectively. These strategies overlook the impact of the driver's behavior patterns and do not take into account the insights garnered from prior experiences. Consequently, these approaches may fail to make informed decisions and miss potential energy-saving opportunities.

The proposed Q-learning based EMS overcomes these drawbacks by learning from interactions with the environment and dynamically adapt the control strategies. Thus, the Q-learning based EMS can optimize the performance under various driving conditions, making it more adaptable compared to conventional approaches. Also, the proposed Q-learning based EMS can make decisions in real-time, considering the current state of the vehicle and the environment. It considers parameters like traffic conditions, road gradients, and driver behavior to optimize energy usage in a dynamic manner. Another advantage of the proposed method is that does not rely on specific system models. It can learn directly from data generated by the vehicle or obtained from historical records. This enables it to

capture the complexities of the electric vehicle system, leading to improved performance. Furthermore, the Q-learning based EMS can leverage historical data to learn from past experiences. By incorporating driver behavior patterns and knowledge gained from previous driving sessions, they can make more informed decisions, leading to enhanced energy efficiency. Overall, the proposed method offers a more adaptable, real-time, and data-driven approach to optimize energy management problems in electric vehicles, overcoming several limitations of conventional strategies.

4.2 Modeling of electric vehicle

4.2.1 dynamic model of vehicle

As the vehicle is on driving status, the formula representing the balance of the force is presented as follows:

$$F_{trac} = F_{inertia} + F_{roll} + F_{aero} + F_{grade} \tag{4.1}$$

where F_{grade} denotes gradient resistance, F_{aero} represents air resistance, F_{trac} means traction force, F_{roll} stands for rolling resistance, and $F_{inertia}$ refers to the inertia force.

The traction force is obtained as follows:

$$F_{trac} = F_m - F_{brake} \tag{4.2}$$

where F_m is the propulsive force generated by the electric motor, whereas F_{brake} refers to the deceleration force, composed of both mechanical braking force and regenerative braking force.

The drag force caused by rolling resistance is computed by making use of the rolling resistance coefficient $f_{roll}(V_{veh}, P_{tire}, \dots)$, the mass of vehicle M_{veh}, and road grade δ.

$$F_{roll} = f_{roll}(V_{veh}, P_{tire}, \dots) M_{veh} g \cos\delta \tag{4.3}$$

where the rolling resistance coefficient is influenced by several variables, including velocity, tire pressure, and temperature. To simplify the calculation, f_{roll} is considered to be a constant.

The aerodynamic resistance is determined through a function that includes variables such as the air drag coefficient (C_d), vehicle frontal area (A_f), air density (ρ_a), and velocity (V_{veh}), as shown in Eq. (4.4)

$$F_{aero} = \frac{1}{2} \rho_a A_f C_d V_{veh}^2 \tag{4.4}$$

The gradient force is computed using a formula that incorporates the pathway grade (δ) and vehicle mass. This calculation is represented in Eq. (4.5).

$$F_{grade} = M_{veh} g \sin\delta \tag{4.5}$$

The EM power is derived from the vehicle dynamics, is calculated as follows:

$$T_{mg} = \frac{F_{trac} d_w}{2 i_0} \tag{4.6}$$

$$\omega_{mg} = \frac{2 V_{veh}}{d_w} i_0 \tag{4.7}$$

$$\eta_{mg} = \mathrm{f}(T_{mg}, \omega_{mg})$$

(4.8)

$$P_{mg} = T_{mg}\omega_{mg}\eta_{mg}Sign(T_{mg}\omega_{mg})$$

(4.9)

$$P_{mg} = P_{cap} + P_{bat}$$

(4.10)

where ω_{mg} stands for the EM rotating speed, η_{mg} represents the EM efficiency is obtained by referencing a predefined lookup table, which maps the motor's torque and rotational speed values to the corresponding efficiency figures. T_{mg} represents the EM output torque, d_w is the wheel diameter, i_0 refers to the final reducer's gear ratio, and P_{cap} and P_{bat} refer to the power to the EM from the SC and LIB.

Table 4.1 illustrates the vehicle parameters adopted in the dissertation.

Table 4.1 vehicle model parameters

Parameters	Value
Curb weight (kg)	1778
Max weight (kg)	2180
Windward area (m^2)	2.34
Air drag coefficient	0.30
Wheelbase (mm)	2870
Wheel diameter(mm)	693.7
Top speed (km/h)	200
0-100 km/h time (s)	8

Parameters	Value
Grade ability (%)	30

4.2.2 Propulsion system model

There are several existing configurations for the LIB and SC HESS. The passive parallel HESS configuration is shown in Fig 4.1, SC and LIB are connected in parallel. This means that both energy storage components share the same voltage level and are connected directly without the need for complex power electronic modules.

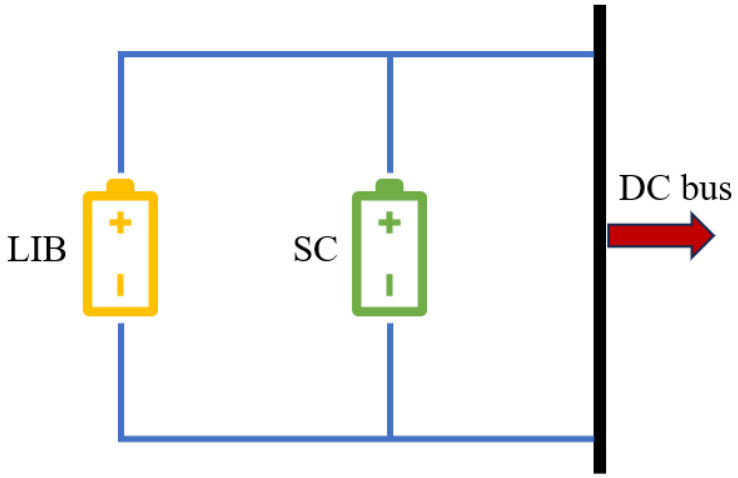

Fig 4.1 Passive parallel HESS configuration

The passive parallel configuration offers several advantages. Firstly, it is the simplest topology, making it easy to implement and maintain. All voltages across the system are equal, simplifying the control and management of the energy storage components. Additionally, supercapacitors in this setup can function as low-pass filters, enhancing the system's efficiency by smoothing out voltage fluctuations. The configuration also boasts high reliability and lower costs due to the absence of power electronic modules, reducing

the chances of component failure and minimizing expenses. Moreover, the absence of a system optimization stage reduces computational complexity, making it easier to design and operate. However, this configuration comes with its set of challenges. One major drawback is the uncontrolled power distribution between LIB and SC. This lack of control can lead to inefficient use of both energy storage components. During cruising and braking, there are significant variations in discharging and charging currents, which can affect the overall performance and efficiency of the system. Moreover, the passive parallel setup has limited utilization of SC, which means that the system cannot achieve an optimum solution for EVs. The inability to fully exploit the potential of SC in this configuration can limit the EV's ability to efficiently handle regenerative braking and rapid acceleration scenarios, impacting its overall energy efficiency and performance.

The SC semi-active HESS configuration is illustrated in Fig 4.2, SC are connected in parallel with LIB, and a DC/DC converter is used to control the energy flow between them, allowing for more flexible and controlled operation.

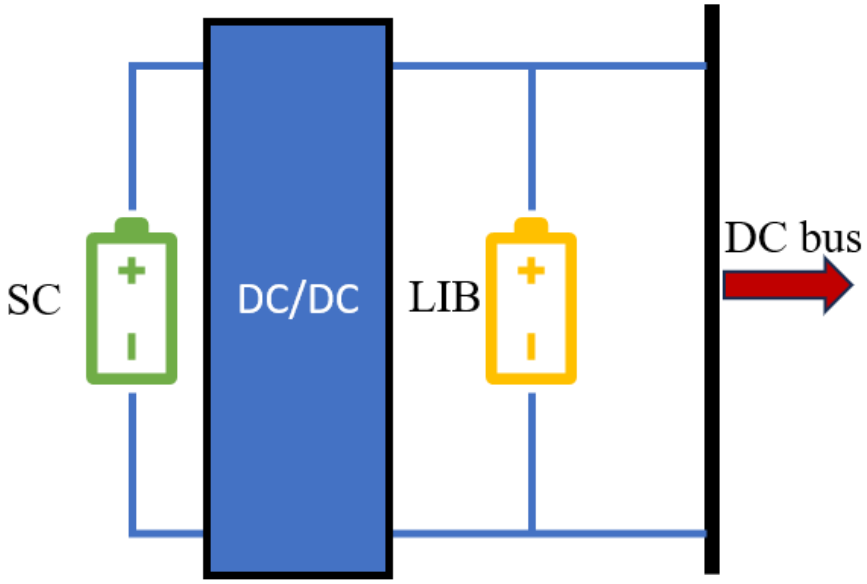

Fig 4.2 SC semi-active HESS configuration

One of the significant advantages of the SC semi-active HESS configuration is its flexible control over supercapacitors. This flexibility enables efficient management of the energy flow, allowing for optimal utilization of both SC and LIB. Moreover, this configuration is suitable for a wide voltage range, accommodating various applications with diverse voltage requirements. The direct connection of LIB to the DC link results in low voltage variation across the link, ensuring stable and consistent power supply. Additionally, the SC semi-active configuration is compact in size and utilizes a low-cost converter, making it an economical choice for energy storage systems. However, there are limitations to this configuration. One drawback is its inability to save maximum regenerative energy in SC. Due to the constraints of the power electronic converter, the system may not capture and store all the energy generated during regenerative braking or other high-energy events. Additionally, the converter in this configuration needs to be rated according to the peak

power rating of SC, which can lead to over-dimensioning and higher costs. Furthermore, while SC are effectively utilized, the same cannot be said for LIB. The system might not fully exploit the potential of LIB due to the limitations imposed by the control strategy and converter design, leading to suboptimal use of the available energy storage capacity.

Fig 4.3 presents the battery semi-active HESS configuration, the LIB is the primary energy storage source, supplemented by SC connected in parallel through a power electronic converter, allowing for controlled energy flow between the two components.

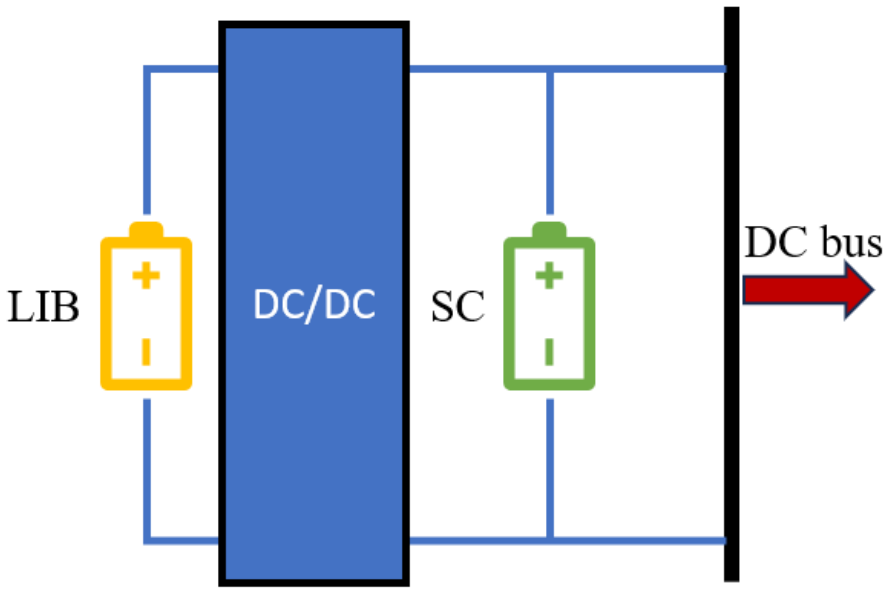

Fig 4.3 Battery Semi-Active HESS Configuration

One of the key advantages of the battery semi-active HESS configuration is that the converter design is based on the average load power, optimizing the system for typical operating conditions and improving overall efficiency. Cell balancing, a crucial aspect of battery management, is required in this configuration, ensuring that all cells within the

battery pack have similar state of charge, thus extending the battery's lifespan. Additionally, this setup results in low current fluctuation across the DC link, enhancing the LIB's life cycle and overall reliability. The presence of SC directly connected to the DC link allows for fast peak power control, enabling rapid energy discharge and absorption during high-demand scenarios. However, there are certain limitations associated with this configuration. High voltage variation across the SC can lead to significant discharge leakage, reducing the overall efficiency of the system. Voltage balancing issues within the individual SC can also arise, requiring additional control mechanisms to ensure uniform charging and discharging, which can add complexity to the system design. Moreover, due to the direct connection of SC with the DC link, there is limited utilization of the SC's potential. The system might not fully exploit the rapid energy absorption capabilities of SC due to constraints imposed by the control strategy and converter design, leading to underutilization of this high-power density energy storage component.

The hybrid diode semi-active HESS configuration is shown in Fig 4.4, SC and LIB are connected in parallel using diodes and a power electronic converter, allowing for controlled energy transfer between the components.

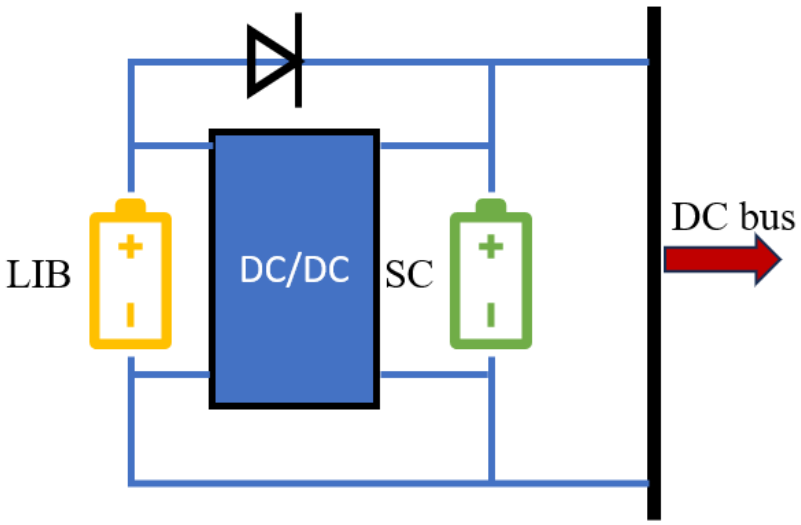

Fig 4.4 Hybrid diode semi-active HESS Configuration

This configuration offers several advantages. One notable benefit is that the voltage of SC is higher than that of lithium-ion batteries, allowing for efficient utilization of SC. This means that supercapacitors can handle high-power bursts and rapid charge/discharge cycles effectively, enhancing the overall energy efficiency of the system. Additionally, LIB load curves tend to be gentler in this configuration, leading to smoother and more stable power delivery. The hybrid diode semi-active configuration also boasts low operating costs, primarily due to the simplicity of its control algorithms, making it an economical choice. Furthermore, this setup results in reduced size and weight of the overall system while maintaining high efficiency, making it suitable for applications where space and weight constraints are critical factors. However, there are challenges associated with this configuration. Reverse current fluctuations may occur from the inverter side, impacting the stability of the system and potentially causing inefficiencies in energy transfer. The high operating voltage, despite delivering the same power, can reduce the current rating of the

components, which may affect the overall performance of the system. Proper sizing of components and their control mechanisms remain a significant challenge in this configuration. Achieving the right balance between component sizes, control strategies, and system requirements is crucial to ensure optimal performance and efficiency, making the proper sizing and control of components a primary concern in the hybrid diode semi-active HESS configuration.

The series fully active HESS configuration is presented in Fig 4.5, SC and LIB are connected in series, and both sources are actively controlled using dedicated power electronic converters. This setup allows for precise control of the energy flow between the components.

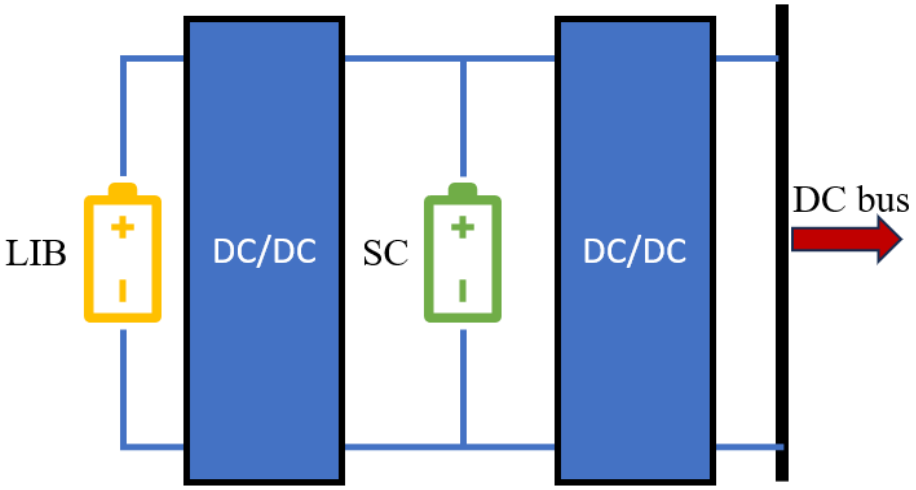

Fig 4.5 Series fully active HESS configuration

One of the main advantages of the series fully active configuration is that both LIB and SC are fully decoupled, leading to improved overall efficiency of the converter. Each source can follow its optimal current and voltage-controlled strategy, maximizing the utilization of

both LIB and SC. This configuration offers more flexibility in controlling both sources, enabling sophisticated energy management algorithms. Additionally, the DC bus voltage is regulated using a voltage-controlled strategy, ensuring stable and consistent power delivery to the load. However, there are several disadvantages associated with this configuration. The system is bulkier, resulting in higher costs due to the additional components and complexity involved in having two power electronic converters in the circuit. High power losses occur due to the two conversion stages up to the DC link, reducing the overall efficiency of the system. The controlling process is more complex compared to other topologies, requiring advanced algorithms and precise synchronization between the converters. Moreover, stability problems can arise across a wide operating voltage range, demanding sophisticated control techniques to maintain system stability under various operating conditions. In summary, the series fully active configuration offers enhanced control and flexibility in managing both lithium-ion batteries and supercapacitors. However, these advantages come at the cost of increased complexity, higher power losses, and potential stability challenges, making it essential to carefully consider the trade-offs before implementing this topology in practical applications.

The parallel fully active HESS configuration is illustrated in Fig 4.6, SC and LIB are connected in parallel, each with its own dedicated power electronic converter, enabling independent and precise control of energy flow from both sources.

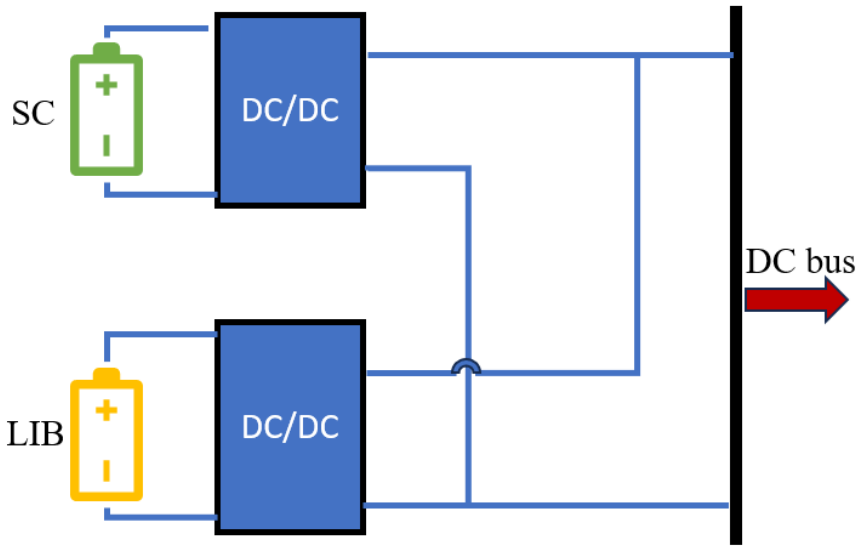

Fig 4.6 Parallel fully active HESS configuration

One of the primary advantages of the parallel fully active HESS configuration is the flexibility it offers in controlling both LIB and SC independently. This independent control allows for optimal utilization of each source, maximizing the efficiency and overall performance of the energy storage system. Additionally, this setup leads to low cell balancing issues for both LIB and SC, ensuring uniform charging and discharging of individual cells within the energy storage components. The DC voltage can be regulated easily across the DC link, providing stability to the system. The stable voltage across the DC link contributes to the system's high performance and efficiency. This configuration is widely adopted in smart grid systems due to the ability to independently control each source, enabling efficient energy management and grid stabilization. However, there are several disadvantages associated with the parallel fully active configuration. The system can be very expensive due to the requirement of two converters, one for each energy source, leading to higher costs in comparison to other topologies. The design and control of the parallel

configuration are more complex compared to cascade models, necessitating advanced control algorithms and precise synchronization between the converters. Moreover, to avoid any interference between LIB and SC, more protection circuits are required, adding to the complexity and cost of the system. In summary, the parallel fully active configuration provides excellent flexibility and control over both LIB and SC, making it suitable for applications where precise energy management and high efficiency are crucial. However, the complexity and cost associated with this configuration require careful consideration and evaluation of the specific requirements of the application before implementation.

The multiple-inputs fully active HESS configuration in Fig 4.7, multiple energy sources, such as LIB and supercapacitors SC, are connected to the system through individual power electronic converters, allowing simultaneous and independent control of each source.

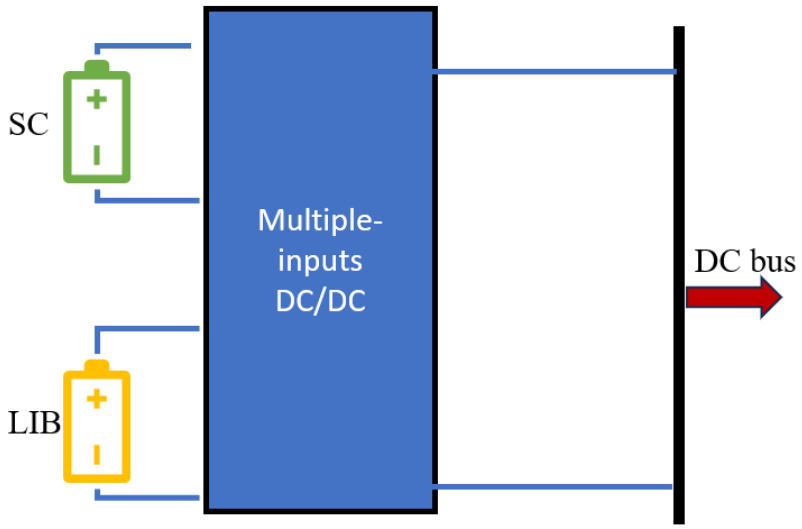

Fig 4.7 Multiple-inputs fully active HESS configuration

One of the key advantages of the multiple-inputs fully active configuration is the ability to control both LIB and SC simultaneously. This simultaneous control enables efficient energy management and maximizes the utilization of multiple energy sources. The configuration allows for multiple inputs to be coupled electrically, magnetically, or electromagnetically, providing flexibility in the design and integration of various energy storage components. Through soft switching techniques, the converter's efficiency can be improved, minimizing power losses during energy conversion. Additionally, this configuration results in a more compact size and lightweight system compared to other fully active topologies, making it suitable for applications with space and weight constraints. However, there are several disadvantages associated with the multiple-inputs fully active configuration. The structure design of this configuration is highly complex, requiring advanced engineering expertise and precise synchronization of multiple energy sources and converters. The use of a large number of switches in the system leads to high switching power losses, reducing the overall efficiency of the energy storage system. The employment of a superfast control system is necessary to manage the complex operation of multiple inputs, adding to the system's cost and complexity. Moreover, due to the presence of multiple energy sources, more protection circuits are required to avoid any interference between LIB and SC, increasing the complexity and cost of the overall system. In general, the multiple-inputs fully active configuration offers the advantage of simultaneous control of multiple energy sources, providing efficient energy management and utilization. However, the complexity of the system design, high switching power losses, cost implications, and the need for extensive protection circuits should be carefully considered

and addressed to ensure the successful implementation of this advanced energy storage topology.

Fig 4.8 presents the hybrid propulsion system diagram.

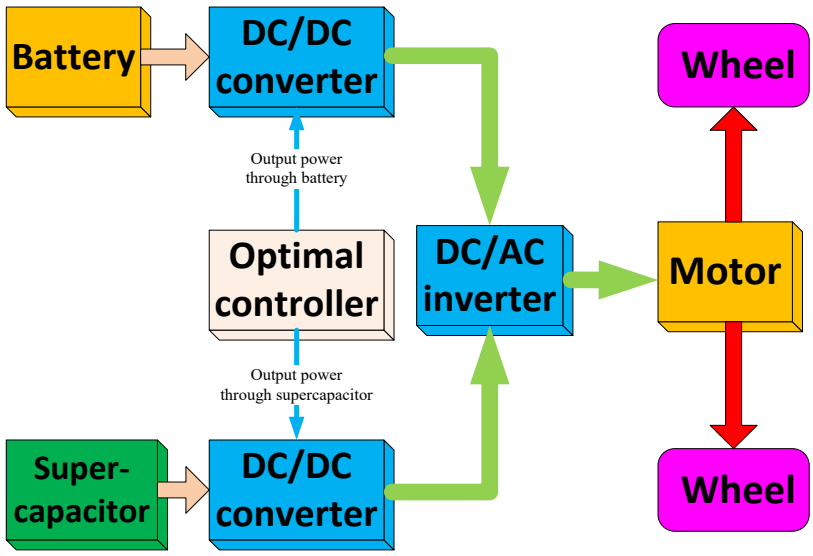

Fig 4.8 The diagram of propulsion system.

The vehicle is propelled by a parallel active HESS comprising a LIB and a SC. The pertinent parameters are listed in Table 4.2. The system powers the rear axle through an EM that is connected to a fixed gear reduction with a ratio of 7.39. During operation, the controller receives power demands from the driver and commands the EM to generate the requisite torque to propel the vehicle. An optimal control strategy determines the power split between the LIB and SC to meet the EM's power needs. The system incorporates a DC/DC converter to interface the SC with the load bus, mitigating potential issues from SC voltage fluctuations. Another DC/DC converter links the LIB pack to the load bus, enabling it to

supply the required current. A DC/AC inverter with 92% efficiency (as per [49]) connects the EM to the power bus.

Table 4.2 Power supply parameters

Parameters	Value
Battery module number	30
Battery pack voltage (V)	342.18~389.70
Battery pack capacity (Ah)	300
SC number	50
SC series connection	50
SC capacitance (F)	200
DC/AC efficiency (%)	92

4.2.3 Model of battery

Fig 4.9 illustrates the modeling approach adopted for the battery, which employs a simplified zero-order equivalent circuit representation. This circuit is composed of an ideal voltage source connected in series with a resistor element.

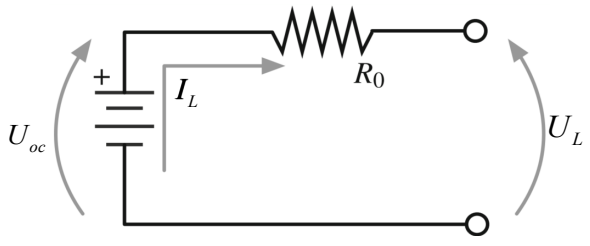

Fig 4.9 The battery internal resistance model.

The voltage and current can be obtained using the followings:

$$U_L = U_{oc} - I_L R_{bat} \tag{4.11}$$

$$I_L = \frac{U_{oc} - \sqrt{U_{oc}^2 - 4R_{bat}P_{bat}}}{2R_{bat}} \tag{4.12}$$

where U_L denotes the load voltage, I_L stands for the current, U_{oc} refers to the open-circuit voltage (OCV), and P_{bat} represents the output power of the battery.

Before initiating the battery charging and discharging tests, the battery was placed inside a temperature chamber held at 25°C for two hours. This procedure aimed to establish consistent internal and external temperatures, mitigating the impact of thermal gradients during subsequent testing. The battery remained within the temperature chamber throughout the charging and discharging cycles. An Arbin test system was employed to perform the charge and discharge procedures, while simultaneously acquiring data at one-second intervals. Fig 4.10 presents the recorded OCV and internal resistance of the battery under test.

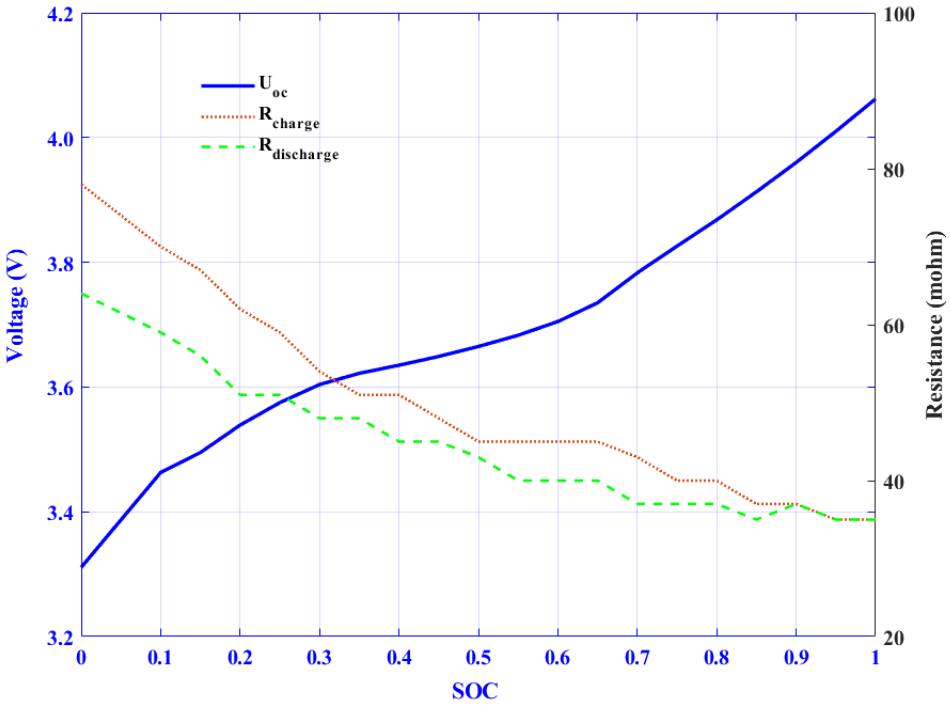

Fig 4.10 Battery OCV and resistance test data

The ampere-hour integral approach offers low implementation costs, high computing efficiency, and straightforward hardware application. Consequently, this dissertation employs the ampere-hour integration approach for calculating the SOC of the battery, as depicted in Eq (4.13)

$$SOC(t) = SOC_0 - \int_0^t I_L(t)dt/Q_{bat} \tag{4.13}$$

where SOC_0 is the SOC at initial point, and Q_{bat} means denotes the battery rated capacity

Throughout the degradation process of LIB, several factors including the ampere-hour throughput, SOC, temperature, and current play significant roles in influencing the aging of

LIBs. Based on reference [30], the LIB aging process can be represented by a semi-empirical model as given below:

$$Ah(n)=DOD(n)*Ah \tag{4.14}$$

$$E_a(n)=31500-152.5*C\text{-}rate \tag{4.15}$$

$$B=\alpha*SOC+\beta \tag{4.16}$$

$$Q_{C\text{-}rate}(n)=B(n)e^{\frac{-E_a(n)}{RT}}Ah^z(n) \tag{4.17}$$

where Ah represents the ampere-hour throughput, n denotes the n-th C-rate being considered, $Ah(n)$ corresponds to the Ah throughput related to the n-th C-rate, and $Q_{C\text{-}rate}(n)$ indicates the battery capacity fade incurred due to the n-th C-rate,. The variables α and β are defined corresponding to SOC values as follows:

$$\begin{cases} \alpha=2896.6,\ \beta=7411.2\ ,\ if\ \ SOC<0.45 \\ \alpha=2694.5,\ \beta=6022.2,\ if\ \ SOC\geq 0.45 \end{cases} \tag{4.18}$$

The cumulated degradation of the LIB is obtained using Eq (4.19).

$$Q_{cycle}=\sum_{1}^{n} Q_{C\text{-}rate}(n) \tag{4.19}$$

Over extended periods of use, the battery cycle repeats daily. Within the i-th cycle of LIB operation, the ampere-hour throughputs of different C-rates are accumulated from the first cycle to the i-th cycle, and the LIB capacity loss is computed based on the total ampere-

hour throughput of different C-rates. The final capacity at the conclusion of each battery cycle serves as the capacity at the initial stage of the subsequent cycle.

4.2.4 A Novel Lithium-Sulfur battery for electric vehicles

Despite the rapid advancement of EV and PHEV technologies ICEVs continue to hold the majority market share a circumstance largely attributable to the elevated costs associated with battery technology. In fact, LIBs are the only viable choice for these EVs currently, which contributes considerably to the increased costs of vehicle purchases. LIBs are associated not only with a costly initial investment, but also a significant expenditure when replacement becomes necessary due to severe degradation. This results in the value of an EV depreciating at a much quicker rate than that of a ICEV. This concern over high LIB costs has interest in alternative battery technologies in both academic and industry. One such alternative, Li-S batteries, capitalizes on the inexpensive availability of sulfur and delivers superiorly high energy density, nearly six times greater than LIBs [79]. The US government also takes a strong interest in Li-S batteries. Several notable projects have been funded by the U.S. Department of Energy to make extensive research on Li-S batteries. These include: 1. Novel Chemistry: Lithium-Selenium and Selenium-Sulfur Couple by Argonne National Laboratory, 2. Development of High Energy Li-S Batteries by Pacific Northwest National Laboratory, 3. Nanostructured Design of Sulfur Cathodes for High Energy Li-S Batteries by Stanford Acceleration Laboratory, 4. Mechanistic Investigation for the Rechargeable Li-S Batteries by University of Wisconsin-Milwaukee, 5. New Electrolytes for Li-S Battery by Lawrence Berkeley National Laboratory, 6. Multifunctional, Self-Healing Polyelectrolyte Gels for Long-Cycle-Life, High-Capacity

Sulfur Cathodes in Li-S Batteries by University of Washington [80]. The limited power density and rapid capacity fade of current commercial Li-S batteries compensate their commercialization in EVs [81]. Industrially developed Li-S cells with a 400 mAh capacity were built and subjected to cycling for 100 cycles, demonstrating a capacity fade of 67% [13]. Though their low power output is a drawback, this can be offset by employing HESS technologies that can enhance the power output when the vehicle is accelerating [82]. The limited lifespan of Li-S batteries is a challenge. The growth of dendrites at the anode and the shuttle effect at the cathode, both resulting from substandard electrodes, separators, and electrolytes, contribute to this problem [79]. Common knowledge asserts that the Solid Electrolyte Interface (SEI), which emerges on the anode from the breakdown of electrolytes and solutes, serves as a protective layer and markedly affects the battery's durability [83]. Nonetheless, the expansion of the SEI layer is also a significant contributor to LIB aging. Recent studies report the development of a innovative Li-S battery that features bilateral SEI layers on both the anode and cathode, represented in Fig 4.11. This innovative battery, by encouraging the concurrent development of SEI layers on both electrodes through the use of composite electrolytes, inhibits the formation of dendrites on the lithium anode and counters the shuttle phenomenon at the sulfur cathode [13]. This state-of-the-art Li-S battery exhibits not only exceptional performance characteristics, such as fast charging, low self-discharging, but also a commendable cycle life and high Coulombic efficiency.

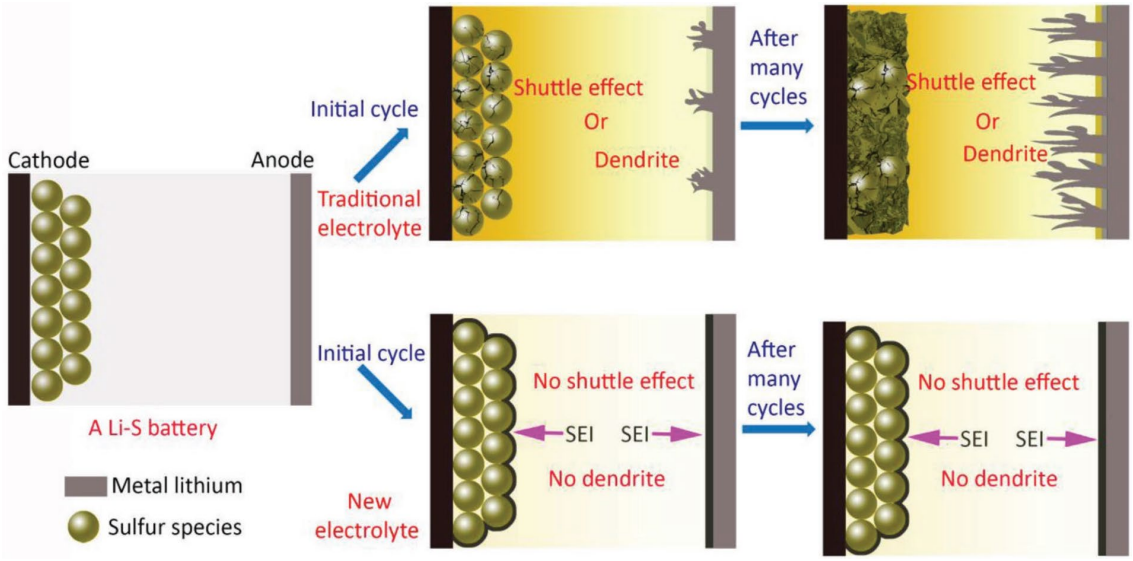

Fig 4.11 Bilateral SEI of the Li-S battery [13]

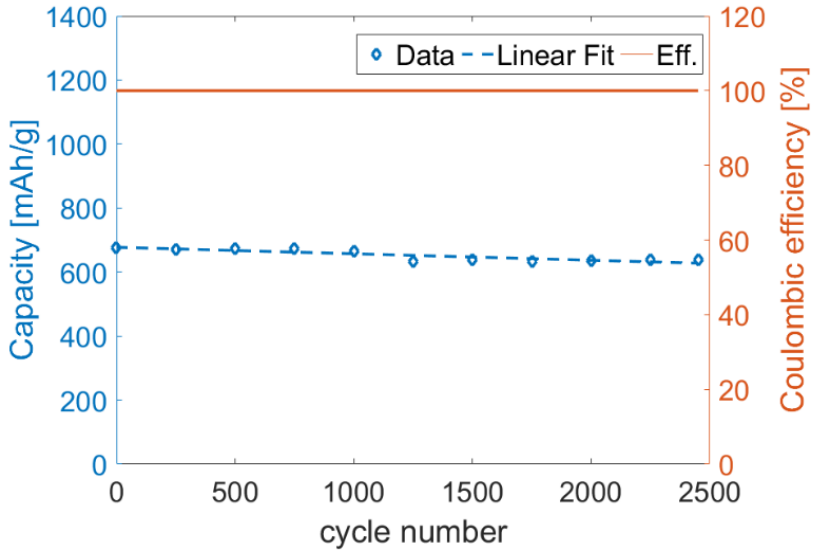

Fig 4.12 Battery degradation

The battery degradation model is developed using a linear fitting function, employing data from test electrolyte materials studied in Li-S batteries. Coulombic efficiency and

capacity measurements over 2500 cycles are depicted in Fig 4.12, along with the corresponding degradation correlation.

$$C_{Li-S} = 677.28 - 0.0203N \tag{4.20}$$

where C_{Li-S} represents the current capacity, N stands for the cycle number. The parameters are derived by linear regression using the data sourced from [13].

4.2.5 Supercapacitor model

The voltage and current of SC is calculated as follows:

$$U_{SC} = U_{SC,OC} - I_{SC}R_{SC} \tag{4.21}$$

$$I_{SC} = \frac{P_{cap}}{U_{SC}} \tag{4.22}$$

where U_{SC} represent the SC load voltage, $U_{SC,OC}$ denotes the open-circuit voltage of SC, I_{SC} stands for the current of SC, and R_{SC} refers to the resistance. The current is determined by the SOC.

$$I_{SC} = C_{SC}U_{SC,max}\Delta SOC_{SC} \tag{4.23}$$

where C_{SC} is the SC capacity, $U_{SC,max}$ is the maximum voltage of the SC, and the SOC_{SC} is the SOC of SC . The SOC_{SC} can be presented as follows:

$$SOC_{SC}(\mathrm{t}) = SOC_{SC}(0) - \frac{\int_0^t I_{SC}(\tau)d\tau}{C_{SC}U_{SC,max}} \tag{4.24}$$

Due to constraints imposed by the testing equipment, the supercapacitor model utilized in this dissertation corresponds to the model outlined in a previously published journal paper [84]. Subsequent stages of this research will involve comprehensive testing of the supercapacitor to construct and validate the employed model.

4.2.6 EM model

In Fig 4.13, the map illustrating the efficiency of the EM is presented, encompassing all losses associated with electromagnetic processes. The EMS derives the EM efficiency by integrating the demanded torque with the rotational speed and feeding this information into the efficiency map. Furthermore, during regenerative braking, the electromagnetic system transitions into a generator to recharge both the LIB and SC, thereby potentially enhancing overall energy efficiency.

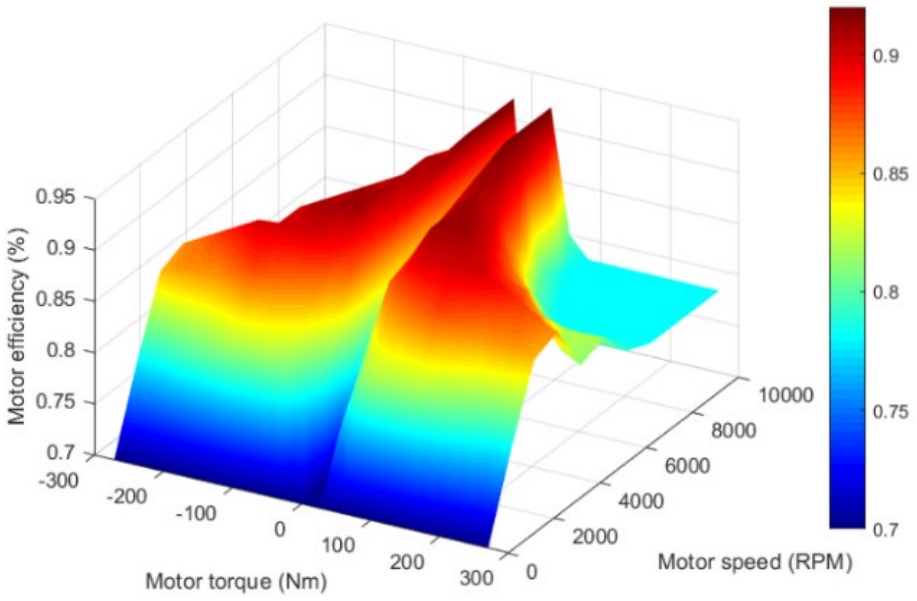

Fig 4.13 Efficiency map of EM

4.2.7 DC/DC converter model

In this dissertation, the integration of the SC and LIB into the propulsion system is achieved using separate DC/DC converters. The provided power and current characteristics, illustrated in Fig 4.14, are influenced by the efficiency of the respective DC/DC converter.

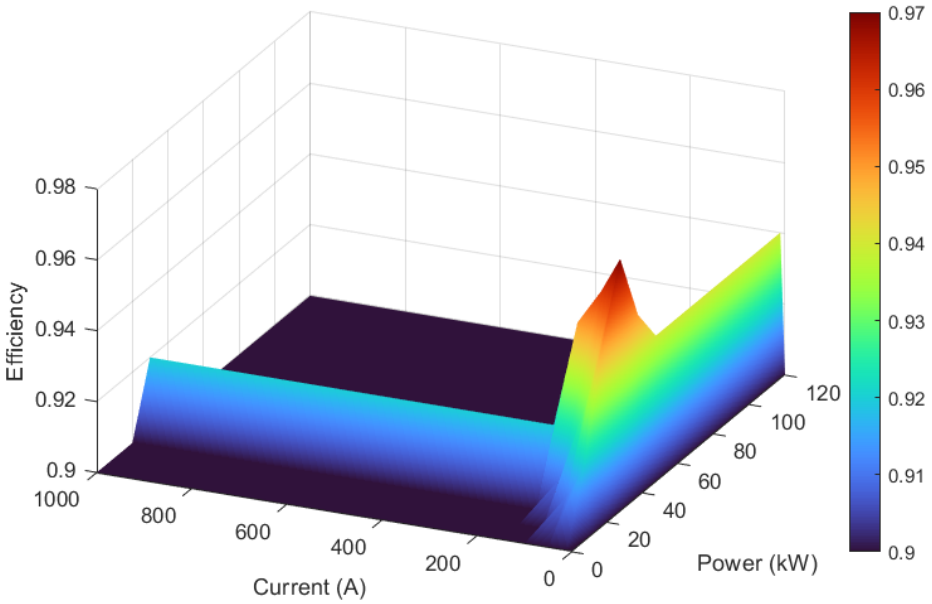

Fig 4.14 Efficiency map of the DC/DC converter

4.3 Q-learning based EMS

Optimization-based EMSs often impose significant computational demands, posing challenges for real-time control applications. Additionally, their implementation typically necessitates prior knowledge of driving cycles, which may not always be accessible or easily estimated. Recently, RL has been adopted as a promising method to address these limitations in conventional EMSs. In RL algorithms, the objective of the agent is to identify actions that maximize the cumulative reward. This process often involves a trial-and-error

approach to determine the correlation between observations and the most effective control strategies. The environment detects the state as observations, leading to the choice of particular actions for controlling the model during the process of trial and error. Subsequently, rewards are adjusted based on the effectiveness of the control. Optimal training in RL hinges on effectively balancing exploration and exploitation. Exploration allows the agent to gather information about the environment, while exploitation leverages existing knowledge to select actions with the highest potential reward. Typically, this balance is achieved through algorithms like ε-greedy action search. In reinforcement learning, the agent's environment is modeled as a Markov Decision Process, where future state probabilities depend solely on the current state, disregarding the sequence of events. The agent undergoes updates through interactions with the environment, wherein it chooses suitable actions at corresponding states to influence the reward for its update. In the EMS control scenario, the environment encompasses multiple factors impacting the vehicle's operational conditions, including velocity, acceleration, and powertrain dynamics. In EVs equipped HESS, the RL agent serves as a power-allocation controller, regulating the output power of both the LIB and SC. The primary goal is to identify a control sequence that decrease battery aging while optimizing energy utilization.

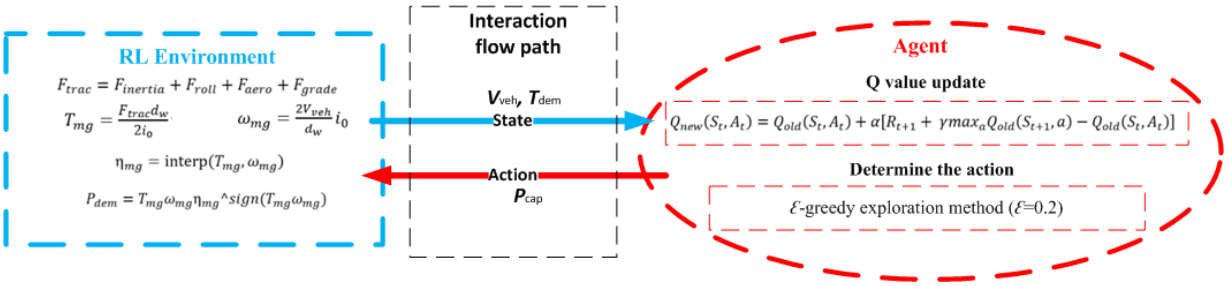

Fig 4.15 Q-leaning training flow

Q-learning is recognized as an off-policy Temporal Difference (TD) learning algorithm, representing a foundational component of RL [85]. This algorithm merges the advantages of Monte Carlo chain and DP techniques. It iteratively refines its estimates through interactions with a predetermined environment, which encapsulates the vehicle model along with other pertinent variables influencing the vehicle's behavior. Fig 4.15 illustrates the interactive process that unfolds during the training process. As a result, the Q-learning algorithm can focus on updating the value function instead of explicitly updating the optimal policy. When the system is in a given condition, the ideal control approach can be derived by evaluating the value or utility of potential actions. This allows determining the optimal sequence of control inputs to apply.

$$Q_{new}(S_t, A_t) = Q_{old}(S_t, A_t) + \alpha[R_{t+1} + \gamma max_a Q_{old}(S_{t+1}, a) - Q_{old}(S_t, A_t)] \tag{4.25}$$

where Q_{new} signifies the updated Q value of the current state S_t and action A_t at the next step, Q_{old} denotes the Q value of the current state S_t and action A_t at the current step, α stands for the learning rate, R_{t+1} represents the instantaneous reward from the environment, the discount factor γ underscores the value of future rewards, $max_a Q_{old}(S_{t+1}, a)$ represents the maximum Q value of the next state S_{t+1} among all possible actions, the learning rate α determines the extent to which the previous Q-value influences the current Q-value.

Throughout the training process, the states encompass the demanded torque (T_{dem}) and the vehicle's velocity. The action taken is the output power of the SC. Subsequently, the reward is computed based on the aging of the LIB and the overall energy usage of the HESS.

$$R = -\omega(E_{bat} + E_{SC}) - (1 - \omega)Q_{loss} + \beta \tag{4.26}$$

where the parameter ω serves as a weighting factor, influencing the balance between energy preservation and battery degradation, Additionally, β denotes a constant bias. E_{bat} represents the energy usage of the LIB, while E_{SC} signifies the energy usage of the SC.

It is essential to notice that Q-learning is traditionally employed for maximization problems, whereas EMS typically address energy minimization objectives. To align these frameworks, negative signs are introduced to battery degradation and energy consumption functions, converting the minimization task into a maximization one. However, this approach presents a challenge: negative rewards may cause selected actions to vanish, as the ε-greedy selection strategy favors actions with higher reward values. To mitigate this issue, an offset parameter β is introduced to ensure the objective functions yield positive values. The rewards obtained from each iteration of the Q-learning-based EMS are stored in a cache, as depicted in Fig 4.16. It is observed that initially, there is a sharp increase in the reward over the first 300 iterations, followed by a gradual deceleration in the rate of increase. After approximately 1000 iterations, the reward value begins to stabilize, fluctuating around a constant value. These fluctuations can be attributed to the ε-greedy exploration strategy, which allows the system to search and learning the environment with a probability of ε, leading to occasional random action selections. Consequently, while the ultimate reward may be lower than that of the previous iteration, the overall trend indicates an increase in reward over time.

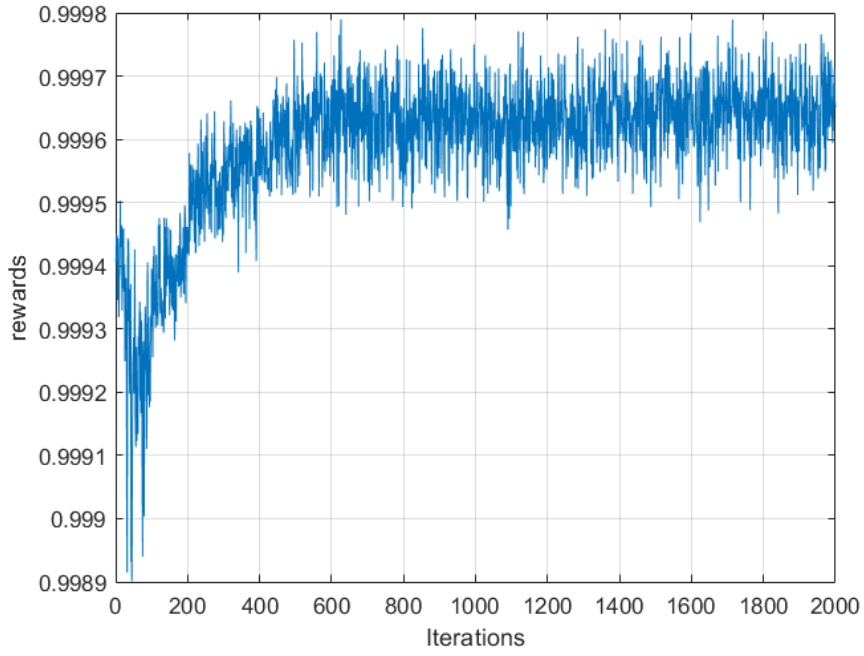

Fig 4.16 Q-learning reward

To assess the control effect of the EMS based on Q-learning and the mitigation of battery degradation using a novel Li-S battery, a series of simulations were conducted to analyze energy consumption and battery aging throughout the entire lifespan of an EV. A total of 1000 simulations were performed, each representing a distinct instance. The observed trend, as depicted in Fig 4.17, demonstrates that the normal LIB capacity degradation significantly impacts the SOC over the course of these 1000 simulations. Comparing the SOC values between the first and the 1000th simulation, it is evident that the SOC decreases at a much faster rate as the LIB's capacity diminishes due to aging effects. Upon concluding the initial test, the SOC value reached 0.4635. Once the 1000th simulation concluded, the final SOC value declined further to 0.4077. This discrepancy reveals the progressive degradation of the LIB's capacity.

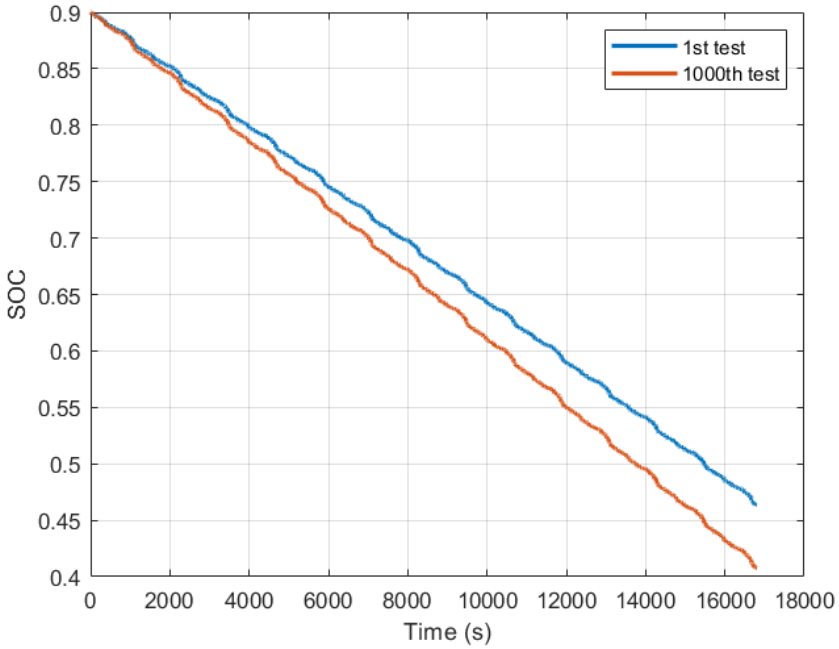

Fig 4.17 LIB SOC trajectories

In the contrast to the observations made with the normal LIB, the Li-S battery exhibits negligible degradation following 1000 simulations. This is evident from the nearly identical SOC trajectory observed between the 1000th simulation and the initial simulation, as illustrated in Fig 4.18. The EV model equipped with the Li-S battery demonstrates remarkable consistency in SOC performance over an extended period. During the first test, the EV model equipped with the Li-S battery achieved a final SOC value of 0.4635. Remarkably, even after 1000 simulations, the final SOC value of the Li-S battery remained high at 0.4545. The variance in SOC, calculated as a percentage after 1000 tests, is a mere 1.94%. However, the normal LIB exhibited a SOC variance of 12.04%. This substantial difference clearly indicates the superior long-term performance and lifespan of the EV model equipped with the novel Li-S battery when compared to the conventional LIB. These

results show the potential of the Li-S battery as a promising alternative to LIB, offering improved performance and a more sustainable solution for electric vehicles in the long run.

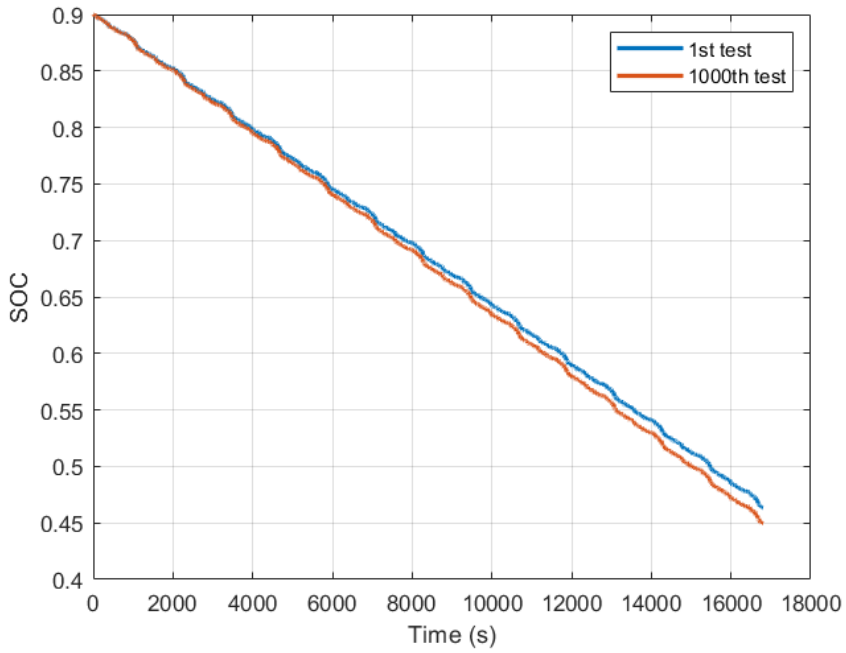

Fig 4.18 Li-S battery SOC trajectories

In accordance with the degradation model outlined in subsection 3.3.3, a comparative analysis of the LIB degradation processes between the initial test and the 1000th test is presented in Fig 4.19.

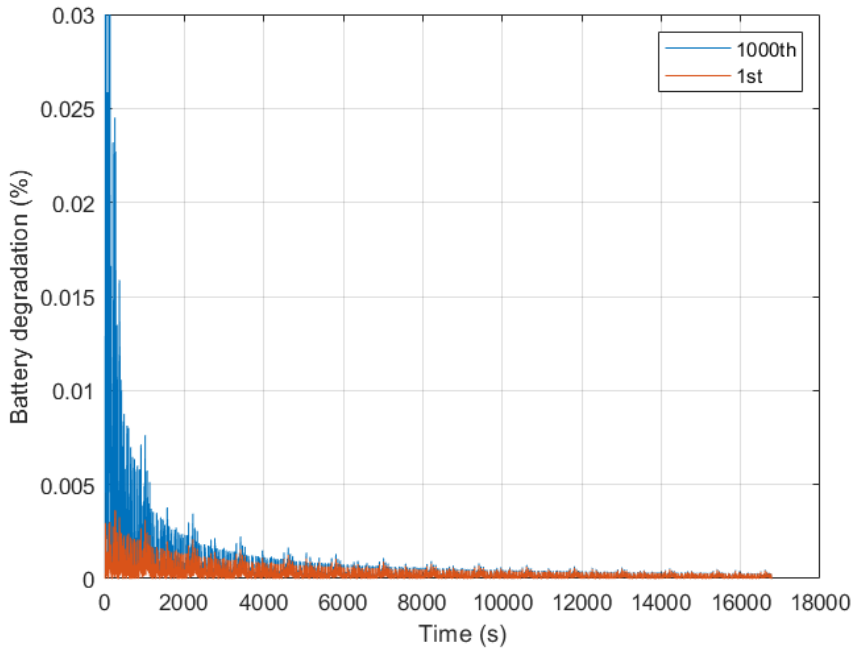

Fig 4.19 LIB degradation comparison

Owing to heightened battery aging effects, a substantial loss of capacity occurs in the LIB prior to the 1000th test. Consequently, during the 1000th test, the LIB experiences a more deeper discharge to extract sufficient energy to meet the driving requirements. This increased depth of discharge is the principal cause for the notably higher degradation observed during the final test in comparison to the initial test. After the 1000 tests, the capacity degradation of the LIB is measured at 11.10%. Furthermore, employing the Li-S battery model detailed in subsection 3.3.4, the degradation observed in the lithium-sulfur (Li-S) battery stands at 2.97%. Notably, this degradation level accounts for only 26.76% of the degradation witnessed in the LIB.

To indicate the control performance of the Q-learning based EMS, the DP-based EMS is regarded as baseline to verify the near-optimal results of the proposed method. Also, the rule-based EMS is adopted to examine the accurate management of the proposed method. The energy consumptions are shown in Table 4.3. Looking at the energy cost in terms of kilowatts per hour (kW/h), the Rule-based EMS achieved an energy cost of 30.76 kW/h when using the Li-S battery, while the LIB version had a slightly higher energy cost of 31.84 kW/h. This indicates a modest improvement of 7.68% when using Li-S battery over LIB. The DP-based EMS demonstrated better performance, with the Li-S battery version consuming only 26.16 kW/h compared to the LIB version at 29.12 kW/h. This shows a more substantial improvement of 11.32% when using Li-S battery over LIB. The Q-learning EMS strategy achieved an energy cost of 27.59 kW/h with the Li-S battery, while the LIB version had an energy cost of 30.46 kW/h. This translates to an improvement of 10.40% when using Li-S battery over LIB. Overall, the comparison reveals that both the DP and Q-learning-based EMS strategies, when coupled with Li-S battery, offer notable energy consumption improvements compared to their LIB counterparts. And the Q-learning based EMS has reached a near-optimal control performance. This underscores the potential of Li-S battery technology to enhance energy efficiency and reduce operational costs in electric vehicles.

Table 4.3 Energy consumption comparison of different EMS

EMS	Energy cost (kW/h)		Improvement
	Li-S battery	LIB	
Rule-based	30.76	31.84	7.68%
DP	26.16	29.12	11.32%
Q-learning	27.59	30.46	10.40%

The battery degradation results are presented in Table 4.4, which presents a comparison of battery degradation among different EMS.

Table 4.4 Battery degradation comparison of different EMS

EMS	Battery degradation (%)		Reduction
	Li-S battery	LIB	
Rule-based	2.97	11.69	78.63%
DP	2.97	10.25	65.51%
Q-learning	2.97	11.10	73.24%

In terms of battery degradation, it is evident that all three EMS strategies exhibit notably lower degradation rates when used with Li-S battery compared to LIB. The Rule-based EMS with Li-S battery demonstrated a degradation rate of 2.97%, whereas the LIB version had a much higher degradation rate of 11.69%. This highlights a substantial reduction in battery degradation of 78.63% when utilizing Li-S battery with the Rule-based EMS. Similarly, the DP-based EMS showed a degradation rate of 2.97% with Li-S battery, whereas the LIB version had a higher degradation rate of 10.25%. This translates to a reduction in battery

degradation of 65.51% when using Li-S battery with the DP-based EMS. For the Q-learning EMS, the Li-S battery exhibited a degradation rate of 2.97%, while the LIB version had a slightly higher rate of 11.10%. This signifies a reduction in battery degradation of 73.24% when using Li-S battery with the Q-learning EMS strategy. These results highlight the significant advantages of Li-S battery technology in mitigating battery degradation across various EMS strategies. The utilization of Li-S battery consistently leads to substantially lower rates of battery degradation, showcasing its potential to enhance the lifespan and overall performance of electric vehicles. This comparison also verify that the proposed Q-learning based EMS has better control effect than the rule-based EMS. Although it cannot be accurately as the DP-based EMS, it still can achieve a near-optimal performance.

4.4 Conclusion

In Chapter 4, a sophisticated Q-learning based EMS for EV has been presented, as an innovative alternative to traditional rule-based and optimization-based systems. Initially, the chapter delves into the dynamics of EVs and the nuances of their propulsion systems, with a focus on modeling an advanced Li-S battery system. I explored different HESS configurations, highlighting their unique benefits and challenges. The crux of the chapter is the introduction of a Q-learning based EMS, designed for its adaptive learning capabilities for optimizing energy distribution between the battery and supercapacitor in real-time, eliminating the need for preset rules or predictive models. Simulations using the EV driving cycle developed in Chapter 2 demonstrate the Q-learning EMS's effectiveness in reducing energy consumption and battery wear. The results showed that the Q-learning based EMS performs on par or better than conventional methods, offering real-time

adaptability without relying on driving condition forecasts. This chapter asserts the significance of Q-learning in advancing EV energy management, suggesting that such intelligent systems are not just theoretically promising but also practically relevant for enhancing the efficiency and longevity of EV energy systems.

CHAPTER 5

IMITATION LEARNING BASED EMS FOR ELECTRIC VEHICLES

In the former chapter, a Q-learning based EMS has been proposed for the EV equipped with a novel Li-S battery. The control performance has been proved when compared with some conventional EMS. However, the time cost and computational burden compensate the real-time implement. In order to mitigate the computational overhead associated with the learning process in the Q-learning based EMS, this chapter introduces an alternative approach known as imitation learning EMS.

5.1 Research gaps and proposed method

While RL-based EMSs have shown advancements in energy management problems, a notable obstacle remains, hindering their practical deployment in real-world settings. Due to a large number of iterations, the long learning time can be too long to update the RL agent and it is not well addressed in existing studies. As previously discussed, the RL agent endeavors to strategize its actions through iterative trial-and-error interactions with a dynamic environment. This learning approach, while effective, is characterized by its relatively low efficiency and demands substantial experiential data gathered from interactions. Existing literature suggests that RL EMSs utilizing table-based method typically require large number of iterative steps, ranging from 2000 to 300000 iterations [37], [47], [86], while RL EMSs employing neural networks may need a range of 15000 to 150000 iterations [44], [87]. Such extensive iteration counts impose a significant computational burden on the system, necessitating lengthy dyno and road testing periods,

which is time-consuming and costly. This dissertation proposes a novel approach to expedite the training process of RL EMS, it introduces a new method: an imitation Q-learning algorithm-based EMS designed specifically for EVs incorporating LIB and SC. Imitation learning, as an algorithmic approach, instructs the agent to emulate expert behavior [56]. In contrast to RL, imitation learning operates without relying on a reward function for agent updates; instead, it involves an expert who provides demonstrations to guide the agent's learning process. By imitating the decisions of the expert, the agent learns an optimal policy. Subsequently, the agent can effectively map states to actions by utilizing the expert's demonstrated experiences [88]. In this dissertation, the imitation techniques is integrated into the training phase of the normal Q-learning algorithm. To the best of our knowledge, this marks the first application of imitation Q-learning in optimizing the control of the EV equipped with HESS.

While imitation learning offers several advantages over traditional RL methods, it also has some challenges. One challenge is the need for high-quality expert demonstrations. If the expert demonstrations are not good, then the agent will not be able to learn a good policy. Another challenge is the difficulty of generalizing to new situations. If the agent is trained on data from a specific environment, then it may not be able to perform well in a different environment. In this dissertation, the imitation learning technique is exploited to short the training phase of EV equipped with LIB and SC with RL-based EMS. The application of imitation learning during the training phase of the conventional Q-learning algorithm marks a novel approach, particularly in optimizing the control of the EV equipped with HESS. This study makes a triple contribution to the existing literature.

Firstly, it introduces an imitation learning-based EMS for HESS, aiming to decrease the quantity of iterations during the training phase and consequently cut down the overall training duration. Secondly, the proposed methodology takes into account both efficiency and battery aging, enhancing the holistic approach to EMS design. Lastly, the control effect of the imitation learning-based EMS is rigorously assessed through comparisons with Q-learning, heuristic rules, and DP-based EMSs, providing validation and verification of its effectiveness.

5.2 Imitation learning based EMS

In order to reduce the duration of training for RL algorithms, multiple strategies can be employed, including selective experience learning, modifications to the learning rate, and imitation techniques. Selective experience learning prioritizes the use of pre-existing experiences, utilizing an experience buffer during decision-making to store and recycle transition experiences. This approach aims to populate the experience buffer with experiences that are nearest to optimal outcomes. However, the repeated reliance on previous experiences might negatively influence control performance. Such a method may prevent the agent from exploring various environmental regions and circumvent actions that lead to subpar performance. It could also overlook certain rare experiences that might emerge under specific circumstances. Adjusting the learning rate involves implementing a flexible learning rate. Eq. (3.25) demonstrates that an increase in the learning rate can expedite the convergence of Q values. Yet, this approach might cause the control effect for the trained agent to become erratic. A higher learning rate certainly quickens the convergence pace, but it simultaneously leads to more substantial updates in rewards. This,

in turn, unjustly inflates the Q values of certain actions. The imitation approach, on the other hand, emphasizes the modification of the initial value. It leverages expertise or heuristic guidelines to modify the Q table during the initial iteration at the commencement of the training for the RL agent. Contrary to methods that balance the considerations of learning duration and control efficacy through selective learning experiences or adjustments in learning rate, the imitation approach delivers an immediate enhancement at the outset, aiming to speed up the training process.

In the development of an EMS for the HESS employing the imitation Q-learning technique, the traditional Q-learning method initializes Q values at zero during the initial phase. Conversely, in the imitation Q-learning approach, the initial Q table is updated by utilizing the application of heuristic principles. When rapid acceleration is required to satisfy the driver's expectations, the propulsion system is tasked with providing sufficient power. When the power demand is met by the battery, it experiences rapid and deep discharges, which could lead to substantial wear and tear. To safeguard the battery against severe deterioration under such circumstances, the EMS is configured to request an increased power contribution from the SC. Additionally, during the slowdown of the vehicle, the HESS can be replenished through the energy harnessed from regenerative braking. In scenarios where braking is aggressive, the generated regenerative energy might exceed the LIB safe absorption capacity, risking safety and accelerating degradation. Consequently, the EMS is programmed to direct the SC to initially capture the regenerated energy, leveraging its rapid charge/discharge capabilities. This approach is guided by heuristic rules detailed in reference [34], as illustrated in Eq (5.1).

$$P_{cap} = \begin{cases} A_1 P_{dem}, & \text{if } P_{dem} > a_{dischg} \text{ and } SOC_{cap} > SOC_{cap,min}\ ; \\ A_2 P_{dem}, & \text{if } P_{dem} < a_{chg} \text{ and } SOC_{cap} < SOC_{cap,max}; \end{cases}$$

(5.1)

where the coefficients A_1 and A_2 serve to calculate the SC output power in proportion to the power demand, whereby P_{dem} represents the power requirement. The heuristic parameters a_{dischg} and a_{chg} define the power demand thresholds during the vehicle's acceleration and deceleration phases, respectively. Furthermore, the operational limits of the SC's SOC are delineated by $SOC_{cap,min}$ and $SOC_{cap,max}$, indicating the minimum and maximum SOC ranges permissible for SC functionality.

In the section on imitation learning, the proposed EV driving cycle is adopted as the speed profile for imitation Q-learning strategy. The diagram of the imitation Q-learning base EMS is shown in Fig 5.1. Firstly, the conventional EMS are designed through the same environment as the Q-learning that create the heuristic rules for improve the energy efficiency and lower the battery degradation. Then, the knowledge of the pre-trained expert is transferred to the Q-learning agent to boost the first training episode. Next, the pre-trained agent is updated through normal training process until the results converged. Finally, the imitation Q-learning based EMS will be delivered to the vehicle model for the optimal control.

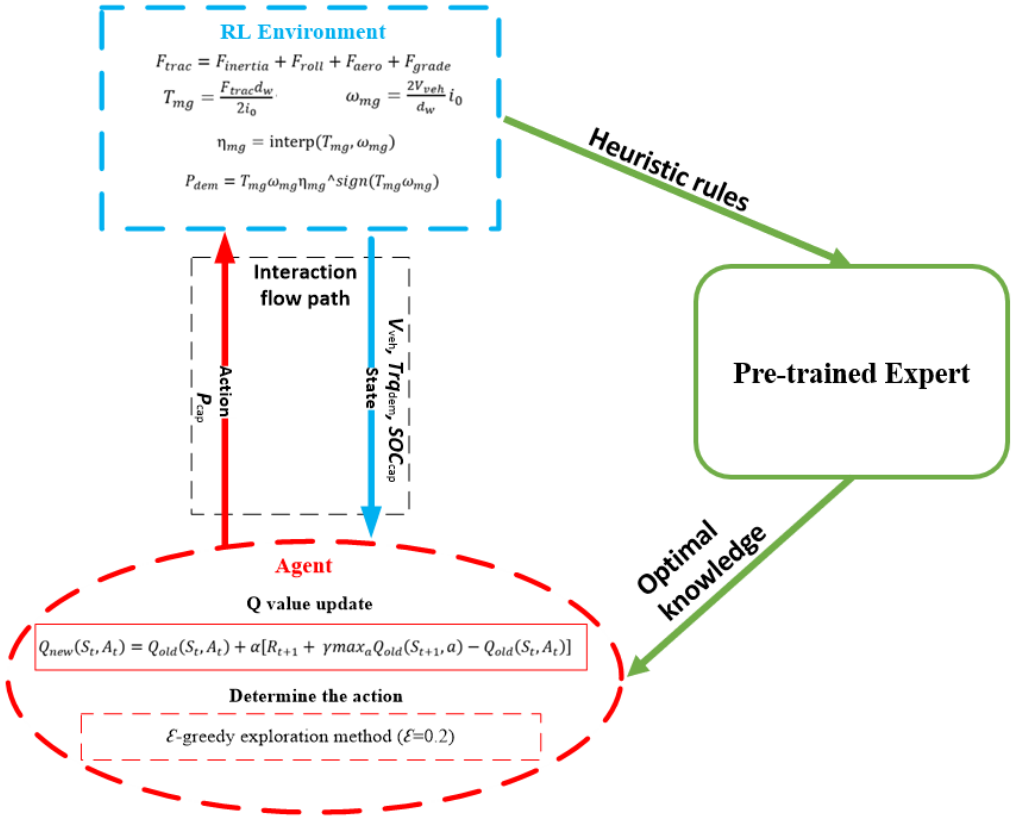

Fig 5.1 Imitation Q-learning diagram

Fig 5.2. displays the initial outcomes. Actions are depicted along the X-axis, where the spectrum extends from the maximum power output during discharging to the peak power during charging, distributed evenly across 50 segments. The Y-axis delineates pairs of states within a 20x20 grid, incorporating two variables: vehicle speed and required torque. The Z-axis, on the other hand, illustrates the Q value associated with each specific action and state combination. Subsequent to the completion of the training phase, the ultimate Q values are illustrated in Fig 5.3.

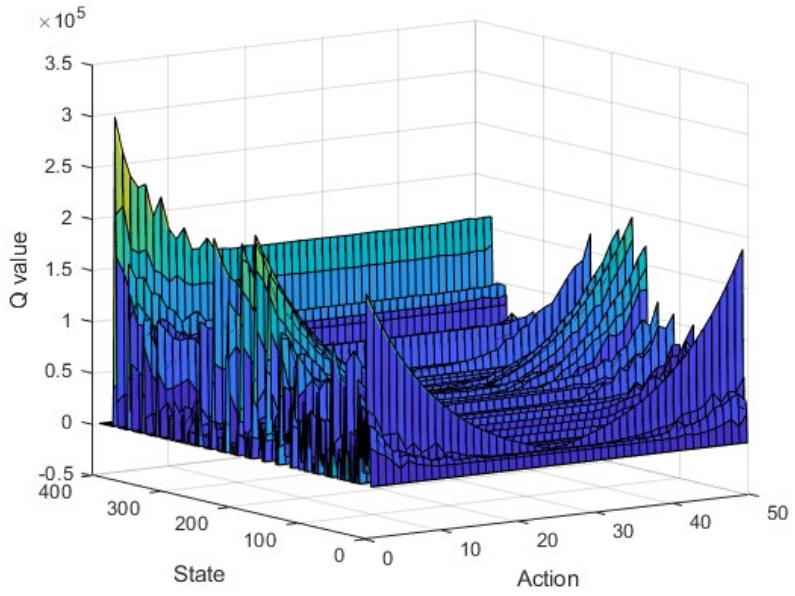

Fig 5.2 Initial imitation result.

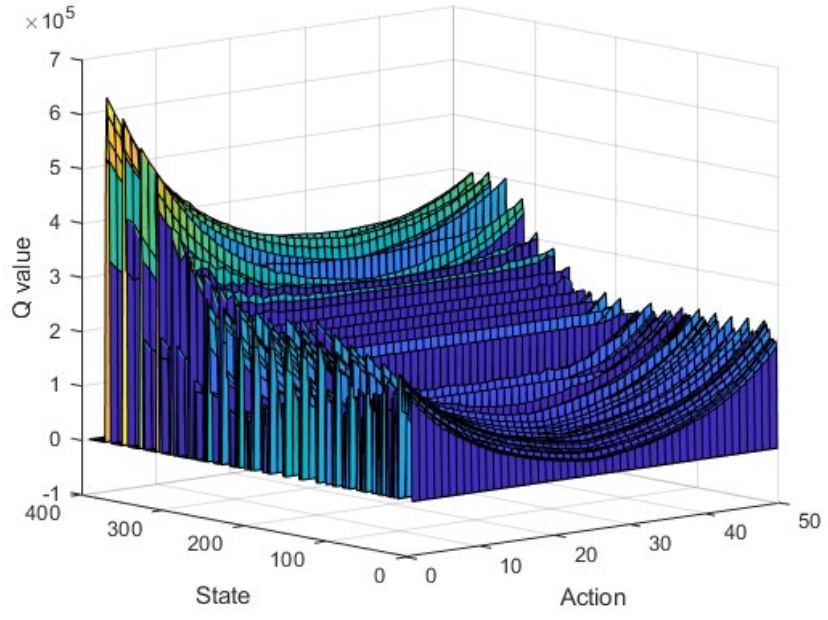

Fig 5.3 Final Q value.

Analyzing the Q values from the starting point to the end reveals that the initial imitation surge pushes the Q-value to $3 * 10^5$, ultimately climbing to $6 * 10^5$. This indicates that the heuristic guidelines provide a significant initial impetus to the Q values, maintaining a consistent trajectory with the eventual outcomes. Throughout the training phase, there is a steady increase in the relative Q values, leading to updates in the RL agent. This analysis highlights how the imitation strategy effectively elevates the baseline for the training process of the RL agent.

Fig 5.4 illustrates how rewards are accumulated through the training process. From the illustrated reward trajectories, it is evident that the initial rewards for the imitation Q-learning approach are significantly greater compared to those of the conventional Q-learning method. Moreover, the imitation Q-learning method reaches a stable state more swiftly.

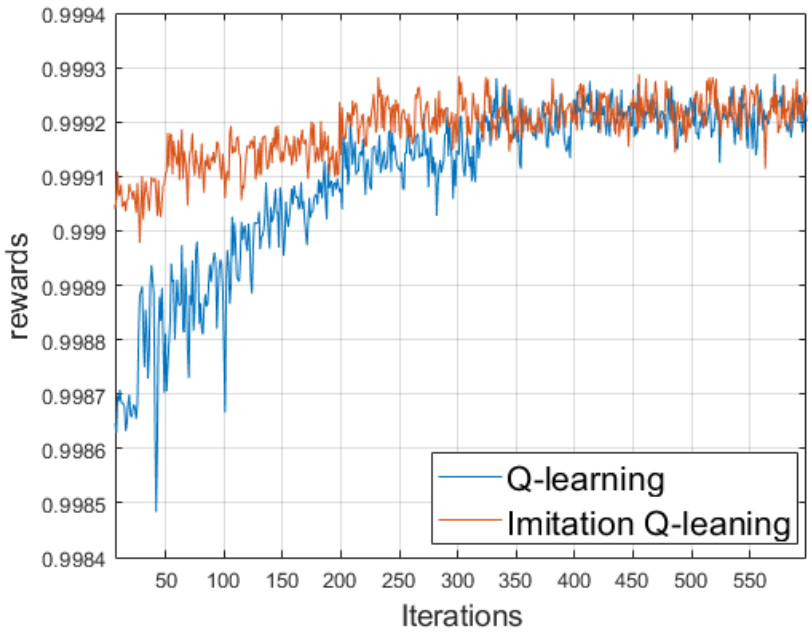

Fig 5.4 Reward trajectories during training.

5.3 Results of imitation Q-learning based EMS

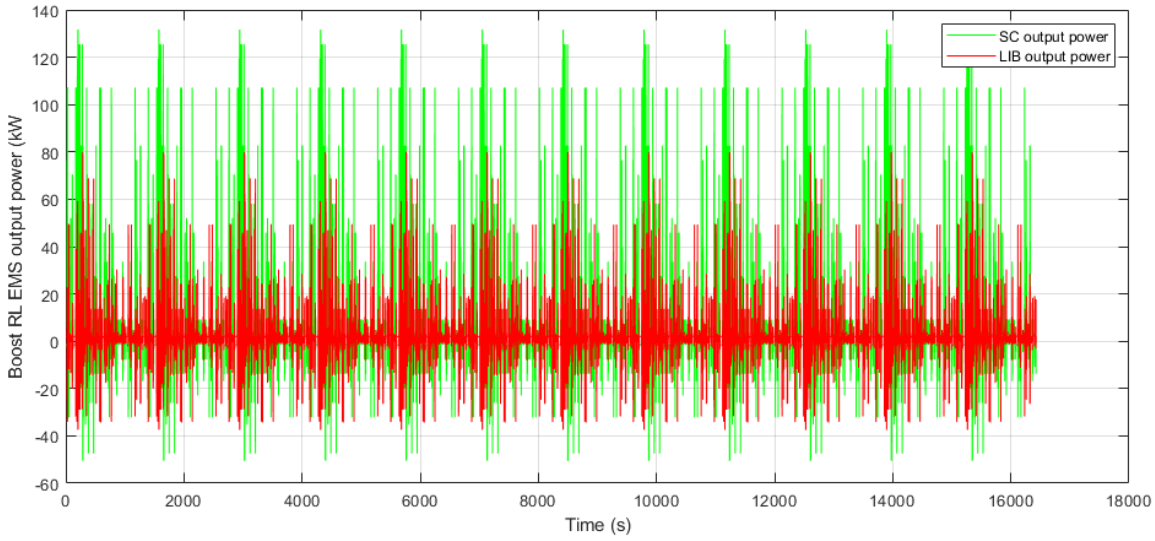

Fig 5.5 HESS output power.

Fig 5.5 displays the power output of the HESS component. The findings indicate that the SC is responsible for handling nearly the entirety of the vehicle's peak power requirements during acceleration and deceleration throughout the simulation phase. The SC generates maximum power output during its discharging phase, particularly when encountering high power demands. Nevertheless, this high output level is not sustainable over extended periods due to the limited energy storage capacity of the SC. The LIB ensures a steady energy supply due to its higher energy density, which allows for more energy transport compared to the SC. During charging, the SC captures the highest peak power. Any additional energy is directed to charge the LIB only after the SC is fully charged. By doing so, the SC significantly mitigates the peaks in battery charging and discharging. After 1000 cycles in the HESS, the LIB experiences a degradation of 21.45%.

When compared to a system with a single ESS, the degradation of the battery within the HESS is lessened by 26.36%, under the control of the imitation Q-learning EMS.

Based on the outcomes of the evaluation, there is an improvement in energy efficiency observed. According to Fig 5.6, the comparison between the SOC for a standalone LIB and a combination of LIB with SC within a HESS demonstrates that the latter maintains a higher charge level following a full driving cycle. When examining Fig 5.5 and Fig 5.6 side by side, it is evident that the HESS configuration is more efficient at capturing energy from regenerative braking compared to a system using only a LIB. This efficiency gain in the HESS-based propulsion system, when compared to the LIB-only setup, is quantified as a 3.83% increase.

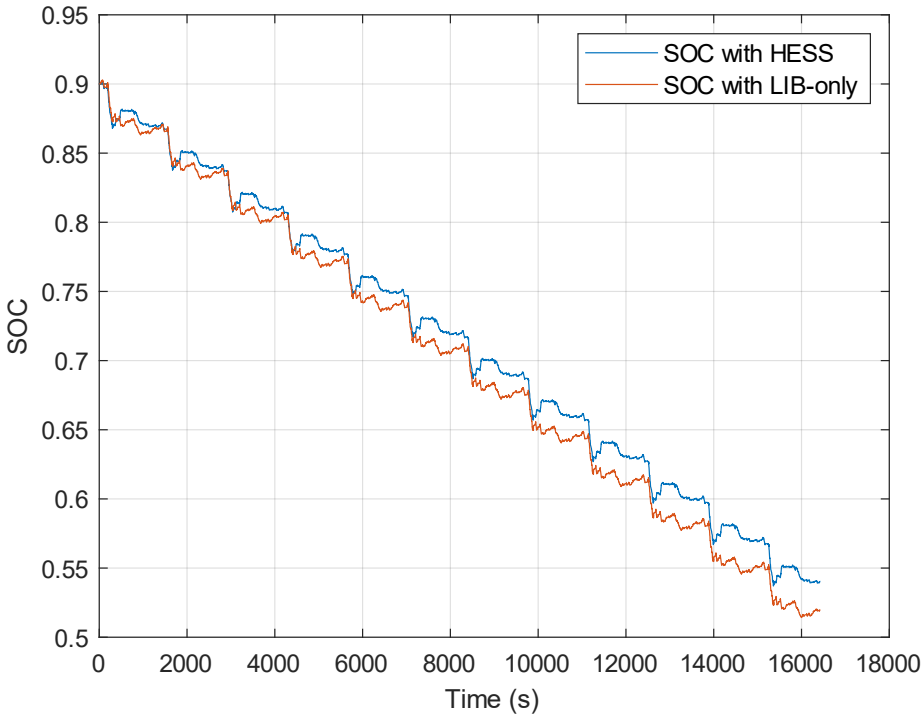

Fig 5.6 Battery SOC trajectories comparison

Figure 5.7 presents the progression of the SOC of the SC throughout the simulation period. The graph displays a starting SOC level of 0.6, diminishing slightly to 0.5 by the end of the simulation. The minimum SOC observed for the supercapacitor is 0.1; at this point, the control strategy utilize the power from the battery to propel the vehicle. Throughout the entire simulation, the SOC values for the SC consistently stay within the range of 0.1 to 0.95.

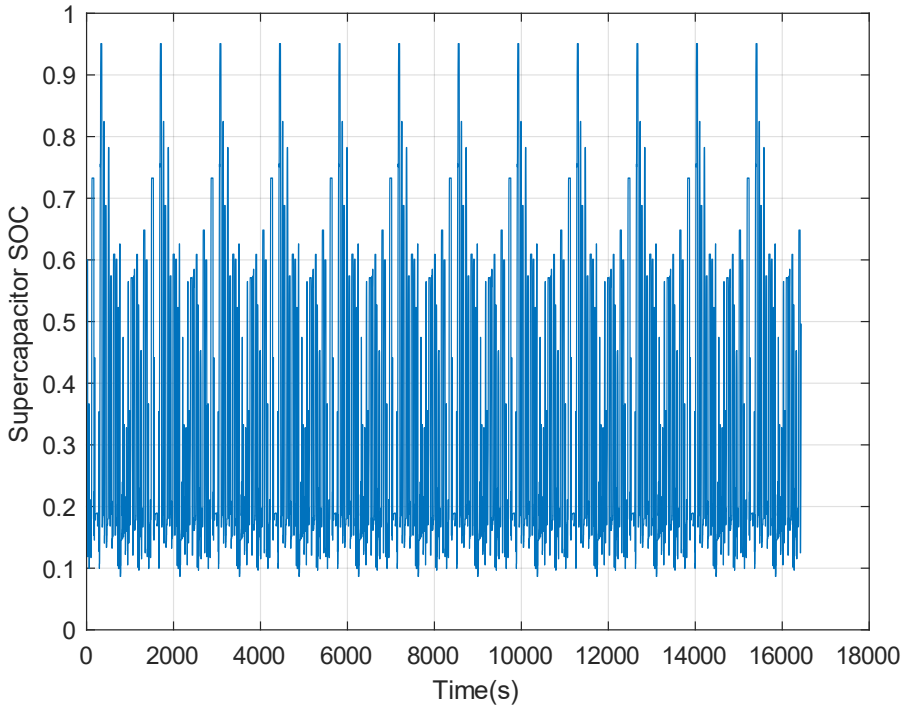

Fig 5.7 Supercapacitor SOC trajectory

This dissertation conducts simulations of the EV model utilizing various EMSs. The study aims to access the impact of the imitative Q-learning EMS by comparing it with the performance and computational efficiency of the rule-based EMS, DP, and traditional Q-learning-based EMSs, all within identical simulation settings. Fig 5.8 depicts the variations

regarding the energy distribution between the LIB and the SC within the HESS, as influenced by the different EMSs.

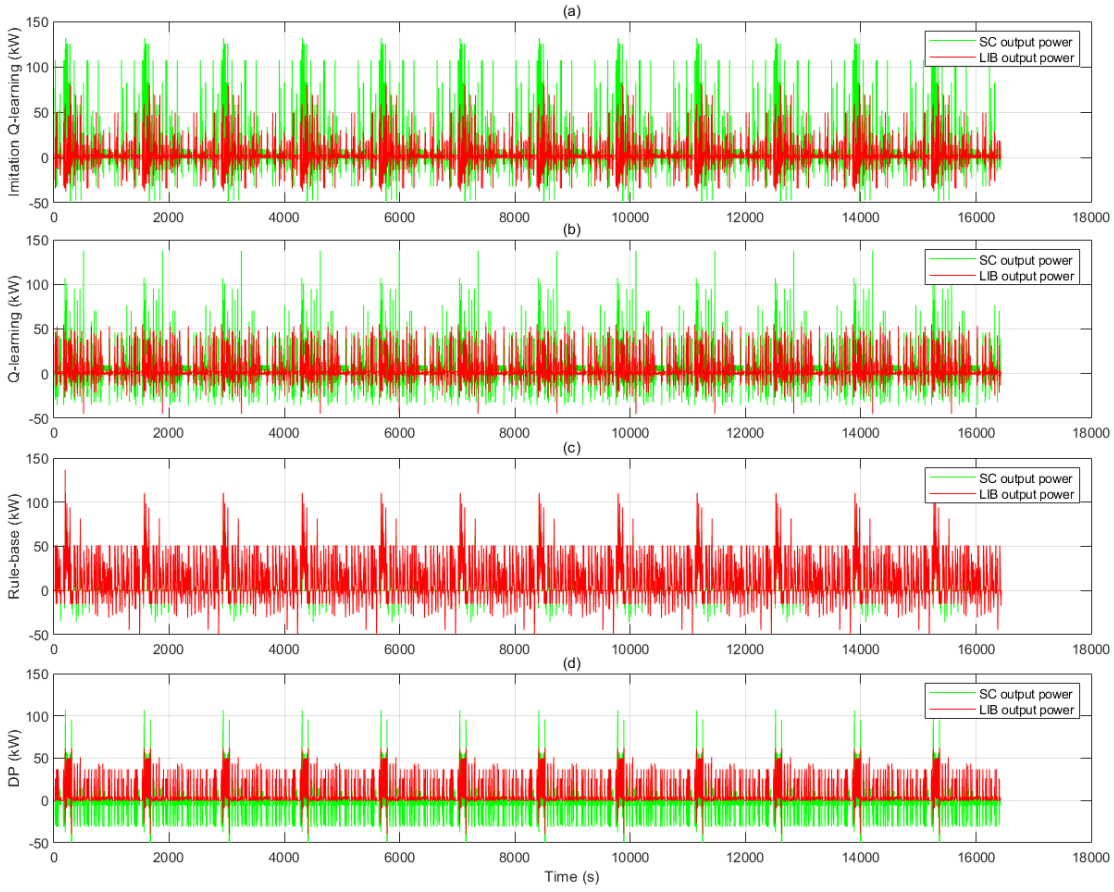

Fig 5.8 EMS comparison

In this analysis, the comparative assessment focuses on the energy distribution between the SC and LIB during the simulated driving cycle, administered under the control of four different EMSs. The findings highlight that the DP based EMS yields the most optimal control efficiency, attributed to DP's capacity to pinpoint the most universally ideal solution. As illustrated in Fig 5.8(d), the SC is tasked with managing peak energy demands through both charging and discharging processes, while the LIB ensures a sustained power supply

that facilitates vehicle operation and mitigates the extents of charging and discharging, thereby preserving battery life. The rule-based EMS, formulated on the foundation of experts' experience and insights, presents a viable approach within specific constraints. However, its performance might falter in scenarios not encompassed by the predefined heuristic rules, compromising its control effectiveness. In the investigation depicted in Fig 5.8©, the strategy involves solely utilizing the SC for the charging operations, with the LIB entrusted with the task of providing both peak charging power and consistent power essential for vehicle propulsion. This rule-based approach is known as less effective in reducing the wear and tear on the battery. In contrast, the imitation Q-learning and the traditional Q-learning methods, exhibit commendable control efficiency that is nearly optimal. These RL methodologies leverage the HESS, primarily employing the SC for managing the bulk of the peak energy demands, which significantly mitigates the 'IB's aging. Despite their overall similarity in performance, a minor distinction is observable between the imitation Q-learning and the conventional Q-learning EMS, as delineated in Fig 5.8(a) and (b), indicating nuanced differences in their operational dynamics.

T©e imitation Q-learning EMS optimizes the use of the SC by allocating it the role of managing all peak power requirements, in contrast to the traditional Q-learning EMS where the LIB is responsible for absorbing peak charging energies. This operational variance underscores a more strategic engagement of th' SC's potential in the imitation Q-learning method, wherein it takes on all high-demand energy situations, thus alleviating stress on the LIB and contributing to its longevity. The effectiveness of this approach in preserving battery health is further supported by the LIB degradation trajectories illustrated in Fig 5.9.

In a single simulation round, the loss in total battery capacity observed was $2.37 * 10^{-6}$ % for rule-based, $7.19 * 10^{-7}$ % for traditional Q-learning, $6.24 * 10^{-7}$ % for imitation Q-learning, and $1.86 * 10^{-7}$ % for DP based EMS.

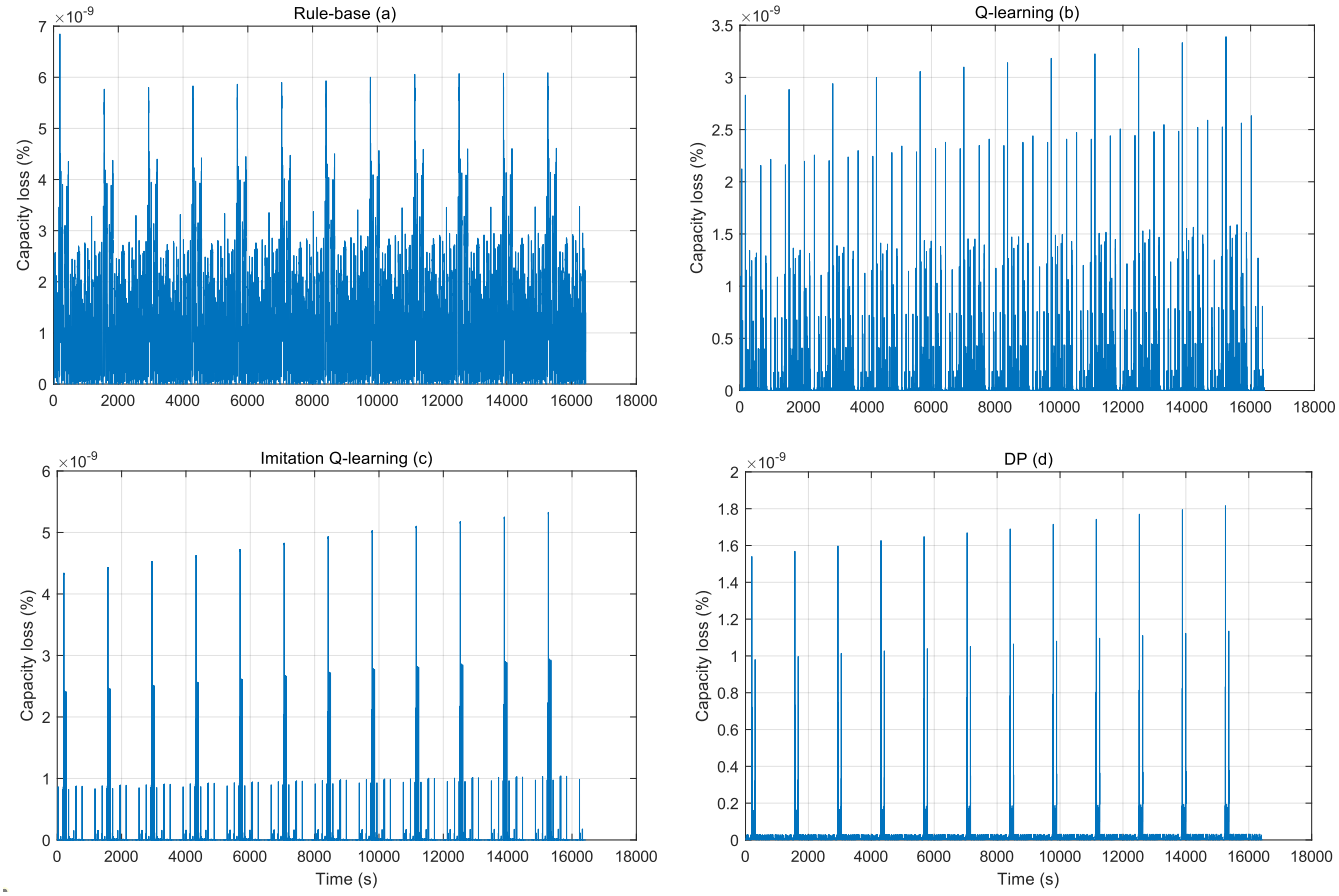

Fig 5.9 LIB degradation comparison

Observations reveal that the DP EMS outperforms in diminishing LIB wear and tear, marking a significant 67.99% reduction in comparison to the counterpart, the imitation Q-learning EMS. Analyzing the degradation patterns between Fig 5.9(c) and (d) proves that, despite their resemblance, the degradation level observed in the imitation Q-learning substantially surpasses that of the DP. This pattern is reiterated in the comparison between

Fig 5.9 (a) and (d). Under the governance of DP EMS, LIB primarily undergoes a consistent discharge process to facilitate vehicle operation, with charging occurrences being infrequent. In contrast, the imitation Q-learning approach subjects the LIB to both charging and discharging activities, resulting in an accelerated aging process when compared to the DP EMS. The imitation Q-learning approach demonstrates a notable reduction in lithium-ion battery degradation compared to the conventional Q-learning method, with the latter exhibiting 11.53% higher degradation. The one-time boosting process in the proposed technique provides a more favorable initial starting point, enabling the imitation Q-learning algorithm to converge more rapidly. Additionally, during the RL agent's training phase, the ε-greedy strategy introduces an element of randomness in the exploration process, contributing to some variance in the final results obtained by imitation Q-learning and conventional Q-learning energy management systems. The rule-based EMS does not fully leverage the potential advantages offered by the HESS configuration. As a result, this rule-based strategy leads to the most severe lithium-ion battery degradation among all the evaluated EMSs. The decline in the health of LIB using traditional Q-learning exceeds that observed with the imitation Q-learning approach by 11.53%. This differential can be attributed to the initial enhancement methodology utilized in imitation Q-learning, effectively reducing the time required for convergence. Moreover, the implementation of the ε-greedy strategy during the RL agent's training phase introduces a degree of unpredictability to the outcomes. Consequently, this variability contributes to the discernible performance disparity between the imitation Q-learning and the conventional Q-learning EMS. In comparison, energy management approaches based on set rules fail to

capitalize on the full potential of HESS, leading to the most pronounced degradation of LIB within this category of EMS.

Table 5.1 Computation time

Method	Value (s)
Rule-base	4.34
Conventional Q-learning	84.79
Imitation Q-learning	39.61
DP	1371.65

Table 5.1 outlines the computational expenses tied to different EMS. Despite the DP EMS demonstrating superior outcomes compared to the suggested technique, its considerable computational demands hinder practical deployment. Within this specific case analysis, the processing duration for the DP EMS is remarkably 34.6 times longer than that required for the imitation Q-learning EMS, rendering the DP approach more suitable as a theoretical benchmark for assessing the efficiency of the advocated imitation Q-learning EMS. According to the investigations conducted in this thesis, the rule-based EMS boasts the fastest result acquisition time; however, it falls short in terms of control effectiveness. In scenarios extending beyond the scope defined by heuristic rules, the system's adaptability is significantly limited. From the analysis conducted earlier, it is evident that both traditional Q-learning and its imitation Q-learning EMS are capable of securing results that closely approach the optimal. Furthermore, imitation Q-learning distinguishes itself by slashing the

computational time required by nearly 53.28% when compared with its traditional Q-learning counterpart in EMS applications.

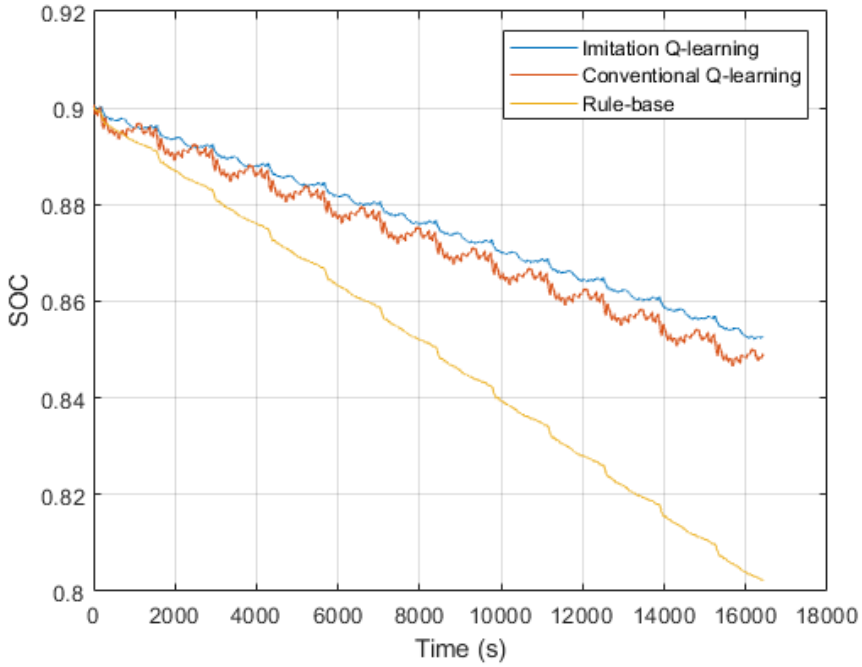

Fig 5.10 LIB SOC trajectories of different EMSs

The comparison of computation times between imitation Q-learning and traditional Q-learning in EMS suggests that these approaches are sufficiently efficient for real-time control applications. To confirm the effectiveness of real-time control, the experiments were performed utilizing the previously mentioned HIL setup. Given the excessive computational demands, DP based EMS was deemed unsuitable for the HIL experimentation. Consequently, a rule-based EMS was implemented in the HIL experiments, serving as a reference point for assessing the manage effect of the discussed method. The outcomes are presented in Fig 5.10, illustrating the SOC patterns for the LIB.

These patterns reveal that the EMSs developed using RL methods outperform the rule-based EMS in terms of energy efficiency. Among the RL-based approaches, the EMS employing imitation Q-learning stands out as the most energy-efficient. Analysis of the SOC trajectories of the LIB confirms that the RL-based EMSs exhibit superior energy efficiency compared to the rule-based system. Each EMS began with an initial SOC of 0.9, with the rule-based EMS concluding at an SOC of 0.8023, marking it as the least efficient among the EMSs evaluated. The SOC for the EMS utilizing imitation Q-learning reached 0.8527 in the end, marking an enhancement of 6.28% over the performance of the rule-based EMS. Meanwhile, the traditional Q-learning approach recorded a final value of 0.8488, exhibiting an improvement of 5.48% when contrasted with the rule-based EMS. Consequently, the EMS based on imitation Q-learning emerges as the most energy-efficient among the discussed EMS setups.

5.4 Conclusion

In Chapter 5, I further evolved the Q-learning based EMS for EVs by incorporating imitation learning to decrease the time-consuming training process inherent in traditional RL. Acknowledging the impracticality of lengthy training times for RL in real-world applications, the chapter introduces a novel imitation Q-learning technique that leverages expert demonstrations for an expedited learning process, significantly reducing the needed iterations. The refined EMS sets initial Q-values based on heuristic understanding, guiding the system to favor SC use during peak power demands, thus alleviating rapid battery wear. Simulation training, leveraging the driving cycle created in prior research, optimizes the collaboration between the battery and SC. This innovative approach demonstrates superior

computational efficiency and control performance compared to both standard Q-learning and traditional EMS strategies. The imitation Q-learning EMS proves capable of reducing battery degradation and improving energy efficiency, achieving these outcomes with a notably reduced computational demand. The system's strategic use of SC to manage peak loads effectively extends battery life. This advancement contributes to the practical implementation of real-time EMS for EVs, enhancing the sustainability and efficiency of electric vehicle technologies.

CHAPTER 6

APPLICATION OF DEEP REINFORCEMENT LEARNING AND DIGITAL TWIN IN EMS

The previous chapters have proposed a RL based EMS which has achieved excellent control performance and reduced the computational cost. But it is only able to handle the discrete optimal control problem. Thus, this chapter proposes deep reinforcement learning based EMS to solve the continuous optimal control problem, and the digital twin technology is integrated into the system to for the real-time implementation.

6.1 Digital twin enhanced Q-learning EMS

Fig 6.1 illustrates the diagram of the digital twin. The development of the virtual model is grounded on the physical model. This virtual model serves the purpose of simulating the physical system, in addition to offering control over the physical model.

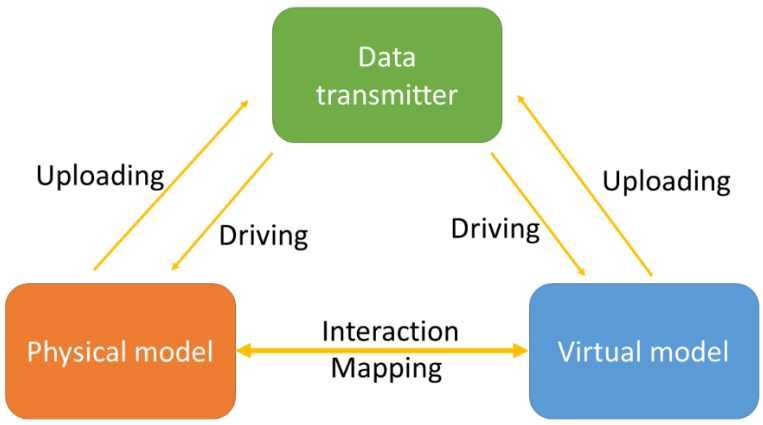

Fig 6.1 Digital twin interaction diagram

Typically, EMS that employ Q-learning are initially trained using conventional driving cycles within simulations [37]. While these cycles aim to mimic real-world traffic scenarios, they only capture a fraction of true traffic behaviors [12]. Consequently, while the initially trained Q-learning agents can offer ideal control tactics under specific conditions, they struggle to manage the unpredictability and sudden changes encountered in actual traffic situations. This dissertation presents an innovative approach by integrating the digital twin concept to upgrade the traditional RL-based EMS in EV. As the EV, equipped with a pre-trained Q-learning based EMS navigates a corresponding digital counterpart operates in parallel within a virtual environment. Data gathered during the drive are transmitted to and from this digital counterpart. This process enables the digital twin to process real-time and historical data, allowing it to refine the EMS. Subsequent enhancements to the Q-learning mechanism are then transferred back to the physical vehicle, thus optimizing the control effectiveness."

the HIL platform is regarded as the basis of the physical model, and presented in Fig 6.2.

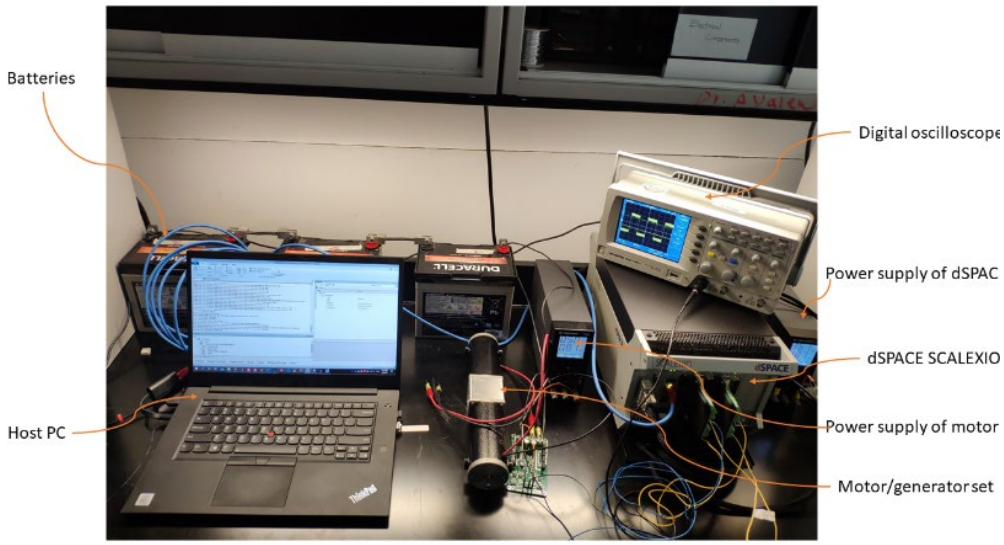

Fig 6.2 HIL platform

A DRL agent, once fully trained, is implemented within the dSPACE SCALEXIO system. Information regarding the drive cycle is managed by the primary computer and is forwarded to the SCALEXIO system. Subsequently, the vehicle's model processes this data to determine the required torque. Upon receiving the current operational status, the DRL agent makes a decision on the appropriate action to take. This action is then communicated to the controller that oversees the motor-generator unit, thus enabling the control over both the motor and the generator. The computational setup for the simulation environment is equipped with an Intel(R) Core(TM) i7-9750H CPU operating at 2.60 GHz, an NVIDIA GeForce RTX 2060 graphics card, and 16.0 GB of RAM memory.

Through the suggested EV driving cycle within the simulation setup, the traditional Q-learning method undergoes training. The training process's reward is illustrated in Fig 6.3. Based on the trajectory of the reward, it is observed that the traditional Q-learning-based

EMS reaches convergence following 600 iterations. Fig 6.4 displays the power output generated by the electric drive system. During the phase of vehicle acceleration, the SC delivers peak power to meet the substantial demand for energy, yet its capacity for energy storage is limited over extended periods. Consequently, the Q-learning EMS leverages the battery, utilizing it as a sustained source of power. In contrast, during deceleration, the SC effectively captures a considerable amount of negative power through regenerative braking. Should the SC reach its capacity, the excess regenerated power is then directed to recharge the battery. This mechanism demonstrates that the SC plays a crucial role in mitigating the intensity of battery charge and discharge cycles, thereby contributing to a reduction in battery aging.

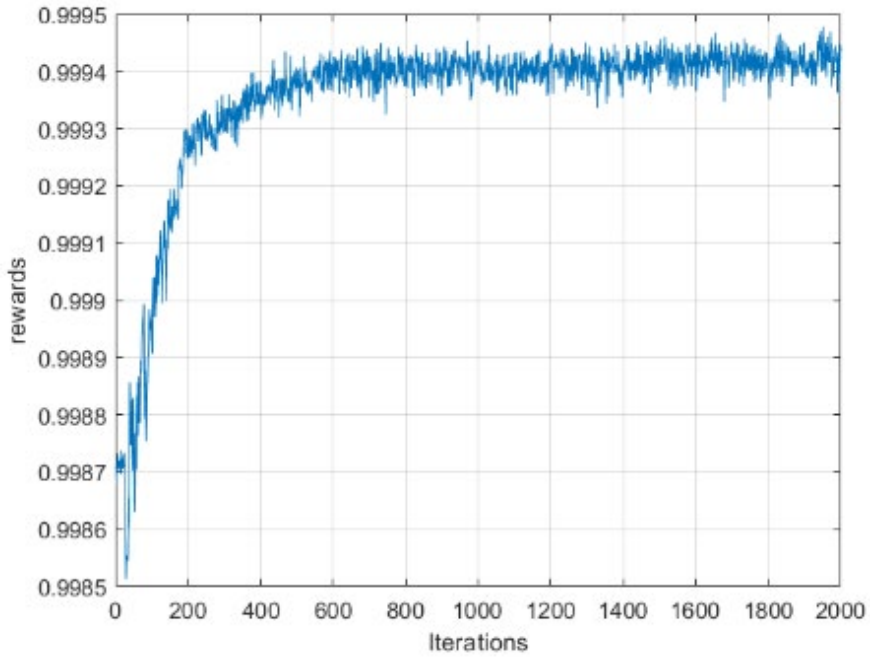

Fig 6.3 Reward trajectory with iterations

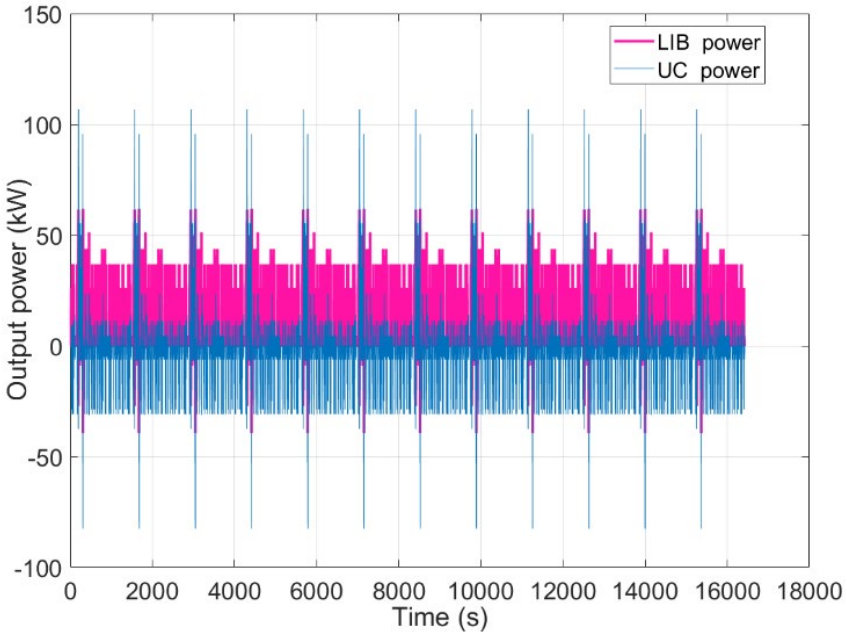

Fig 6.4 Output of electric drive system

While traditional Q-learning EMS have approached near-optimal outcomes within predetermined driving cycles during pre-training, their efficacy diminishes when applied to actual driving scenarios. To address this shortfall, the introduction of a digital twin-augmented Q-learning EMS presents a viable solution. This method bridges the gap by mirroring real-world driving conditions within a digital twin framework, thereby facilitating updates to the Q-learning policy to better reflect these conditions. When subjected to varying driving cycles, the energy efficiency benefits from the enhancements applied to the Q-learning EMS. Comparative SOC trajectories for the battery, managed via Q-learning both with and absent the digital twin intervention, are illustrated in Fig 6.5.

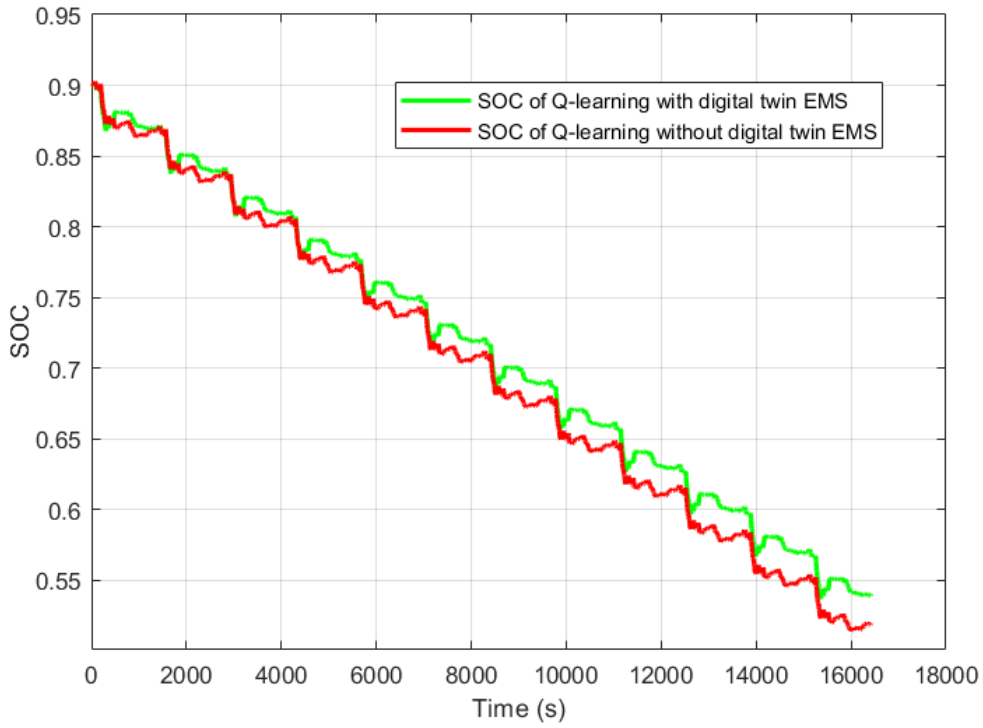

Fig 6.5 SOC trajectories

The utilization of a digital twin approach enhances the adaptability of Q-learning algorithms to evolving environmental conditions. Within the electric drive system, energy stored in the SC represents a minimal fraction; thus, the battery SOC serves as the primary metric for assessing stored energy for both the battery and SC. When governed by the improved Q-learning EMS, the electric drive system's battery exhibits a 4.36% higher charge retention compared to systems managed by traditional Q-learning EMS techniques.

The degradation of battery capacity is also mitigated. As demonstrated in the battery wear-and-tear model, the loss in battery capacity is depicted in Fig 6.6. LIB managed by the standard Q-learning EMS experience more pronounced degradation compared to those managed by the advanced Q-learning EMS. Moreover, the rate of degradation under the

conventional Q-learning EMS accelerates over time. In comparison, the advanced Q-learning EMS demonstrates the ability to maintain a more stable rate of battery degradation over time. By leveraging a digital twin model, this sophisticated Q-learning approach succeeds in mitigating the extent of lithium-ion battery capacity fade.

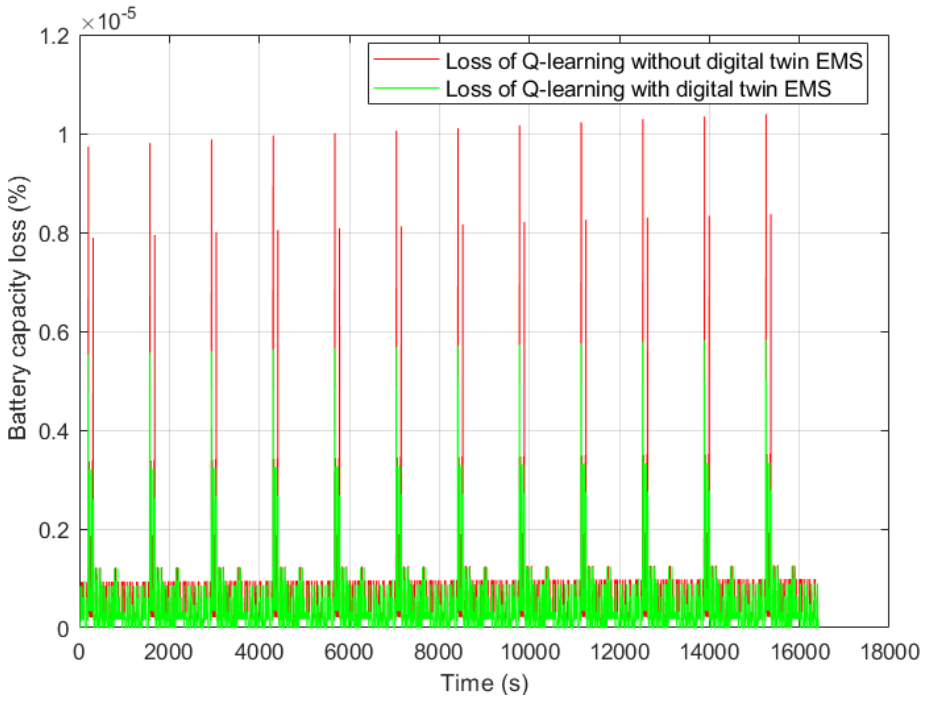

Fig 6.6 Battery capacity loss comparison

6.2 Deep reinforcement learning based EMS

6.2.1 Deep Q-networks

A DQN comprises a multilayered neural network that processes an initial state s, generating an action values vector $Q(s, a; \theta)$, with θ representing the network's parameters. Illustrated in Fig 6.7 is the foundational architecture for DQN-driven DRL algorithms, inclusive of a replay buffer along with two distinct networks: the evaluation network and

the target network. Interactions between the agent and the EV environment lead to the storage of transactional data $T_t = (s_t, a_t, r_t, s_{t+1})$ in the replay buffer $B_t = \{T_1, T_2, \dots, T_t\}$. In the process, mini-batches are randomly selected from the replay buffer. The evaluation network then retrieves a mini-batch to compute the state-action value, denoted as $Q(s, a; \theta_i)$. Concurrently, the target network utilizes data from the same mini-batch to create a target Q value, denoted y_i^{DQN}. The difference between these two neural network outputs informs the design of the loss function. This loss function is crucial for updating the network's parameters. The formula employed for calculating and refining the loss function during each iteration, i, is as follows:

$$L_i(\theta_i) = E_{(s,a,r,s') \sim U(D)} \left[\left(y_i^{DQN} - Q(s, a; \theta_i) \right)^2 \right], \tag{6.1}$$

with

$$y_i^{DQN} = r + \gamma \max_{a'} Q(s', a'; \theta^-), \tag{6.2}$$

where θ^- symbolizes the parameters for a distinct, unchanging target [89]. These parameters remain constant over multiple iterations and are periodically refreshed with the parameters from the evaluation network [90].

Utilizing experience replay enhances data utilization by allowing samples to be reused across several updates. Moreover, it contributes to lowering variability because uniformly drawing samples from the replay buffer decreases the interdependence of the samples applied during the update process. Furthermore, experience replay has evolved, with

iterations such as Prioritized Experience Replay [91], depicted in Fig 6.7, and Hindsight Experience Replay [92]. These iterations find application in varying contexts and offer improvements over the traditional method of experience replay.

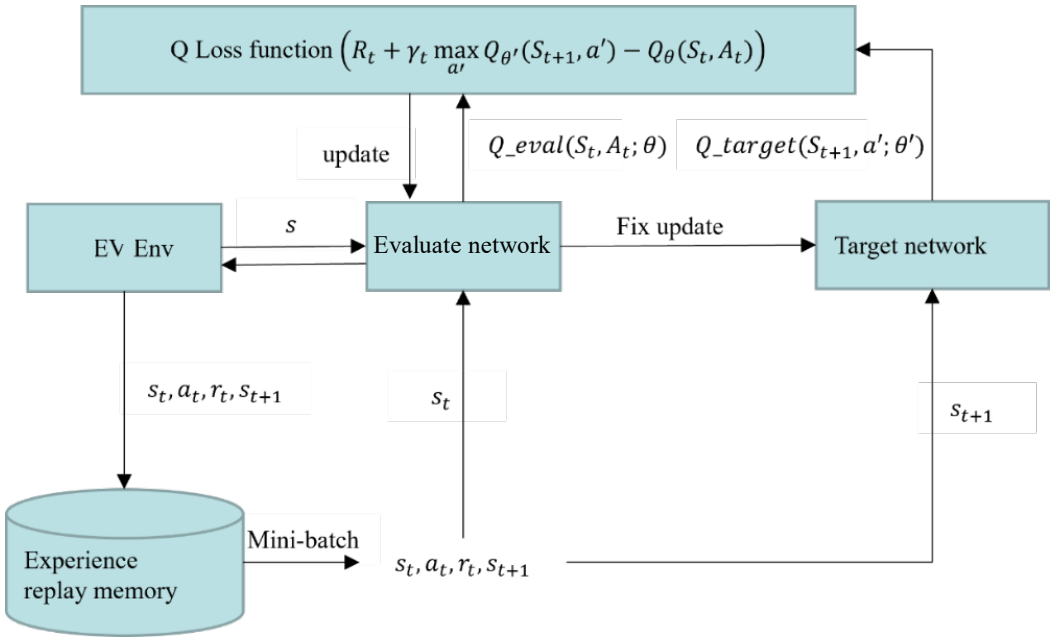

Fig 6.7 Architecture of DQN-based Algorithm.

To address the issue of overestimation encountered in DQN [93], Double Deep Q-Networks (DDQN) employs a strategy where the action, determined through the evaluation network, serves as the input for the target network. Consequently, the target Q value is determined based on this input, ensuring a more accurate estimation.

$$y_i^{DDQN} = r + \gamma Q(s', arg\max_{a'} Q(s', a'; \theta_i); \theta^-). \tag{6.3}$$

The diversity between DDQN and DQN is that different target networks y_i^{DDQN} and y_i^{DQN} are used [90]. As shown in Fig 6.7 and Fig 6.8, this utilization of separate target networks results in varying loss functions.

6.2.2 Rainbow Deep Q-networks

Rainbow integrates the DQN algorithm with six enhancements aimed at overcoming its constraints and enhancing performance [94]. These enhancements include double Q-learning, prioritized experience replay, multi-step learning, dueling architecture, distributional reinforcement learning, and noisy networks. Fig 6.8 illustrates a diagram that represents how Rainbow functions. As indicated in the diagram, Rainbow adopts a prioritized experience replay mechanism, differing from the conventional experience replay buffer, for storing and sampling experiences from the EV setting to train the neural network. This approach ensures more efficient learning by focusing on more significant experiences. Rainbow employs a unique dueling network architecture at its core. It leverages the structure of double-Q networks along with a multi-step targeting strategy to compute the multi-step targets Q value y_i^{DDQN}. DDQN performs better than DQN since the prioritized experience replay (PER) is used in the DDQN [91]. The key advancement offered by PER lies in its ability to boost the likelihood of experiencing highly-anticipated outcomes following rewards tied to TD error. This enhancement has the potential to streamline the training process and elevate the precision of eventual outcomes.

As illustrated in Fig 6.8, Rainbow introduces an adaptation in its learning approach through the employment of multi-step targets, diverging from the traditional single-step target method. Rather than solely relying on a singular step for reward accumulation

followed by bootstrap from the subsequent step, the multi-step learning approach leverages the outcomes of the forthcoming n steps [95]. This technique calculates the n-steps return as follows:

$$R_t^{(n)} = \sum_{k=0}^{n-1} \gamma_t^{(k)} R_{t+k+1} \tag{6.4}$$

Utilizing the multi-step target approach, the loss function for Rainbow is described through the equation presented in reference [95],

$$\left(R_t^{(n)} + \gamma_t^{(n)} Q_{\theta'}\left(S_{t+n}, \underset{a'}{argmax}\, Q_\theta(S_{t+n}, a')\right) - Q_\theta(S_t, A_t)\right)^2 \tag{6.5}$$

The rate at which learning progresses can be effectively adjusted by manipulating the n parameter within the multi-step learning framework [96]. This adjustment allows for a tailored approach to learning speed, accommodating different learning curves and enhancing the efficiency of the process.

In gaming scenarios requiring a multitude of actions before attaining the initial reward, the inefficacy of employing ϵ-greedy policies becomes evident. Noisy Nets introduce an innovative approach by integrating a noisy linear layer that merges both deterministic and stochastic elements [97].

$$y = (b + Wx) + \left(b_{noisy} \odot \epsilon^b + \left(W_{noisy} \odot \epsilon^w\right)x\right) \tag{6.6}$$

When random variables, designated as ϵ^b and ϵ^w, are implemented within the network's architecture, they are combined through element-wise multiplication ⊙. This

novel technique substitutes the conventional linear relationship, expressed as $y = b + Wx$. Progressively, the network acquires the capability to diminish the impact of this introduced noise, albeit at varying degrees across distinct regions of the state space. This method facilitates exploration that is conditioned by the state while enabling a mechanism for self-adjustment over time.

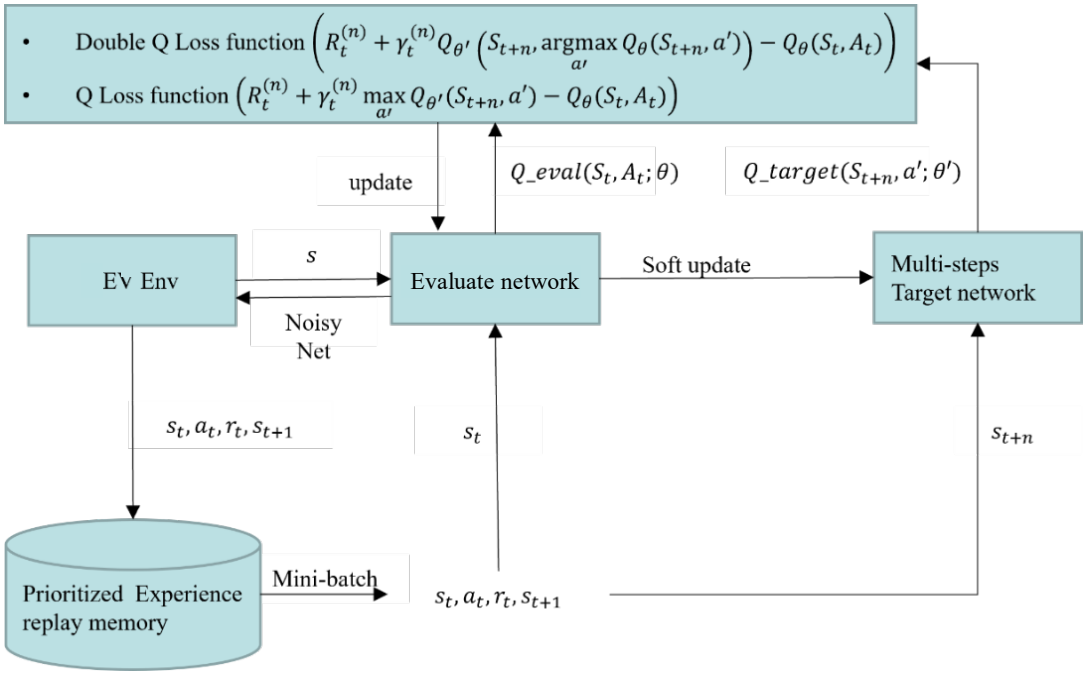

Fig 6.8 Architecture of Rainbow Algorithm.

6.2.3 Deep Deterministic Policy Gradient

Fig 6.9 illustrates that the DDPG framework is composed of dual neural network structures alongside an experience replay mechanism. These neural networks are identified as the Actor and Critic networks, with each one further divided into an online network that directly interacts with input data, and a target network designed for stability during the

learning process [98]. In operation, the Actor network engages with the environment for EVs, logging interactions—specifically, states s_t, actions a_t, rewards r_t, and subsequent states s_{t+1} into the experience replay. This replay mechanism then selects a random subset of these interactions, forming mini-batches (s_i, a_i, r_i, s_{i+1}), which are processed by both the Actor and Critic networks. Notably, the Critic's target network is tasked with forecasting the future reward y_i, utilizing the action $(\mu'(s_{i+1}))$ determined by the Actor's target network [98].

$$y_i = r_i + \gamma Q'\big(s_{i+1}, \mu'(s_{i+1})\big) \tag{6.7}$$

In this framework, γ represents the discount factor. The terms Q' and μ' denote the Critic and Actor target networks, respectively. Given these definitions, one can calculate the loss experienced by the Critic using the equation below, as detailed in [98].

$$L = \frac{1}{N}\sum\nolimits_i \big(y_i - Q(s_i, a_i)\big)^2 \tag{6.8}$$

where N represents the total number of mini-batches. Utilizing feedback from the Critic network, the Actor's policy gets refined through the application of the sampled policy gradient, as outlined in equation [98],

$$\nabla_{\theta^\mu} J = \frac{1}{N}\sum\nolimits_i [\nabla_a Q(s_i, \mu(s_i)) \nabla_{\theta^\mu} \mu(s_i|\theta^\mu)] \tag{6.9}$$

where θ^μ represents the parameters of the online Actor network. To enhance learning stability, the target network undergoes gradual updates using a small hyperparameter ε,

$$\theta' \leftarrow \varepsilon\theta + (1-\varepsilon)\theta' \tag{6.10}$$

$$\theta^{\mu'} \leftarrow \varepsilon\theta^{\mu} + (1-\varepsilon)\theta^{\mu'} \tag{6.11}$$

where θ' denotes the parameters of the Critic target network, θ represents the parameters of the Critic online network, and $\theta^{\mu'}$ refers to the parameters of the Actor target network.

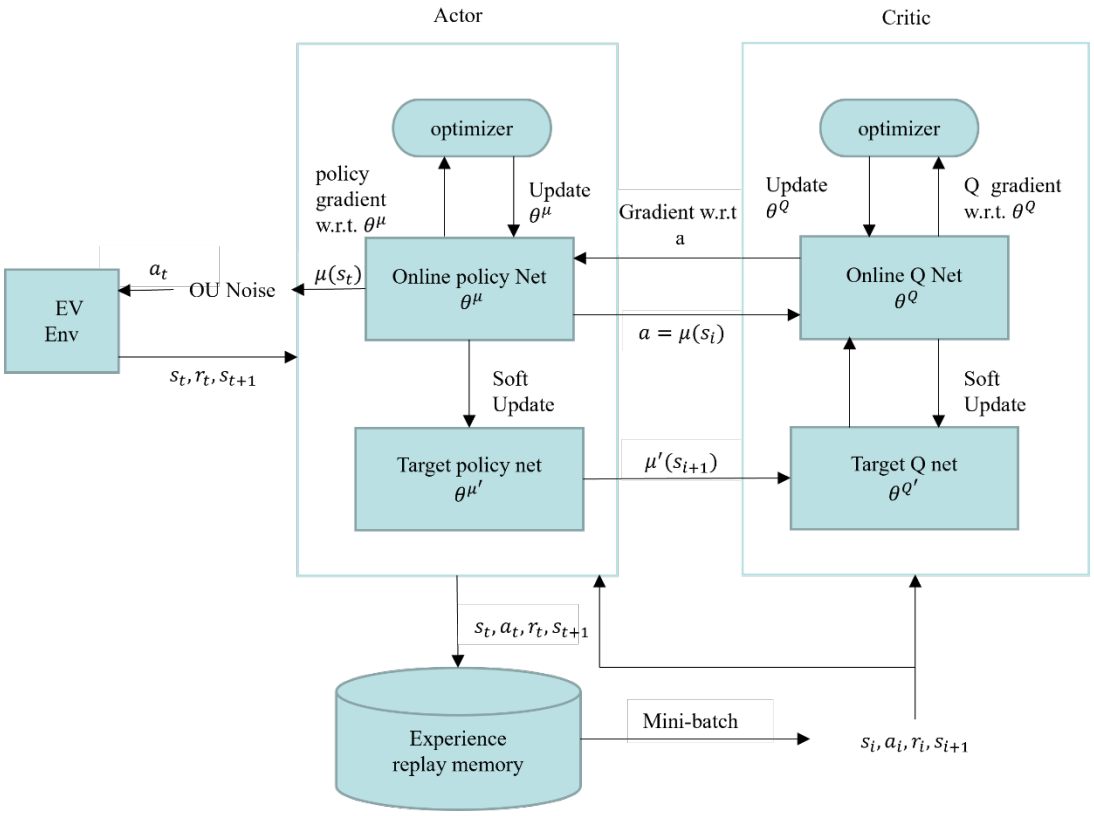

Fig 6.9 Architecture of DDPG

6.2.4 Twin-delayed DDPG

TD3, an extension of DDPG, addresses approximation error shortcomings and enhances stability [99]. It integrates continuous Double Q learning, Policy Gradient, and Actor-Critic techniques. Unlike DDPG, TD3 features dual Critic networks comprising two online

networks (Q_1, Q_2) and two target networks (Q'_1, Q'_2). The expected target value is determined as described by the equation [99],

$$y_1 = r + \gamma Q'_1(s', \bar{a}) \tag{6.12}$$

$$y_2 = r + \gamma Q'_2(s', \bar{a}) \tag{6.13}$$

where μ' represents the Actor target network. To mitigate overestimation issues, TD3 calculates the expected target value by selecting the minimum of two estimates, as illustrated in the equation provided [99],

$$y = r + \gamma \min_{i=1.2} Q'_i(s', \bar{a}) \tag{6.14}$$

A crucial enhancement in TD3 involves target policy smoothing, which functions as a controller to address overfitting issues in Q value calculation by introducing noise ε, as depicted in equation [99],

$$\bar{a} = \mu'(s') + \varepsilon, \quad \varepsilon \sim \mathrm{clip}(\mathrm{N}(0, \sigma), -c, c) \tag{6.15}$$

When provided with the target value, the Critic network's loss function can be determined using the equations presented in [99],

$$L_1 = \frac{1}{N} \sum_i \left(y_i - Q_1(s_i, a_i)\right)^2 \tag{6.16}$$

$$L_2 = \frac{1}{N}\sum_i \left(y_i - Q_2(s_i, a_i)\right)^2$$

(6.17)

where L_1, L_2 denote the loss functions of the first and second Critic networks respectively. Following the backpropagation of losses and the subsequent update of the two Critic networks, the Actor network is then updated through gradient ascent using the Critic network output, as described below.

$$\nabla_{\theta^\mu} J = \frac{1}{N}\sum_i [\nabla_a Q_1(s_i, \mu(s_i)) \nabla_{\theta^\mu} \mu(s_i|\theta^\mu)]$$

(6.18)

where θ^μ represents the parameter of the online Actor network. TD3 also implements a soft update approach, which is described by the equations provided in reference [99],

$$\theta'_i \leftarrow \varepsilon\theta_i + (1-\varepsilon)\theta'_i$$

(6.19)

$$\theta^{\mu\prime} \leftarrow \varepsilon\theta^\mu + (1-\varepsilon)\theta^{\mu\prime}$$

(6.20)

where θ'_i denotes the parameter associated with the Critic target network, while θ_i refers to the Critic online network, and $\theta^{\mu\prime}$ stands for the parameter of the Actor target network.

6.2.5 Trust Region Policy Optimization (TRPO)

In TRPO, the core concept involves adjusting the policy through a nuanced balance between exploration and constraint. The objective is to steer the policy in a direction that maximizes progress while adhering to specified limits to maintain proximity with the previous policy. This balance is enforced through a constraint typically quantified by KL divergence. The agent actively engages with the EV environment, gathering sequential data

points termed trajectories $D = \{s_0, a_0, \dots, a_{T-1}, s_T\}$ using a policy network. Subsequently, the Critic network calculates the advantage value based on the trajectories, as described by the equation [100],

$$\hat{A}_t = -V(s_t) + r_t + \gamma r_{t+1} + \cdots + \gamma^{T-t+1} r_{T-1} + \gamma^{T-t} V(s_T) \tag{6.21}$$

The time index t within the range [0, T], and considering V as the current value function, the policy is adjusted using the advantage value, following the equations provided in [100],

$$\theta_k = \underset{\theta}{\operatorname{argmax}}\, \hat{E}_t \left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)} \hat{A}_t \right] \tag{6.22}$$

$$\text{s.t. } \hat{E}_t \left[D_{KL}(\pi_\theta(\cdot\,|s_t) || \pi_{\theta_{old}}(\cdot\,|s_t)) \right] \leq \delta \tag{6.23}$$

The policy $\pi_{\theta_{old}}$ refers to the previous policy prior to the update, while D_{KL} represents the KL divergence. To simplify the theoretical TRPO method, second-order Taylor series approximations are utilized for faster learning. The corresponding loss and KL-divergence are detailed in the equations presented in reference [100],

$$L(\theta) \approx g^T(\theta - \theta_k) \tag{6.24}$$

$$\bar{D}_{KL}(\theta||\theta_k) \approx \frac{1}{2}(\theta - \theta_k)^T H(\theta - \theta_k), \quad \bar{D}_{KL} \leq \delta \tag{6.25}$$

where g represents the policy gradient, and H signifies the correlation between the policy and parameter θ. The resolution of this quadratic equation is elaborated in reference [100]:

$$\theta_{k+1} = \theta_k + \alpha^j \sqrt{\frac{2\delta}{g^T H^{-1} g}} H^{-1} g \tag{6.26}$$

where α represents the backtracking coefficient, and j denotes the smallest non-negative integer that ensures policy $\pi_{\theta_{k+1}}$ meets the KL-divergence constraint. Instead of directly calculating and storing H^{-1}, the conjugate gradient algorithm is employed to find $x = H^{-1} g$, resulting in the equation for θ_{k+1} as described in reference [100].

$$\theta_{k+1} = \theta_k + \alpha^j \sqrt{\frac{2\delta}{x_k^T H x_k}} x_k \tag{6.27}$$

The MSE is used to update Critic network, given by the following equation [100],

$$\emptyset_{k+1} = \underset{\emptyset}{\operatorname{argmin}} E[V(s_t|\emptyset) - R_t] \tag{6.28}$$

where $\emptyset_{k+1}$ denotes the Critic network parameters.

6.2.6 The Proximal Policy Optimization Algorithm (PPO)

PPO strives to manage the challenge of maximizing policy improvement effectively while avoiding overstepping, which can result in collapse [101]. As presented in Fig 6.10, agents engage with the EV environment to gather trajectories $D = \{s_0, a_0, \ldots, a_{T-1}, s_T\}$ utilizing a policy network. Upon obtaining the trajectories, the Critic network computes the advantage value using the equation provided in reference [101],

$$\hat{A}_t = -V(s_t) + r_t + \gamma r_{t+1} + \cdots + \gamma^{T-t+1} r_{T-1} + \gamma^{T-t} V(s_T)$$

(6.29)

In the given time frame indexed by t within [0, T], and with V representing the current value function, the policy's loss function is determined based on the advantage value, as shown in the equation referenced in [101],

$$L(\theta) = \hat{E}_t \left[\min \left(\frac{\pi_\theta(a_t, s_t)}{\pi_{\theta_{old}}(a_t, s_t)} \hat{A}_t, clip \left(\frac{\pi_\theta(a_t, s_t)}{\pi_{\theta_{old}}(a_t, s_t)}, 1 - \epsilon, 1 + \epsilon \right) \hat{A}_t \right) \right]$$

(6.30)

where π_θ and $\pi_{\theta_{old}}$ represent the new and old policies, respectively. The hyperparameter ε is utilized to constrain the probability ratio within the range $[1 - \epsilon, 1 + \epsilon]$, thereby managing the extent to which the new policy diverges from the old policy. This approach introduces a first-order method for trust region optimization, preventing the agent from excessively favoring positive value actions or hastily dismissing negative value actions. Subsequently, the policy is adjusted through stochastic gradient ascent, as outlined in equation [101].

$$\theta_{k+1} = \underset{\theta}{\operatorname{argmax}} E[L(\theta)]$$

(6.31)

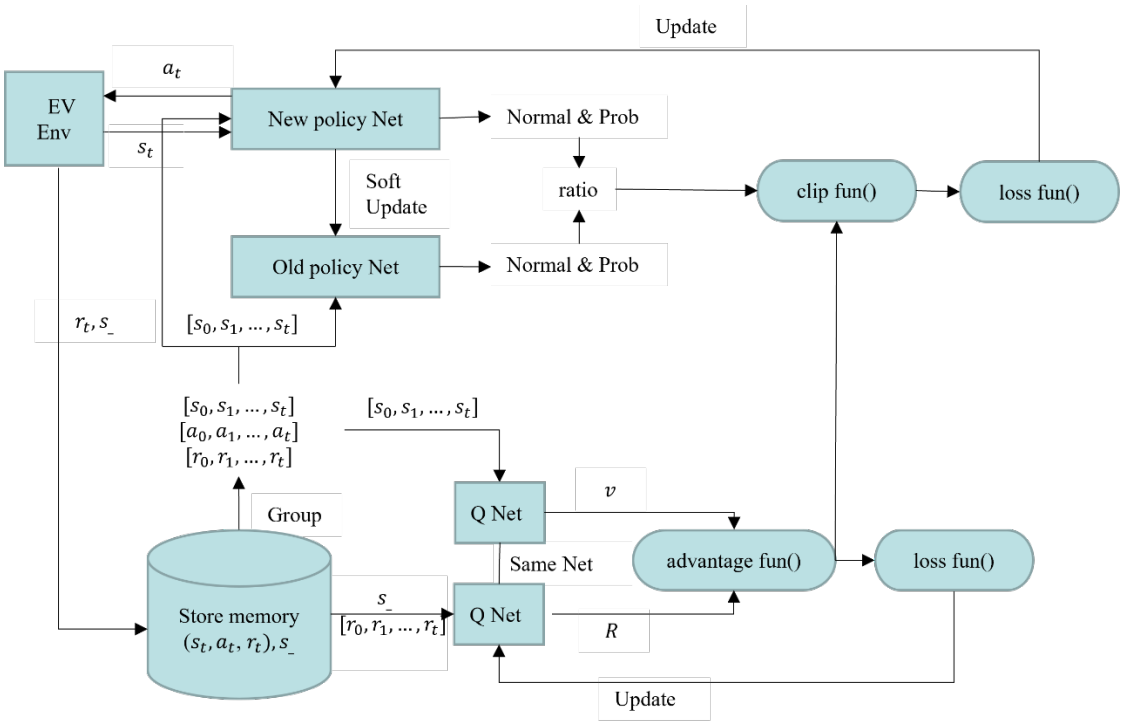

Fig 6.10 Architecture of PPO

The MSE is used to update critic network, shown the following equation [101],

$$\emptyset_{k+1} = \underset{\emptyset}{\operatorname{argmin}} E[V(s_t|\emptyset) - R_t] \tag{6.32}$$

where $\emptyset_{k+1}$ stands for the Critic network parameters.

6.2.7 Deep reinforcement learning EMS comparison

The trajectories of the cumulative reward metric is depicted for four different DRL algorithms operating in a continuous action space environment, as presented in Fig 6.11. To mitigate fluctuations, the plotted curves have been smoothed using a 10-point moving average filter. The computation time parameter refers to the total duration required for completing the training process. The convergence reward value indicates the average

reward attained by the agent upon reaching a stable convergence point, while the convergence episode specifies the particular training episode at which this convergence condition was satisfied. Convergence time, on the other hand, indicates the time taken by the agent to reach the convergence episode. The findings indicate that all four continuous DRL-based EMS systems achieve convergence within 50 episodes. The training process begins with different initial conditions due to the stochastic nature of the parameters. A comparison of the results suggests that DDPG and TD3 exhibit faster initial learning rates compared to other algorithms. In contrast, PPO demonstrates a slower increase in rewards during training. Analysis of Table 6.1 reveals that PPO achieves the highest convergence reward among the DRL algorithms studied. However, the convergence speed of PPO lags behind other algorithms. Specifically, DDPG stands out for its rapid convergence compared to the other algorithms. DDPG requires significantly less time to reach convergence than TD3, TRPO, and PPO, with reductions of 33.7%, 46.1%, and 65.2%, respectively. Moreover, PPO's convergence reward surpasses TRPO, TD3, and DDPG by small margins of 0.18%, 0.26%, and 0.15% respectively. Moreover, an observation of the performance after convergence indicates that DDPG exhibits more stable learning curves than TD3, PPO, and TRPO due to their flatness. Notably, when considering the same total number of training episodes, DDPG outperforms TRPO, TD3, and PPO in terms of time efficiency, achieving time savings of 10.1%, 10.2%, and 7.9% respectively. Additionally, TD3 demonstrates slightly higher test rewards compared to TRPO, DDPG, and PPO, with improvements of 0.08%, 0.3%, and 0.78% respectively.

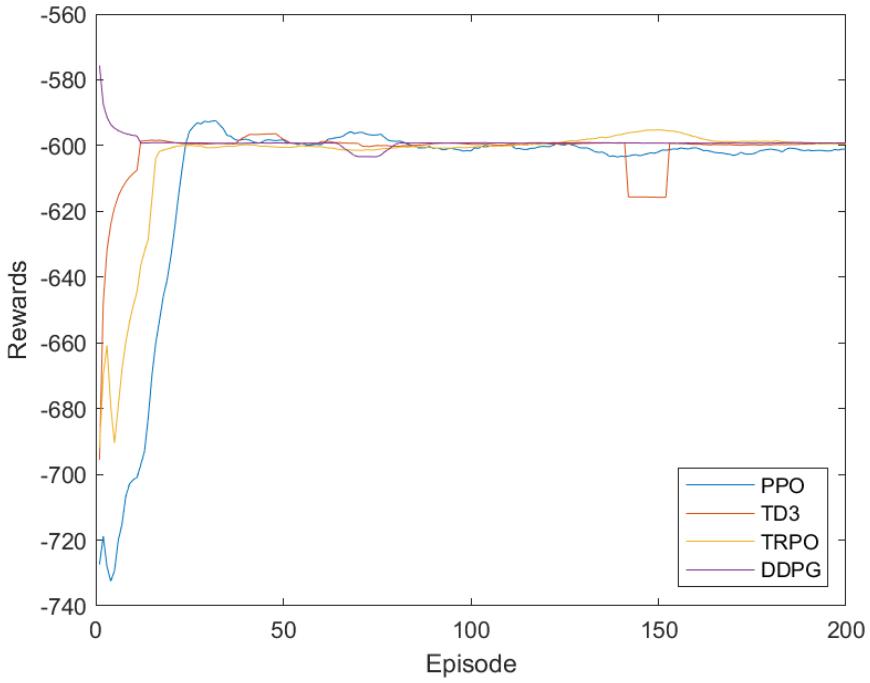

Fig 6.11 Comparison of DRL algorithms in continuous action space

Table 6.1 Parameters of DRL algorithm in continuous action space

Agent	Computation time (s)	Convergence reward	Convergence episode	Convergence time (s)	Final test reward
DDPG	722	598.5	43	141	601.5
TD3	740	599.6	25	91	597.3
TRPO	741	600.1	21	74	596.8
PPO	665	599.4	16	49	599.1

6.3 Digital twin-enhanced DRL-based EMS

RL-based EMS have made advancements in optimizing the control of EVs' energy usage, there remain unresolved challenges in this area. The utilization of table-based RL

algorithms restricts the ability to manage discrete action and state spaces within EMS. In the preceding section, it was observed that the Q-learning-based EMS operates with discrete state and action variables, suggesting limitations in the range of available states and actions. The effectiveness of the system is influenced by the discretization process applied to environmental states and action spaces. As the dimensions of state and action spaces increase, the complexity of training also rises, leading to the challenge known as the 'curse of dimensionality' [102]. However, in various fields like robotics and energy management, discretization is not an ideal approach due to its adverse effects on solution quality. Additionally, fine discretization demands substantial memory and computing resources. Unlike traditional Q-learning, the DRL method employs multi-layer neural networks to estimate the Q-values, offering significant advancements in handling continuous state spaces more effectively. According to the comparison of different DRL method in Subsection5.3, this dissertation adopts the DDPG to develop the DRL-based EMS for the EV equipped with HESS and the diagram of digital twin-enhanced DDPG-EMS is shown in Fig 6.12.

Within the DDPG framework, the actor network interacts with a virtual digital twin model representing the electric vehicle's dynamics. The interactions between the actor and the model are recorded and stored in an experience replay buffer. During training, this buffer randomly samples small batches of prior experiences, which are then fed as inputs to the actor and critic networks. The critic target network evaluates the expected long-term return based on the actions proposed by the actor target network.

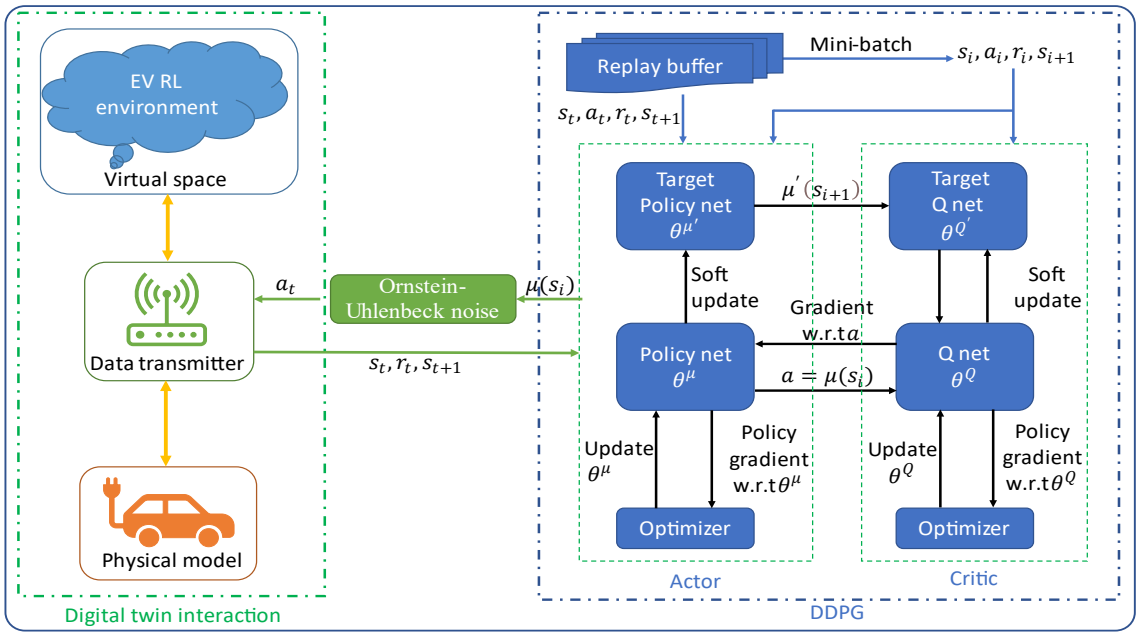

Fig 6.12 Digital twin-enhanced DDPG-EMS diagram

DDPG is a model-free, policy-based reinforcement learning algorithm used to solve continuous control tasks. DDPG is an off-policy algorithm that combines concepts from DQN and policy gradient methods. It utilizes deep neural networks to approximate the policy (actor) and the action-value function (critic). Since the DDPG uses an actor-critic architecture, it has better convergence performance than Q-learning. The critic in DDPG learns an action-value function, which helps to reduce the variance in the policy gradient estimation. This can bring in more stable learning and improved convergence compared to Q-learning, particularly in environments with high variance. When combined with the digital twin technology, the DDPG is more scalable than Q-learning for problems with large or continuous action spaces. The digital twin model has higher fidelity, and the virtual space contains more information than the conventional RL environment, such as real-time

traffic data. That means the action spaces grows rapidly, which dramatically increases Q-learning's computational complexity. But DDPG avoids this issue and can handle more complex problems with larger action spaces by directly learning a policy. Besides, the DDPG allows for more efficient exploration in continuous action spaces compared to Q-learning. In Q-learning, exploration is typically achieved through ε-greedy exploration, which can be inefficient in continuous action spaces. DDPG uses noise added to the output of the actor-network for exploration, such as Ornstein-Uhlenbeck noise in this dissertation, which can lead to more effective exploration strategies in continuous action spaces.

As shown in Algorithm 1, a high-fidelity digital twin of the physical model is mapped in the virtual space, accurately representing the powertrain components, battery system, driving conditions, and environmental factors that affect energy usage and driving range. The action of the DDPG agent is the ratio of the energy distribution between LIB and SC. The velocity and torque demand still be chosen as the states. But it is different from the conventional DDPG-based EMS, whose speed and torque demand are designed on the fixed driving cycle. The states in the proposed method are collected from the physical model to improve the adaptability of the EMS in real-world traffic situations.

Algorithm 1 Digital twin-enhanced DDPG-EMS

```
Map the physical model to establish digital twin model
Import data from physical model to initialize the digital twin environment (states, actions, rewards)
Initialize actor network (policy network) with random weights
Initialize target actor network with the same weights as the actor network
Initialize critic network (Q-value network) with random weights
Initialize target critic network with the same weights as the critic network
Initialize replay buffer (memory for experience storage)
for episode in range(max_episodes) do
    Reset digital twin environment to its initial state
    Get initial state from the digital twin environment
    Reset noise process
    episode_reward ← 0
    for t in range(max_timesteps) do
        Select action using the actor network and add exploration noise
        Perform action in digital twin environment
        Observe next state, reward, and done flag (termination)
        Store experience (state, action, reward, next_state, done) in the replay buffer
        if len(replay_buffer) ≥ batch_size then
            Sample a minibatch of experiences from the replay buffer
            Update critic network using target actor and target critic networks
            Update actor network using critic network's gradients
            Update target networks with a soft update (polyak averaging)
        end if
        episode_reward ← episode_reward + reward
        state ← next_state
        if done then
            break
        end if
    end for
    if episode % evaluation_interval == 0 then
        Evaluate actor network performance on digital twin environment
        Save network weights if performance is improved
    end if
end for
Discard the original EMS and deploy the trained DDPG EMS to physical model
```

Besides, to fully exploit the scalability of DDPG, the battery SOC is introduced into the state space. This helps the EMS absorb more information from the environment to improve control accuracy. The reward function for the DDPG is the same as the Q-learning, which

considers energy consumption and battery degradation. Once the DDPG agent is trained, evaluate its performance by simulating the control policy in the digital twin environment. The obtained results are also compared to existing control strategies or benchmarks. If the performance is unsatisfactory, refine the DDPG algorithm or reward function and retrain the agent. After achieving satisfactory performance in the digital twin environment, the learned control policy is deployed in the physical model.

6.4 Results of digital twin enhanced DRL-base EMS

In the pre-train phase, the proposed EV driving cycle is adopted to initialize the DRL agent at the starting point. The results of the pre-trained DDPG-based EMS are show in Fig 6.13.

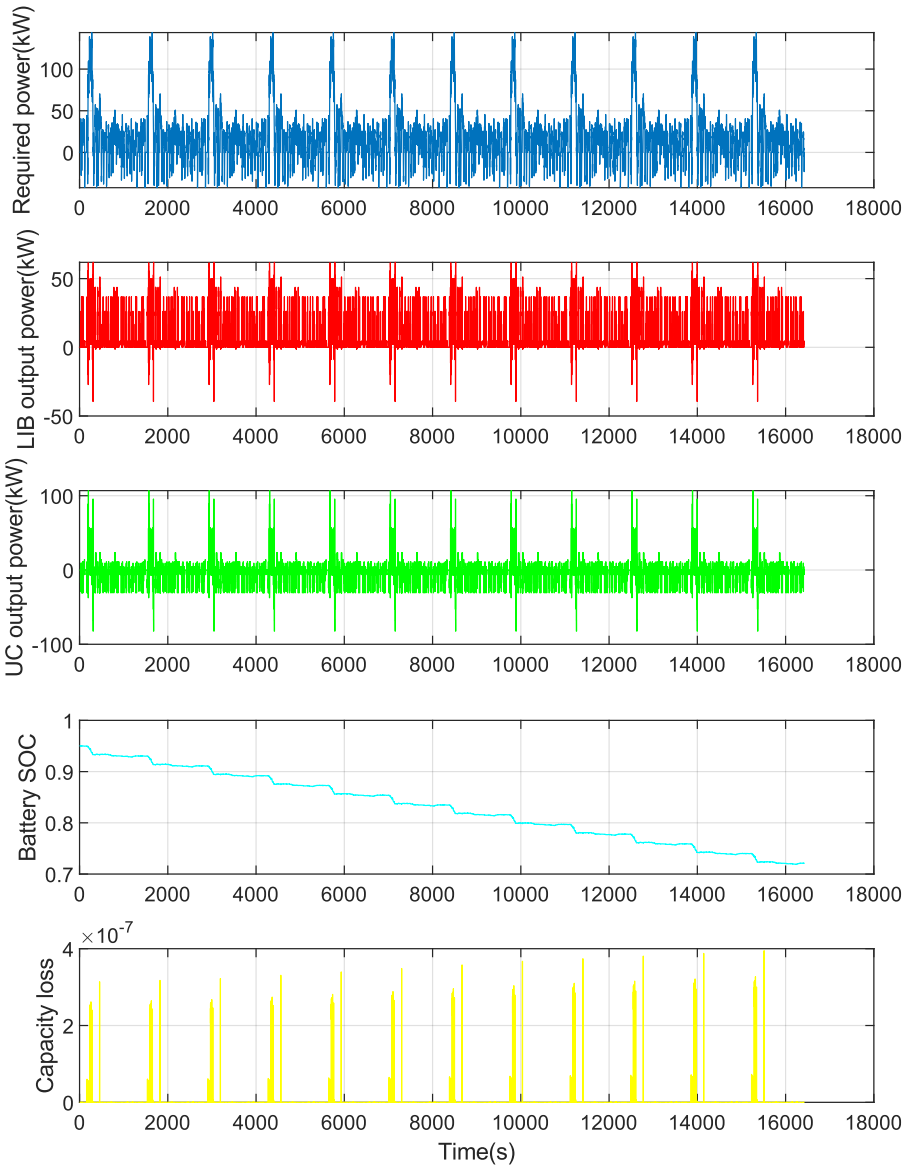

Fig 6.13 Results of conventional DDPG-based EMS

During the vehicle acceleration phase, the SC delivers high power output to meet the increased power demand. However, the SC has limitations in storing energy for long-term operations. To address this issue, a Q-learning EMS leverages the battery as a continual power source. During deceleration, the SC captures excess power from regenerative braking. When the SC reaches full capacity, any surplus regenerated energy is directed to charge the battery. This process demonstrates the SC's ability to effectively manage peak power demands on the battery, thereby reducing battery degradation.

Fig 6.13 illustrates the effectiveness of the DDPG-based EMS in maximizing the utilization of the SC to lessen the strain on the LiB. During the vehicle's acceleration phase, the SC delivers high-power output to meet the substantial power demand. However, due to its limited energy storage capacity for prolonged operations, the DDPG-based EMS optimally utilizes the battery as a continuous power source. Moreover, during deceleration, the SC efficiently captures excess power from regenerative braking. Any surplus regenerated energy beyond the SC's capacity is directed towards charging the battery. This dynamic demonstrates the SC's ability to effectively manage the peaks of battery charging and discharging power, thereby mitigating battery degradation.

The traditional DDPG-based EMS has shown near-optimal performance for a specific predefined driving cycle, its operational efficacy tends to diminish in real-world driving scenarios. To enhance adaptability and responsiveness, a digital twin is incorporated into the EMS framework to leverage current traffic conditions. This dissertation employs the Worldwide Harmonized Light Vehicles Test Procedure (WLTP) to replicate real-world

driving environments. A comparative analysis between the conventional DDPG-based EMS and the newly proposed approach is outlined below:

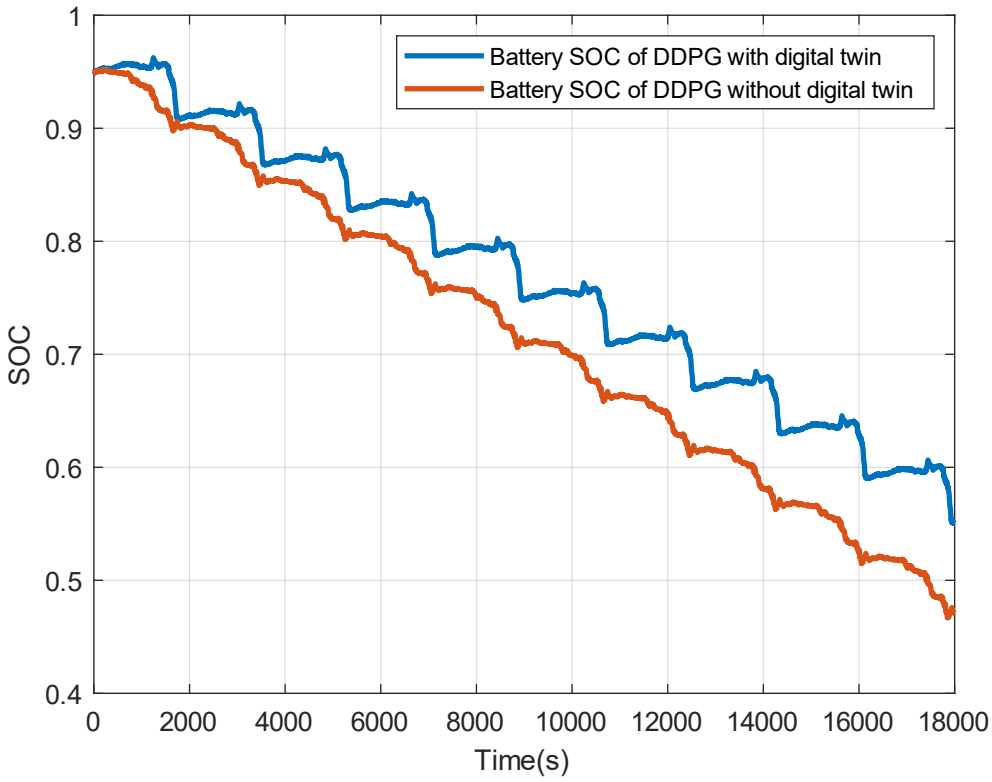

Fig 6.14 SOC trajectories

Fig 6.14 illustrates the SOC of conventional DDPG-based EMS and the digital twin-enhanced DDPG-based EMS. The evaluation reveals that the suggested approach demonstrates superior performance compared to the traditional DDPG-based EMS. Based on the final SOC value, the digital twin-augmented DDPG-based EMS exhibits 17.08% higher energy efficiency than the standalone DDPG-based EMS. Furthermore, in terms of battery degradation, the proposed method exhibits enhanced control capabilities, effectively mitigating battery capacity loss.

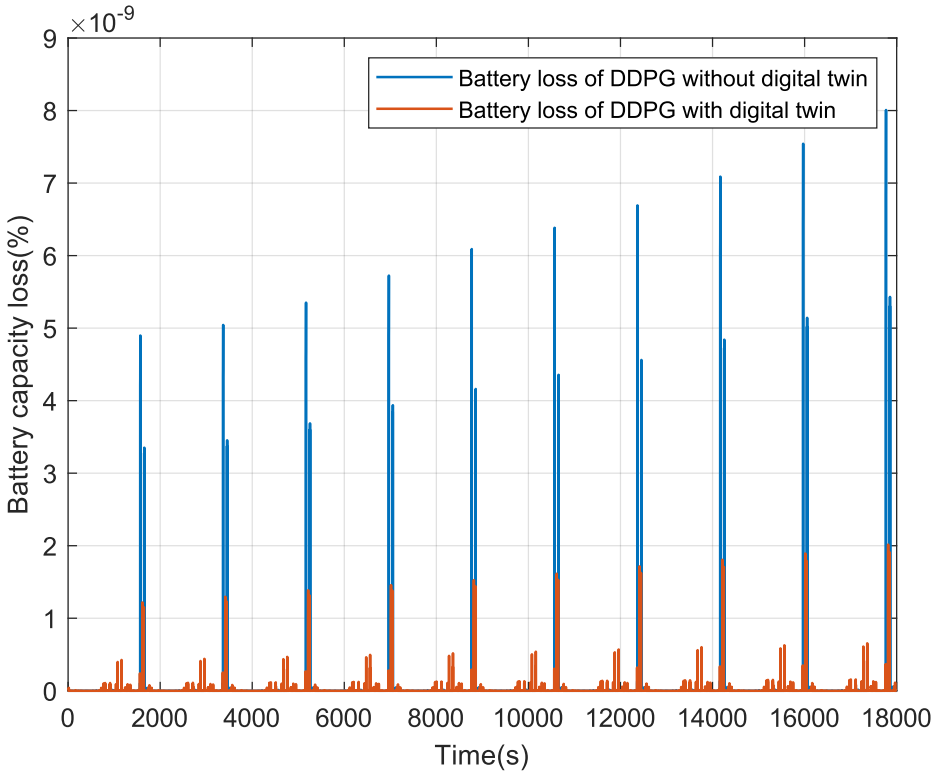

Fig 6.15 Battery capacity loss trajectories

Fig 6.15 indicates the battery aging during driving in the WLTP. The total capacity losses in the simulation of DDPG-based EMS are $3.41 * 10^{-7}\%$ and $1.32 * 10^{-7}\%$ without and with the digital twin, respectively. The results show that the digital twin-enhanced DDPG-based EMS has better adaptability and control performance. The Agents of both methods are pre-trained through UDDS. When deployed in the physical model under the WLTP, the proposed method achieves higher energy efficiency and lower battery degradation.

6.5 Comparative study with other EMSs

In this dissertation, EV HIL simulations incorporating HESS are conducted. The effectiveness of various EMS, namely the digital twin-enhanced DDPG-based EMS, digital twin-enhanced Q-learning-based EMS, Q-learning-based EMS, and rule-based EMS, is analyzed in a unified setting to evaluate effectiveness of the control strategy and its computational requirements. The performance of the LIB and SC within the HESS under different EMS strategies is visualized in Fig 6.16.

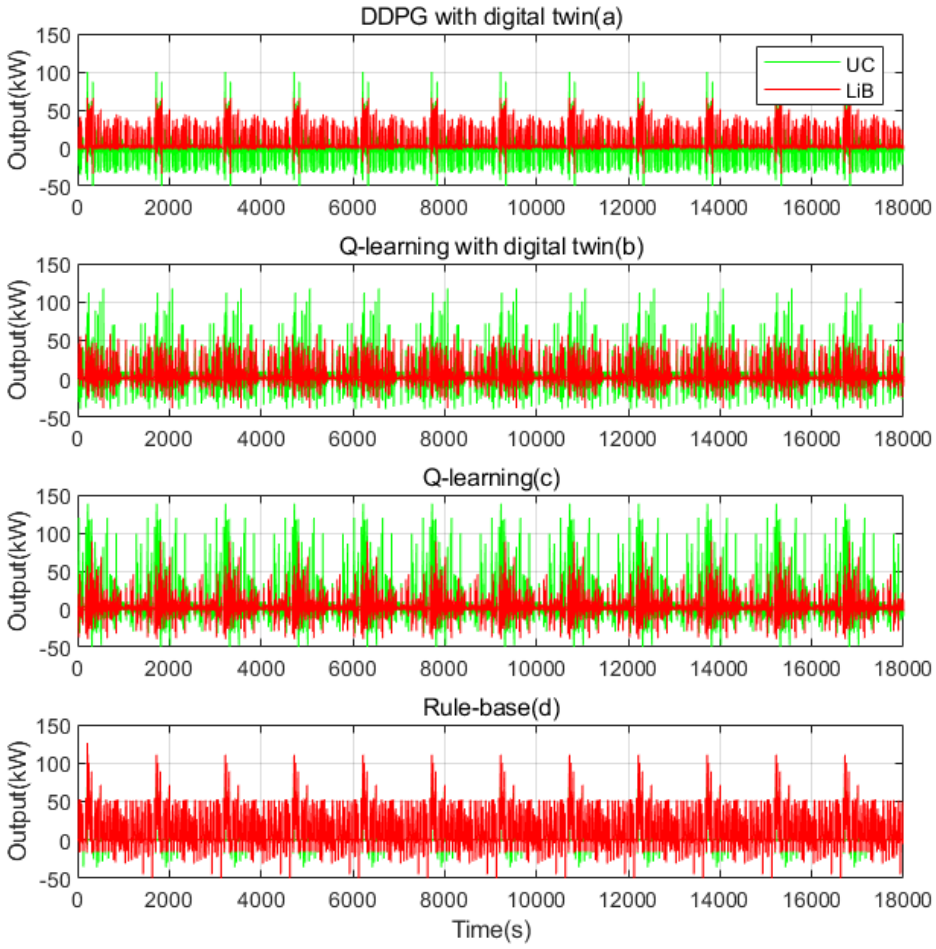

Fig 6.16 EMS comparison

The comparison of output power displays the performance of the SC and LIB outputs during the simulated driving cycle across under the control four different EMSs. By the results, the digital twin-enhanced DDPG-based EMS have the best manage effect since the developed method takes advantage of the digital twin to infuse real-time traffic information. In Fig 6.16(a), the SC exhibits high peak power for charging and discharging, while the LIB sustains continuous power to support vehicle operation and mitigate battery degradation. The rule-based EMS, developed based on expert knowledge, may have limitations in addressing diverse application scenarios. Fig 6.16(d) illustrates that the SC primarily participates in charging, with the LIB providing peak charging power. Moreover, the LIB is responsible for providing continuous driving power, making the current rule-based approach inefficient in preserving battery health. Comparing the digital twin-enhanced Q-learning and traditional Q-learning methods, both demonstrate effective control performance. These RL based EMSs leverage the HESS by assigning the SC to manage peak power demands, thus reducing LIB degradation. However, Fig 6.16 (b) and (c) highlight a subtle contrast between the digital twin-enhanced Q-learning EMS and the traditional Q-learning approach. In the former, the SC handles all peak power requirements, while the latter assigns peak charging tasks to the LIB. Fig 6.17 illustrates the LIB degradation trends to assess the effectiveness of our proposed control strategy.

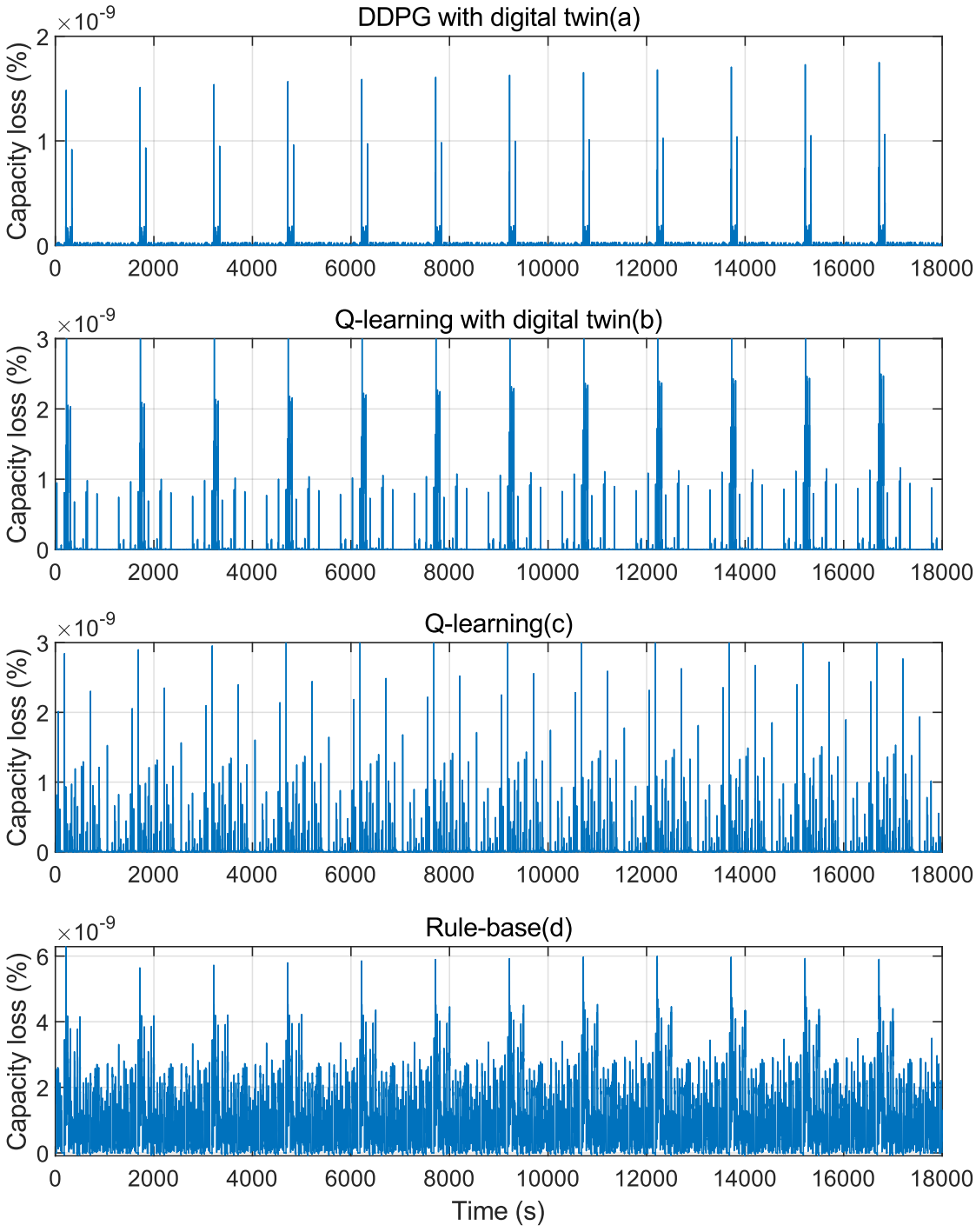

Fig 6.17 Battery capacity loss comparison

Cumulative battery capacity decrease after completing a single simulation using distinct control approaches such as rule-based, traditional Q-learning, digital twin-enhanced Q-learning, and digital twin-enhanced DDPG-based EMS, is measured at low rates of $1.57 * 10^{-6}$ %, $6.48 * 10^{-7}$ %, $5.81 * 10^{-7}$ %, and $1.32 * 10^{-7}$ %, respectively. The digital twin-enhanced DDPG-based EMS demonstrates superior performance in mitigating LIB degradation, showing a reduction of 67.99% compared to the digital twin-enhanced Q-learning-based EMS. Upon examining Fig 6.17((a) and (b), the LIB degradation patterns exhibited by the digital twin-augmented Q-learning-based EMS and the digital twin-enhanced DDPG-based EMS are comparable. However, the values observed in the digital twin-enhanced Q-learning-based EMS are significantly higher. This observation is further corroborated by Fig 6.17(a) and (d). Under the control of the digital twin-enhanced DDPG-based EMS, the LIB primarily operates in discharge mode to power the vehicle, with infrequent charging cycles. Conversely, the digital twin-enhanced Q-learning-based EMS engages the LIB in both charging and discharging operations, resulting in increased cumulative battery aging compared to the digital twin-enhanced DDPG-based EMS. Notably, the LIB degradation witnessed in the traditional Q-learning EMS surpasses that of the digital twin-enhanced Q-learning-based EMS by 11.53%. Furthermore, the rule-based EMS fails to fully utilize the benefits of HESS, resulting in the most significant degradation of the LIB among all EMS configurations. From the results of battery degradation, the digital twin technology enables the RL-based EMS to exploit the potential for the optimal solution fully. Both DDPG and the Q-learning methods

outperform the conventional counterpart. The digital twin-enhanced DDPG-based EMS also achieves a better result for alleviating battery degradation.

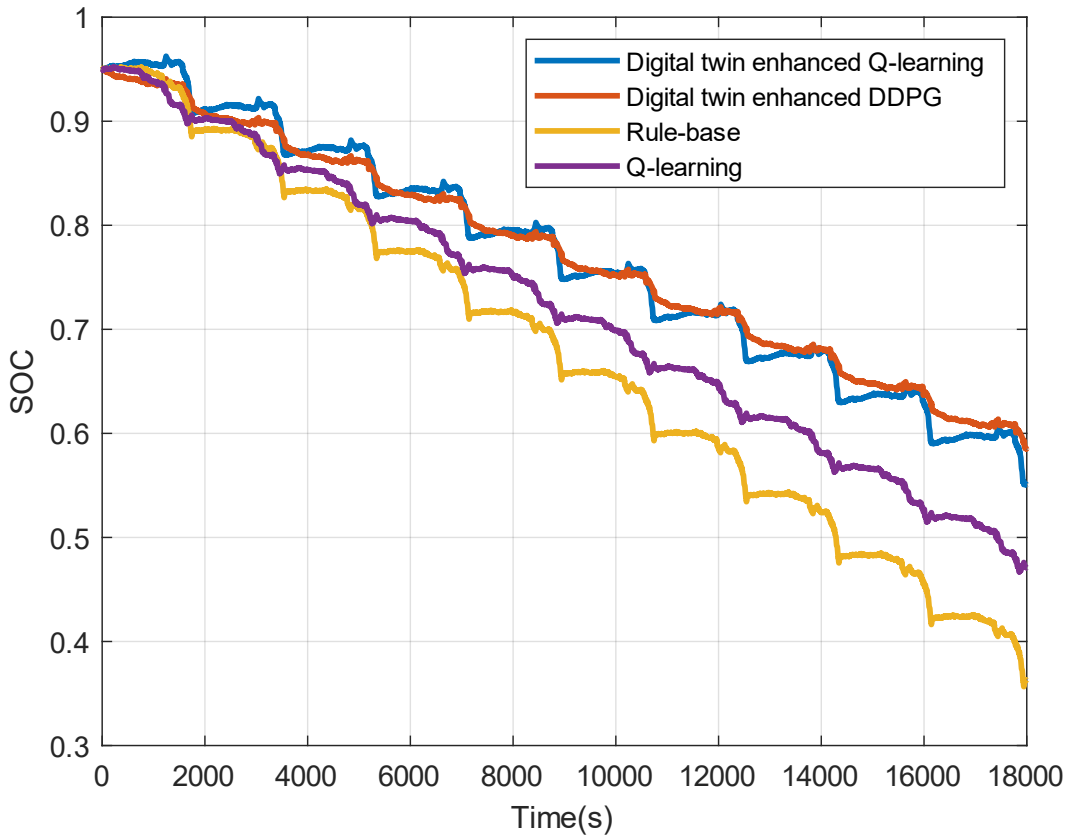

Fig 6.18 SOC trajectories comparison

The energy reserve of the SC is relatively small in the electric drive system, making the SOC of the battery a key metric to assess the stored energy for both the battery and SC. Fig 6.18 illustrates the comparison of SOC trajectories. It can be observed from the LIB SOC trajectories that the energy efficiency of RL-based EMS configurations surpasses that of the rule-based EMS. All EMS start with an SOC of 0.95, and the rule-based EMS exhibits the lowest final SOC of 0.3627 among the tested EMS configurations. The digital

twin-enhanced DDPG-based EMS has the highest energy efficiency, whose final SOC value is 0.5851. The digital twin-enhanced Q-learning-based EMS also achieves good control performance, and its final SOC value is 0.5509, which is lower than the proposed method by 6.21%. The final SOC of conventional Q-learning-based EMS is 0.4708, which is much lower than the Q-learning with the digital twin. This phenomenon is the same as we discussed in the last subsection, and the comparison between the digital twin-enhanced DDPG and conventional DDPG shows that the digital twin can enhance the performance of RL-based EMS.

6.6 Conclusion

In Chapter 6, the challenges of table-based RL in managing continuous control tasks are addressed by introducing DRL algorithms such as DQN, DDPG, and PPO, which excel in continuous spaces. The chapter explores how these methods enhance computational efficiency and control performance within EV EMS. Furthermore, the integration of digital twin technology represents a paradigm shift for traditional Q-learning EMS, allowing it to adapt to real-time traffic conditions for improved performance. The digital twin, equipped with high-fidelity simulations, accurately mirrors the EV's powertrain and environmental interactions, facilitating superior energy management decisions. Simulation results demonstrate that DRL combined with a digital twin outperforms conventional EMS in both energy efficiency and battery longevity. This chapter proposes a robust framework for real-time EMS application in EVs, where the synergy of DRL and digital twin technology offers substantial improvements in adaptability and optimization of energy use.

CHAPTER 7

CONCLUSIONS AND FUTURE WORK

This dissertation delineates the critical research problem—creating a robust energy management system using DRL tailored for battery and SC EVs, and the necessity of an EV-specific driving cycle for system training. It addresses the predominant challenges in this domain, including the construction of realistic driving cycles, real-time application of RL algorithms, battery longevity, and computational efficiency. This sets the stage for the dissertation to explore these avenues, promising significant strides in the field of EV energy management.

7.1 Conclusions

In CHAPTER three, a representative urban driving cycle for EV is established, which is crucial for accurately evaluating EV performance and energy management systems. The chapter outlines a methodical approach for this construction, starting with the strategic selection of urban routes to capture a comprehensive array of driving conditions. It then details the meticulous collection and processing of vehicular operation data to ensure a robust data set reflective of true urban driving patterns. Leveraging sophisticated data analysis techniques, the chapter introduces the use of principal component analysis to effectively reduce data dimensionality, simplifying the complex data set while preserving its most critical characteristics. This is complemented by a hybrid classification algorithm combining SOM and SVM to categorize the driving conditions, ensuring the driving cycle's relevance to diverse urban scenarios. The core of the chapter is the innovative application

of Markov chains and Monte Carlo simulations, which are employed to synthesize a driving cycle that not only mirrors the stochastic nature of urban driving but also maintains statistical fidelity to the collected data. The validation of this driving cycle is thorough, utilizing metrics such as relative error and speed-acceleration probability distributions, which confirm that the cycle accurately reflects real-world driving conditions. This Chapter encapsulates the significance of the proposed driving cycle, emphasizing its potential to bridge the gap between theoretical research and practical application in EV energy management. It underscores the meticulousness of the approach and the careful consideration of statistical representativeness, offering a solid foundation for subsequent efforts in optimizing EV energy consumption and battery performance. This chapter not only contributes a new methodological framework to the field but also provides a validated tool for EV technology.

In CHAPTER four, a Q-learning based EMS for EVs is provided, positioning it as a strategic alternative to the more conventional rule-based and optimization-based systems. The chapter initiates by providing an overview, elucidating the complex dynamics of EVs and the structure of their propulsion systems. This understanding is crucial for the formulation of an efficient EMS. It places particular emphasis on the modeling of a Li-S battery, spotlighting this emerging technology as a cost-effective and energy-dense alternative to the widely used LIBs. The dissertation then navigates through the various configurations of the HESS, dissecting the advantages and challenges associated with each setup. Crucially, the chapter introduces a Q-learning based EMS, which stands out for its ability to autonomously learn and optimize energy distribution between the EV's battery

and supercapacitor without the need for pre-defined rules or models. This EMS is tested through simulation, using the proposed EV driving cycle developed in Chapter 2, to assess its effectiveness in reducing energy consumption and mitigating battery degradation. This Chapter emphasizes the potential of the Q-learning based EMS as an innovation in the realm of EV energy management. The results of the simulations present a compelling case for the Q-learning approach, which not only meets but in certain respects exceeds the performance of traditional rule-based methods, achieving near-optimal results. Notably, the approach does not rely on anticipating driving conditions, but instead focuses on making real-time energy management decisions that optimize the current state of the EV's energy storage systems. This feature exemplifies the system's robustness and reliability, underscoring the practicality of the Q-learning based EMS in real-world applications. Through these advancements, the chapter contributes an efficient and sustainable EV technologies, fostering a more resilient energy framework for future mobility solutions.

In CHAPTER five, an Imitation Q-learning based EMS for EV is built upon the previous chapters by enhancing EMS through imitation learning, designed to reduce the training time cost. The chapter begins by discussing the limitations of traditional RL methods, particularly the extensive training time due to numerous iterations, which is not practical for real-world applications. To address this, it introduces an innovative imitation Q-learning approach that uses expert demonstrations to kickstart the learning process, thereby significantly reducing the number of iterations required. The chapter then details the design of the EMS using the imitation Q-learning method, where the initial Q values are set based on heuristic rules rather than starting from zero, as is typical in conventional

Q-learning. This strategic adjustment allows the system to prioritize SC usage during high power demand scenarios to prevent rapid battery degradation. The EMS is further refined through simulation-based training using a previously developed EV driving cycle, optimizing the power sharing between the battery and SC. This Chapter 4 presents an imitation Q-learning based EMS that shows improvements in computational efficiency and control performance over traditional Q-learning, rule-based EMSs. The approach is validated through a series of simulations and comparisons, demonstrating that it can effectively reduce battery degradation and enhance energy efficiency, with the added advantage of requiring less computational time. The SC is effectively utilized to handle peak power demands, thereby preserving the battery's lifespan. This method presents a real-time applicable EMS for EVs, contributing a crucial piece to the puzzle of sustainable and efficient electric vehicle technology.

CHAPTER six of the dissertation marks an advancement in the realm of EMS for EVs by integrating DRL with digital twin technology. This chapter begins by examining the limitations of table-based reinforcement learning algorithms, which struggle with continuous control tasks due to their discrete nature. To overcome this, the dissertation proposes the application of DRL techniques, such as DQN, DDPG, and PPO, which are capable of handling continuous control problems. The chapter details the benefits of each method and their implementation in the context of EV energy management, focusing on the enhancement of computational efficiency and control performance. The digital twin technology is introduced as a breakthrough to enhance the conventional Q-learning EMS, enabling the system to adapt to real-time traffic conditions, thereby improving the

adaptability and accuracy of the EMS. The digital twin-enhanced DRL-based EMS utilizes high-fidelity simulation models to replicate the EV's powertrain, battery system, and environmental factors, leading to better decision-making in energy management. The simulations show that the digital twin-enhanced EMS not only improves energy efficiency but also significantly reduces battery degradation compared to the conventional Q-learning and rule-based EMS. This chapter presents a comprehensive solution for the real-time application of EMS in EVs. It demonstrates that the combination of DRL and digital twin technology significantly optimizes energy consumption and battery health in a more efficient and adaptive manner than previous approaches. The chapter provides evidence that the proposed approach can lead to more sustainable and cost-effective EV operations, contributing to the advancement of intelligent energy management systems in the automotive industry.

In summary, this dissertation presents an investigation into EV optimal energy management through some techniques. It commences with an overview of the challenges and motivations for improved EV energy systems. Successive chapters enhance the framework for an EMS, initially by developing an EV driving cycle, then by implementing a Q-learning based EMS, which is further refined through imitation learning to improve efficiency and practicality. The work culminates in the integration of DRL with digital twin technology, yielding an adaptive and efficient EMS. The obtained results provide an EMS framework that promises to elevate the sustainability and performance of future EVs, marking a stride in the advancement of eco-friendly transportation technology.

7.2 Future work

While this dissertation has made progresses, several areas for future research and improvement have been identified. Firstly, in the driving data collection, the impact of temperature and other weather conditions was not considered, potentially affecting vehicle performance consistency and introducing errors into the data. To enhance real-world applicability, future studies will incorporate weather-condition variables in the creation of driving cycles, addressing this limitation. This research will focus on investigating and validating more sophisticated models for the battery and supercapacitor, thereby bolstering the accuracy of the energy component in EMS testing and design. Extensive validation procedures will be employed to ensure the chosen models accurately reflect real-world performance. In the analysis of Li–S batteries, this study focused on performance aspects without delving into the low power density drawback. Future work will conduct additional experiments covering a broader spectrum of battery temperatures and current scenarios. Moreover, the application of bilateral SEI Li–S batteries in battery EVs, a critical aspect for future markets, will be thoroughly investigated. The integration of a heuristic rule-based EMS has bolstered the proposed method's effectiveness. Future studies will explore alternative optimization-based EMSs, such as DP and Pontryagin's Minimum Principle, to enhance the efficiency of imitation Q-learning. The computational model will be expanded to include the SOC of battery and SC. Another aspect to be addressed in future work is the consideration of passenger variations in the Q-learning EMS. Changes in passenger numbers significantly affect the vehicle's total weight, influencing energy consumption under similar traffic conditions. Additionally, the development and application of a more

comprehensive full-scale HIL platform will be a focal point in future research, ensuring a more accurate representation of real-world conditions.

APPENDIX A

PUBLICATIONS DURING PHD STUDY

- Y. Ye, J. Zhang, S. Pilla, A. M. Rao, and B. Xu, "Application of a new type of lithium-sulfur battery and reinforcement learning in plug-in hybrid electric vehicle energy management," *Journal of Energy Storage*, vol. 59, p. 106546, Mar. 2023, doi: 10.1016/j.est.2022.106546.
- Y. Ye, H. Wang, B. Xu, and J. Zhang, "An imitation learning-based energy management strategy for electric vehicles considering battery aging," *Energy*, vol. 283, p. 128537, Nov. 2023, doi: 10.1016/j.energy.2023.128537
- Y. Ye, X. Zhao, and J. Zhang, "Driving cycle electrification and comparison," *Transportation Research Part D: Transport and Environment*, vol. 123, p. 103900, Oct. 2023, doi: 10.1016/j.trd.2023.103900.
- Y. Ye and J. Zhang, "A Reconfigurable Battery Topology for Cell Balancing," SAE International, Warrendale, PA, SAE Technical Paper 2023-01–1683, Oct. 2023. doi: 10.4271/2023-01-1683.
- Y. Ye, B. Xu, J. Zhang, B. Lawler, and B. Ayalew, "Reinforcement Learning-Based Energy Management System Enhancement Using Digital Twin for Electric Vehicles," in *2022 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Nov. 2022, pp. 1–6. doi: 10.1109/VPPC55846.2022.10003411.
- Y. Ye, J. Zhang, S. Pilla, and A. M. Rao, "Application of a new type of Lithium-Sulfur battery in plug-in hybrid electric vehicle cruise control," Volume 9: Sustainable

Energy Solutions for Changing the World: Part I, preprint, Mar. 2021. doi: 10.46855/energy-proceedings-7109.

- H. Wang, Y. Ye, J. Zhang, and B. Xu, "A comparative study of 13 deep reinforcement learning based energy management methods for a hybrid electric vehicle," *Energy*, vol. 266, p. 126497, Mar. 2023, doi: 10.1016/j.energy.2022.126497.
- X. Zhao, Y. Ye, J. Ma, P. Shi, and H. Chen, "Construction of electric vehicle driving cycle for studying electric vehicle energy consumption and equivalent emissions," *Environ Sci Pollut Res*, vol. 27, no. 30, pp. 37395–37409, Oct. 2020, doi: 10.1007/s11356-020-09094-4.
- H. Wang, Z. Arjmandzadeh, Y. Ye, J. Zhang, and B. Xu, "FlexNet: A warm start method for deep reinforcement learning in hybrid electric vehicle energy management applications," *Energy*, vol. 288, p. 129773, Feb. 2024, doi: 10.1016/j.energy.2023.129773.
- H. Wang, Z. Arjmandzadeh, Y. Ye, J. Zhang, and B. Xu, "Automated Expert Knowledge-Based Deep Reinforcement Learning Warm Start via Decision Tree for Hybrid Electric Vehicle Energy Management," *SAE Int. J. Elec. Veh.*, vol. 13, no. 1, Art. no. 14-13-01–0006, Aug. 2023, doi: 10.4271/14-13-01-0006.

BIBLIOGRAPHY

[1] Y. Ye, J. Zhang, S. Pilla, A. M. Rao, and B. Xu, "Application of a new type of lithium-sulfur battery and reinforcement learning in plug-in hybrid electric vehicle energy management," *J. Energy Storage*, vol. 59, p. 106546, Mar. 2023, doi: 10.1016/j.est.2022.106546.

[2] Z. Chen, B. Xia, C. You, and C. C. Mi, "A novel energy management method for series plug-in hybrid electric vehicles," *Appl. Energy*, vol. 145, pp. 172–179, May 2015, doi: 10.1016/j.apenergy.2015.02.004.

[3] Z. Chen, C. C. Mi, B. Xia, and C. You, "Energy management of power-split plug-in hybrid electric vehicles based on simulated annealing and Pontryagin's minimum principle," *J. Power Sources*, vol. 272, pp. 160–168, Dec. 2014, doi: 10.1016/j.jpowsour.2014.08.057.

[4] J. Peng, H. He, and R. Xiong, "Rule based energy management strategy for a series–parallel plug-in hybrid electric bus optimized by dynamic programming," *Appl. Energy*, vol. 185, pp. 1633–1643, Jan. 2017, doi: 10.1016/j.apenergy.2015.12.031.

[5] P. Seers, G. Nachin, and M. Glaus, "Development of two driving cycles for utility vehicles," *Transp. Res. Part Transp. Environ.*, vol. 41, pp. 377–385, Dec. 2015, doi: 10.1016/j.trd.2015.10.013.

[6] J. C. Kurnia, A. P. Sasmito, and T. Shamim, "Performance evaluation of a PEM fuel cell stack with variable inlet flows under simulated driving cycle conditions," *Appl. Energy*, vol. 206, pp. 751–764, Nov. 2017, doi: 10.1016/j.apenergy.2017.08.224.

[7] J. Brady and M. O'Mahony, "Development of a driving cycle to evaluate the energy economy of electric vehicles in urban areas," *Appl. Energy*, vol. 177, pp. 165–178, Sep. 2016, doi: 10.1016/j.apenergy.2016.05.094.

[8] W. T. Hung, H. Y. Tong, C. P. Lee, K. Ha, and L. Y. Pao, "Development of a practical driving cycle construction methodology: A case study in Hong Kong," *Transp. Res. Part Transp. Environ.*, vol. 12, no. 2, pp. 115–128, Mar. 2007, doi: 10.1016/j.trd.2007.01.002.

[9] W. Naranjo Lourido, L. E. Munoz, and J. E. Pereda, "A Methodology to Obtain a Synthetic Driving Cycle through GPS Data for Energy Analysis," in *2015 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Montreal, QC, Canada: IEEE, Oct. 2015, pp. 1–5. doi: 10.1109/VPPC.2015.7352876.

[10] A. Manthiram, "A reflection on lithium-ion battery cathode chemistry," *Nat. Commun.*, vol. 11, no. 1, Art. no. 1, Mar. 2020, doi: 10.1038/s41467-020-15355-0.

[11] A. U. Rahman, I. Ahmad, and A. S. Malik, “Variable structure-based control of fuel cell-supercapacitor-battery based hybrid electric vehicle,” *J. Energy Storage*, vol. 29, p. 101365, Jun. 2020, doi: 10.1016/j.est.2020.101365.

[12] B. Xu *et al.*, “Q-Learning-Based Supervisory Control Adaptability Investigation for Hybrid Electric Vehicles,” *IEEE Trans. Intell. Transp. Syst.*, pp. 1–10, 2021, doi: 10.1109/TITS.2021.3062179.

[13] L. Fan *et al.*, “Simultaneous Suppression of the Dendrite Formation and Shuttle Effect in a Lithium–Sulfur Battery by Bilateral Solid Electrolyte Interface,” *Adv. Sci.*, vol. 5, no. 9, p. 1700934, Sep. 2018, doi: 10.1002/advs.201700934.

[14] J. I. Huertas, J. Díaz, D. Cordero, and K. Cedillo, “A new methodology to determine typical driving cycles for the design of vehicles power trains,” *Int. J. Interact. Des. Manuf. IJIDeM*, vol. 12, no. 1, pp. 319–326, Feb. 2018, doi: 10.1007/s12008-017-0379-y.

[15] H. Achour and A. G. Olabi, “Driving cycle developments and their impacts on energy consumption of transportation,” *J. Clean. Prod.*, vol. 112, pp. 1778–1788, Jan. 2016, doi: 10.1016/j.jclepro.2015.08.007.

[16] H. Y. Tong and W. T. Hung, “A Framework for Developing Driving Cycles with On-Road Driving Data,” *Transp. Rev.*, vol. 30, no. 5, pp. 589–615, Sep. 2010, doi: 10.1080/01441640903286134.

[17] R. Günther, T. Wenzel, M. Wegner, and R. Rettig, “Big data driven dynamic driving cycle development for busses in urban public transportation,” *Transp. Res. Part Transp. Environ.*, vol. 51, pp. 276–289, Mar. 2017, doi: 10.1016/j.trd.2017.01.009.

[18] L. Berzi, M. Delogu, and M. Pierini, “Development of driving cycles for electric vehicles in the context of the city of Florence,” *Transp. Res. Part Transp. Environ.*, vol. 47, pp. 299–322, Aug. 2016, doi: 10.1016/j.trd.2016.05.010.

[19] A. Fotouhi, “Tehran driving cycle development using the k-means clustering method,” *Sci. Iran.*, p. 8, 2013.

[20] P. Yuhui, Z. Yuan, and Y. Huibao, “Development of a representative driving cycle for urban buses based on the K-means cluster method,” *Clust. Comput.*, vol. 22, no. S3, pp. 6871–6880, May 2019, doi: 10.1007/s10586-017-1673-y.

[21] A. Esteves-Booth, T. Muneer, H. Kirby, J. Kubie, and J. Hunter, “The measurement of vehicular driving cycle within the city of Edinburgh,” *Transp. Res. Part Transp. Environ.*, vol. 6, no. 3, pp. 209–220, May 2001, doi: 10.1016/S1361-9209(00)00024-9.

[22] M. Knez, T. Muneer, B. Jereb, and K. Cullinane, “The estimation of a driving cycle for Celje and a comparison to other European cities,” *Sustain. Cities Soc.*, vol. 11, pp. 56–60, Feb. 2014, doi: 10.1016/j.scs.2013.11.010.

[23] M. A. Pouresmaeili, I. Aghayan, and S. A. Taghizadeh, “Development of Mashhad driving cycle for passenger car to model vehicle exhaust emissions calibrated using on-board measurements,” *Sustain. Cities Soc.*, vol. 36, pp. 12–20, Jan. 2018, doi: 10.1016/j.scs.2017.09.034.

[24] U. Galgamuwa, L. Perera, and S. Bandara, “Development of a driving cycle for Colombo, Sri Lanka: an economical approach for developing countries: Development of a Driving Cycle for Colombo, Sri Lanka,” *J. Adv. Transp.*, vol. 50, no. 7, pp. 1520–1530, Nov. 2016, doi: 10.1002/atr.1414.

[25] Z. Jing, G. Wang, S. Zhang, and C. Qiu, “Building Tianjin driving cycle based on linear discriminant analysis,” *Transp. Res. Part Transp. Environ.*, vol. 53, pp. 78–87, Jun. 2017, doi: 10.1016/j.trd.2017.04.005.

[26] R. Smith, S. Shahidinejad, D. Blair, and E. L. Bibeau, “Characterization of urban commuter driving profiles to optimize battery size in light-duty plug-in electric vehicles,” *Transp. Res. Part Transp. Environ.*, vol. 16, no. 3, pp. 218–224, May 2011, doi: 10.1016/j.trd.2010.09.001.

[27] S. H. Kamble, T. V. Mathew, and G. K. Sharma, “Development of real-world driving cycle: Case study of Pune, India,” *Transp. Res. Part Transp. Environ.*, vol. 14, no. 2, pp. 132–140, Mar. 2009, doi: 10.1016/j.trd.2008.11.008.

[28] N. H. Arun, S. Mahesh, G. Ramadurai, and S. M. Shiva Nagendra, “Development of driving cycles for passenger cars and motorcycles in Chennai, India,” *Sustain. Cities Soc.*, vol. 32, pp. 508–512, Jul. 2017, doi: 10.1016/j.scs.2017.05.001.

[29] S.-H. Ho, Y.-D. Wong, and V. W.-C. Chang, “Developing Singapore Driving Cycle for passenger cars to estimate fuel consumption and vehicular emissions,” *Atmos. Environ.*, vol. 97, pp. 353–362, Nov. 2014, doi: 10.1016/j.atmosenv.2014.08.042.

[30] Q. Shi, Y. Zheng, R. Wang, and Y. Li, “The study of a new method of driving cycles construction,” *Procedia Eng.*, vol. 16, pp. 79–87, 2011, doi: 10.1016/j.proeng.2011.08.1055.

[31] Q. Wang, H. Huo, K. He, Z. Yao, and Q. Zhang, “Characterization of vehicle driving patterns and development of driving cycles in Chinese cities,” *Transp. Res. Part Transp. Environ.*, vol. 13, no. 5, pp. 289–297, Jul. 2008, doi: 10.1016/j.trd.2008.03.003.

[32] H. He, J. Guo, N. Zhou, C. Sun, and J. Peng, “Freeway Driving Cycle Construction Based on Real-Time Traffic Information and Global Optimal Energy

Management for Plug-In Hybrid Electric Vehicles," *Energies*, vol. 10, no. 11, p. 1796, Nov. 2017, doi: 10.3390/en10111796.

[33] U. Galgamuwa, L. Perera, and S. Bandara, "A Representative Driving Cycle for the Southern Expressway Compared to Existing Driving Cycles," *Transp. Dev. Econ.*, vol. 2, no. 2, p. 22, Oct. 2016, doi: 10.1007/s40890-016-0027-4.

[34] S. Shi *et al.*, "Research on Markov property analysis of driving cycles and its application," *Transp. Res. Part Transp. Environ.*, vol. 47, pp. 171–181, Aug. 2016, doi: 10.1016/j.trd.2016.05.013.

[35] X. Zhao, Y. Ye, J. Ma, P. Shi, and H. Chen, "Construction of electric vehicle driving cycle for studying electric vehicle energy consumption and equivalent emissions," *Environ. Sci. Pollut. Res.*, vol. 27, no. 30, pp. 37395–37409, Oct. 2020, doi: 10.1007/s11356-020-09094-4.

[36] X. Zhao, Q. Yu, J. Ma, Y. Wu, M. Yu, and Y. Ye, "Development of a Representative EV Urban Driving Cycle Based on a k-Means and SVM Hybrid Clustering Algorithm," *J. Adv. Transp.*, vol. 2018, pp. 1–18, Nov. 2018, doi: 10.1155/2018/1890753.

[37] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," *Appl. Energy*, vol. 211, pp. 538–548, Feb. 2018, doi: 10.1016/j.apenergy.2017.11.072.

[38] J. P. Trovão, P. G. Pereirinha, H. M. Jorge, and C. H. Antunes, "A multi-level energy management system for multi-source electric vehicles – An integrated rule-based meta-heuristic approach," *Appl. Energy*, vol. 105, pp. 304–318, May 2013, doi: 10.1016/j.apenergy.2012.12.081.

[39] Z. Chen, H. Hu, Y. Wu, Y. Zhang, G. Li, and Y. Liu, "Stochastic model predictive control for energy management of power-split plug-in hybrid electric vehicles based on reinforcement learning," *Energy*, vol. 211, p. 118931, Nov. 2020, doi: 10.1016/j.energy.2020.118931.

[40] L. Kouchachvili, W. Yaïci, and E. Entchev, "Hybrid battery/supercapacitor energy storage system for the electric vehicles," *J. Power Sources*, vol. 374, pp. 237–248, Jan. 2018, doi: 10.1016/j.jpowsour.2017.11.040.

[41] Q. Zhang, L. Wang, G. Li, and Y. Liu, "A real-time energy management control strategy for battery and supercapacitor hybrid energy storage systems of pure electric vehicles," *J. Energy Storage*, vol. 31, p. 101721, Oct. 2020, doi: 10.1016/j.est.2020.101721.

[42] S. East and M. Cannon, “Optimal Power Allocation in Battery/Supercapacitor Electric Vehicles Using Convex Optimization,” *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 12751–12762, Nov. 2020, doi: 10.1109/TVT.2020.3023186.

[43] Z. Song *et al.*, “Multi-objective optimization of a semi-active battery/supercapacitor energy storage system for electric vehicles,” *Appl. Energy*, vol. 135, pp. 212–224, Dec. 2014, doi: 10.1016/j.apenergy.2014.06.087.

[44] X. Qi, Y. Luo, G. Wu, K. Boriboonsomsin, and M. Barth, “Deep reinforcement learning enabled self-learning control for energy efficient driving,” *Transp. Res. Part C Emerg. Technol.*, vol. 99, pp. 67–81, Feb. 2019, doi: 10.1016/j.trc.2018.12.018.

[45] Y. Wu, H. Tan, J. Peng, H. Zhang, and H. He, “Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus,” *Appl. Energy*, vol. 247, pp. 454–466, Aug. 2019, doi: 10.1016/j.apenergy.2019.04.021.

[46] X. Lin, Y. Wang, P. Bogdan, N. Chang, and M. Pedram, “Reinforcement learning based power management for hybrid electric vehicles,” in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov. 2014, pp. 33–38. doi: 10.1109/ICCAD.2014.7001326.

[47] X. Qi, G. Wu, K. Boriboonsomsin, M. J. Barth, and J. Gonder, “Data-Driven Reinforcement Learning–Based Real-Time Energy Management System for Plug-In Hybrid Electric Vehicles,” *Transp. Res. Rec.*, vol. 2572, no. 1, pp. 1–8, Jan. 2016, doi: 10.3141/2572-01.

[48] C. Liu and Y. L. Murphey, “Power management for Plug-in Hybrid Electric Vehicles using Reinforcement Learning with trip information,” in *2014 IEEE Transportation Electrification Conference and Expo (ITEC)*, Jun. 2014, pp. 1–6. doi: 10.1109/ITEC.2014.6861862.

[49] B. Xu, J. Shi, S. Li, H. Li, and Z. Wang, “Energy consumption and battery aging minimization using a Q-learning strategy for a battery/ultracapacitor electric vehicle,” *Energy*, vol. 229, p. 120705, Aug. 2021, doi: 10.1016/j.energy.2021.120705.

[50] T. Liu, X. Hu, S. E. Li, and D. Cao, “Reinforcement Learning Optimized Look-Ahead Energy Management of a Parallel Hybrid Electric Vehicle,” *IEEEASME Trans. Mechatron.*, vol. 22, no. 4, pp. 1497–1507, Aug. 2017, doi: 10.1109/TMECH.2017.2707338.

[51] Y. Zou, T. Liu, D. Liu, and F. Sun, “Reinforcement learning-based real-time energy management for a hybrid tracked vehicle,” *Appl. Energy*, vol. 171, pp. 372–382, Jun. 2016, doi: 10.1016/j.apenergy.2016.03.082.

[52] Y. Ye, J. Zhang, and B. Xu, “A Fast Q-learning Energy Management Strategy for Battery/Supercapacitor Electric Vehicles Considering Energy Saving and Battery Aging,” in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*, Cape Town, South Africa: IEEE, Dec. 2021, pp. 1–6. doi: 10.1109/ICECET52533.2021.9698682.

[53] Y. Hu, W. Li, H. Xu, and G. Xu, “An Online Learning Control Strategy for Hybrid Electric Vehicle Based on Fuzzy Q-Learning,” *Energies*, vol. 8, no. 10, Art. no. 10, Oct. 2015, doi: 10.3390/en81011167.

[54] J. Wu, Y. Zou, X. Zhang, T. Liu, Z. Kong, and D. He, “An Online Correction Predictive EMS for a Hybrid Electric Tracked Vehicle Based on Dynamic Programming and Reinforcement Learning,” *IEEE Access*, vol. 7, pp. 98252–98266, 2019, doi: 10.1109/ACCESS.2019.2926203.

[55] B. Xu, F. Malmir, and Z. Filipi, “Real-Time Reinforcement Learning Optimized Energy Management for a 48V Mild Hybrid Electric Vehicle,” *SAE Tech. Pap.*, Apr. 2019, doi: 10.4271/2019-01-1208.

[56] J. Ho and S. Ermon, “Generative Adversarial Imitation Learning,” p. 9.

[57] G. Bhatti, H. Mohan, and R. Raja Singh, “Towards the future of smart electric vehicles: Digital twin technology,” *Renew. Sustain. Energy Rev.*, vol. 141, p. 110801, May 2021, doi: 10.1016/j.rser.2021.110801.

[58] Y. Xie *et al.*, “Microsimulation of electric vehicle energy consumption and driving range,” *Appl. Energy*, vol. 267, p. 115081, Jun. 2020, doi: 10.1016/j.apenergy.2020.115081.

[59] Z. Zhang, Y. Zou, T. Zhou, X. Zhang, and Z. Xu, “Energy Consumption Prediction of Electric Vehicles Based on Digital Twin Technology,” *World Electr. Veh. J.*, vol. 12, no. 4, p. 160, Sep. 2021, doi: 10.3390/wevj12040160.

[60] S. Inuzuka, F. Xu, B. Zhang, and T. Shen, “Reinforcement Learning Based on Energy Management Strategy for HEVs,” in *2019 IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2019, pp. 1–6. doi: 10.1109/VPPC46532.2019.8952511.

[61] J. Wu, Z. Wei, W. Li, Y. Wang, Y. Li, and D. U. Sauer, “Battery Thermal- and Health-Constrained Energy Management for Hybrid Electric Bus Based on Soft Actor-Critic DRL Algorithm,” *IEEE Trans. Ind. Inform.*, vol. 17, no. 6, pp. 3751–3761, Jun. 2021, doi: 10.1109/TII.2020.3014599.

[62] J. Zhou, S. Xue, Y. Xue, Y. Liao, J. Liu, and W. Zhao, “A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning,” *Energy*, vol. 224, p. 120118, Jun. 2021, doi: 10.1016/j.energy.2021.120118.

[63] J. Lin and D. A. Niemeier, “An exploratory analysis comparing a stochastic driving cycle to California’s regulatory cycle,” *Atmos. Environ.*, vol. 36, no. 38, pp. 5759–5770, Dec. 2002, doi: 10.1016/S1352-2310(02)00695-7.

[64] J. Lin and D. A. Niemeier, “Estimating Regional Air Quality Vehicle Emission Inventories: Constructing Robust Driving Cycles,” *Transp. Sci.*, Aug. 2003, doi: 10.1287/trsc.37.3.330.16045.

[65] S. Rangaraju, L. De Vroey, M. Messagie, J. Mertens, and J. Van Mierlo, “Impacts of electricity mix, charging profile, and driving behavior on the emissions performance of battery electric vehicles: A Belgian case study,” *Appl. Energy*, vol. 148, pp. 496–505, Jun. 2015, doi: 10.1016/j.apenergy.2015.01.121.

[66] Z. Yu *et al.*, “Statistical inference-based research on sampling time of vehicle driving cycle experiments,” *Transp. Res. Part Transp. Environ.*, vol. 54, pp. 114–141, Jul. 2017, doi: 10.1016/j.trd.2017.04.029.

[67] T. L. Saaty, “Analytic Heirarchy Process,” in *Wiley StatsRef: Statistics Reference Online*, John Wiley & Sons, Ltd, 2014. doi: 10.1002/9781118445112.stat05310.

[68] X. Zhao, X. Zhao, Q. Yu, Y. Ye, and M. Yu, “Development of a representative urban driving cycle construction methodology for electric vehicles: A case study in Xi’an,” *Transp. Res. Part Transp. Environ.*, vol. 81, p. 102279, Apr. 2020, doi: 10.1016/j.trd.2020.102279.

[69] J. D. K. Bishop, C. J. Axon, and M. D. McCulloch, “A robust, data-driven methodology for real-world driving cycle development,” *Transp. Res. Part Transp. Environ.*, vol. 17, no. 5, pp. 389–397, Jul. 2012, doi: 10.1016/j.trd.2012.03.003.

[70] D. Huang, H. Xie, H. Ma, and Q. Sun, “Driving cycle prediction model based on bus route features,” *Transp. Res. Part Transp. Environ.*, vol. 54, pp. 99–113, Jul. 2017, doi: 10.1016/j.trd.2017.04.038.

[71] B. Yue, S. Shi, N. Lin, P. Guo, Z. Li, and Z. Zhang, “Study on the Design Method of Driving Cycle with Road Grade Based on Markov Chain Model,” in *2015 IEEE Vehicle Power and Propulsion Conference (VPPC)*, Oct. 2015, pp. 1–5. doi: 10.1109/VPPC.2015.7353026.

[72] J. Gonder and T. Markel, “Energy Management Strategies for Plug-In Hybrid Electric Vehicles,” SAE International, Warrendale, PA, SAE Technical Paper 2007-01–0290, Apr. 2007. doi: 10.4271/2007-01-0290.

[73] X. Hu, J. Jiang, B. Egardt, and D. Cao, “Advanced Power-Source Integration in Hybrid Electric Vehicles: Multicriteria Optimization Approach,” *IEEE Trans. Ind. Electron.*, vol. 62, no. 12, pp. 7847–7858, Dec. 2015, doi: 10.1109/TIE.2015.2463770.

[74] L. Serrao, S. Onori, and G. Rizzoni, “ECMS as a realization of Pontryagin’s minimum principle for HEV control,” in *2009 American Control Conference*, Jun. 2009, pp. 3964–3969. doi: 10.1109/ACC.2009.5160628.

[75] H. Khayyam and A. Bab-Hadiashar, “Adaptive intelligent energy management system of plug-in hybrid electric vehicle,” *Energy*, vol. 69, pp. 319–335, May 2014, doi: 10.1016/j.energy.2014.03.020.

[76] C.-C. Lin, H. Peng, J. W. Grizzle, and J.-M. Kang, “Power management strategy for a parallel hybrid electric truck,” *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 6, pp. 839–849, Nov. 2003, doi: 10.1109/TCST.2003.815606.

[77] L. Xu, M. Ouyang, J. Li, F. Yang, L. Lu, and J. Hua, “Optimal sizing of plug-in fuel cell electric vehicles using models of vehicle performance and system cost,” *Appl. Energy*, vol. 103, pp. 477–487, Mar. 2013, doi: 10.1016/j.apenergy.2012.10.010.

[78] E. Vinot, R. Trigui, Y. Cheng, C. Espanet, A. Bouscayrol, and V. Reinbold, “Improvement of an EVT-Based HEV Using Dynamic Programming,” *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 40–50, Jan. 2014, doi: 10.1109/TVT.2013.2271646.

[79] D. Mourembles, B. Buergler, L. Gajewski, A. Cooke, and C. Barchasz, “Li-S Cells for Space Applications (LISSA),” in *2019 European Space Power Conference (ESPC)*, Sep. 2019, pp. 1–5. doi: 10.1109/ESPC.2019.8931976.

[80] M. Taggougui *et al.*, “Batteries Annual Progress Report (FY2019),” Lawrence Livermore National Lab. (LLNL), Livermore, CA (United States); Sandia National Lab. (SNL-NM), Albuquerque, NM (United States); Argonne National Lab. (ANL), Argonne, IL (United States). Argonne Leadership Computing Facility (ALCF); Brookhaven National Lab. (BNL), Upton, NY (United States); Lawrence Berkeley National Lab. (LBNL), Berkeley, CA (United States). National Energy Research Scientific Computing Center (NERSC), DOE/EE-1987, May 2020. doi: 10.2172/1637433.

[81] A. Fotouhi, D. J. Auger, K. Propp, and S. Longo, “Lithium–Sulfur Battery State-of-Charge Observability Analysis and Estimation,” *IEEE Trans. Power Electron.*, vol. 33, no. 7, pp. 5847–5859, Jul. 2018, doi: 10.1109/TPEL.2017.2740223.

[82] L. Sun, “A Novel Battery-Supercapacitor Power Supply For Electric Vehicles (EVs) – Design, Simulation And Experiment”.

[83] S. Dörfler, H. Althues, P. Härtel, T. Abendroth, B. Schumm, and S. Kaskel, “Challenges and Key Parameters of Lithium-Sulfur Batteries on Pouch Cell Level,” *Joule*, vol. 4, no. 3, pp. 539–554, Mar. 2020, doi: 10.1016/j.joule.2020.02.006.

[84] Y. Ye, H. Wang, B. Xu, and J. Zhang, “An imitation learning-based energy management strategy for electric vehicles considering battery aging,” *Energy*, p. 128537, Jul. 2023, doi: 10.1016/j.energy.2023.128537.

[85] D. Bertsekas, *Reinforcement Learning and Optimal Control*. Athena Scientific, 2019.

[86] J. Cao and R. Xiong, “Reinforcement Learning-based Real-time Energy Management for Plug-in Hybrid Electric Vehicle with Hybrid Energy Storage System,” *Energy Procedia*, vol. 142, pp. 1896–1901, Dec. 2017, doi: 10.1016/j.egypro.2017.12.386.

[87] J. Wu, H. He, J. Peng, Y. Li, and Z. Li, “Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus,” *Appl. Energy*, vol. 222, pp. 799–811, Jul. 2018, doi: 10.1016/j.apenergy.2018.03.104.

[88] A. Attia and S. Dayan, “Global overview of Imitation Learning.” arXiv, Jan. 19, 2018. Accessed: Nov. 25, 2022. [Online]. Available: http://arxiv.org/abs/1801.06503

[89] V. Mnih *et al.*, “Playing Atari with Deep Reinforcement Learning.” arXiv, Dec. 19, 2013. doi: 10.48550/arXiv.1312.5602.

[90] V. Mnih *et al.*, “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, Art. no. 7540, Feb. 2015, doi: 10.1038/nature14236.

[91] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized Experience Replay.” arXiv, Feb. 25, 2016. doi: 10.48550/arXiv.1511.05952.

[92] M. Andrychowicz *et al.*, “Hindsight Experience Replay,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2017. Accessed: Jan. 17, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/453fadbd8a1a3af50a9df4df899537b5-Abstract.html

[93] H. Hasselt, “Double Q-learning,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2010. Accessed: Jan. 17, 2023. [Online]. Available: https://proceedings.neurips.cc/paper/2010/hash/091d584fced301b442654dd8c23b3fc9-Abstract.html

[94] M. Hessel *et al.*, “Rainbow: Combining Improvements in Deep Reinforcement Learning.” arXiv, Oct. 06, 2017. Accessed: Jan. 17, 2023. [Online]. Available: http://arxiv.org/abs/1710.02298

[95] S. Sutton, “Predicting and Explaining Intentions and Behavior: How Well Are We Doing?,” *J. Appl. Soc. Psychol.*, vol. 28, no. 15, pp. 1317–1338, 1998, doi: 10.1111/j.1559-1816.1998.tb01679.x.

[96] R. S. Sutton and A. G. Barto, *Reinforcement Learning, second edition: An Introduction*. MIT Press, 2018.

[97] M. Fortunato *et al.*, “Noisy Networks for Exploration.” arXiv, Jul. 09, 2019. doi: 10.48550/arXiv.1706.10295.

[98] T. P. Lillicrap *et al.*, “Continuous control with deep reinforcement learning.” arXiv, Jul. 05, 2019. doi: 10.48550/arXiv.1509.02971.

[99] S. Fujimoto, H. Hoof, and D. Meger, “Addressing Function Approximation Error in Actor-Critic Methods,” in *Proceedings of the 35th International Conference on Machine Learning*, PMLR, Jul. 2018, pp. 1587–1596. Accessed: Jan. 17, 2023. [Online]. Available: https://proceedings.mlr.press/v80/fujimoto18a.html

[100] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, “Trust Region Policy Optimization,” in *Proceedings of the 32nd International Conference on Machine Learning*, PMLR, Jun. 2015, pp. 1889–1897. Accessed: Jan. 17, 2023. [Online]. Available: https://proceedings.mlr.press/v37/schulman15.html

[101] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms.” arXiv, Aug. 28, 2017. doi: 10.48550/arXiv.1707.06347.

[102] B. Xu *et al.*, “Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle,” *Appl. Energy*, vol. 259, p. 114200, Feb. 2020, doi: 10.1016/j.apenergy.2019.114200.