Modelling and Optimization of Multi-Energy System Operation Based on Deep Reinforcement Learning

Zhang, Bin

DOI (link to publication from Publisher):
10.54337/aau763295893

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

MODELLING AND OPTIMIZATION OF MULTI-ENERGY SYSTEMS OPERATION BASED ON DEEP REINFORCEMENT LEARNING

BY
BIN ZHANG

PhD Thesis 2024

AALBORG UNIVERSITY
DENMARK

MODELLING AND OPTIMIZATION OF MULTI-ENERGY SYSTEMS OPERATION BASED ON DEEP REINFORCEMENT LEARNING

by

Bin Zhang

AALBORG UNIVERSITY
DENMARK

Thesis submitted to
the Faculty of Engineering and Science at Aalborg University

for the degree of
Doctor of Philosophy in Electrical Engineering

PhD Thesis 2024

Submitted: September 2024

Main Supervisor: Professor Zhe Chen
Aalborg University

Co-supervisor: Zhou Liu
Aalborg University

Assessment: Associate Professor Zhenyu Yang (Chair)
Aalborg University, Denmark

Professor Frank Phillipson
Maastricht University, The Netherlnads

Professor Andreas Sumper
Universitat Politècnica de Catalunya, Spain

PhD Series: Faculty of Engineering and Science, Aalborg University

Department: AAU Energy

ISSN: 2446-1636
ISBN: 978-87-85239-51-8

Published by:
Aalborg University Open Publishing
Kroghstræde 1-3
DK – 9220 Aalborg Øst
aauopen@aau.dk

CV

Bin Zhang obtained a Bachelor of Science degree in Automation from Hohai University in 2017, and a Master of Science degree in Electrical Engineering from the University of Electronics Science and Technology of China in 2021. He is now pursuing his Ph.D. at AAU Energy, Aalborg University, Denmark, as a member of the Renewable Energy Research Group.

His research interests focus on application of artificial intelligent in power systems, including machine learning, multi-energy operation optimization, microgrids, electric vehicles.

ABSTRACT

The growth of the global economy has led to noticeable issues of energy crisis and environmental issues, hence driving the advancement of alternative energy production. Renewable energy generation may effectively access the power system by utilizing dispersed access to the power grid. Renewable energy generation exhibits significant unpredictability and instability, and the inclusion of numerous renewable energy sources (RESs) leads to substantial alterations in the functioning of the power grid. Moreover, the unpredictable variability of energy costs and the inconsistent patterns of controlled loads, such as electric vehicles (EVs), introduce a level of uncertainty to the functioning of the power system.

Another ongoing energy transition is the integration of different forms of energy. Multi-energy system (MES), have the potential to yield significant sustainable, efficient, economic and resiliency benefits. However, intermittent RES generation, uncertain and heterogeneous load demands, and balance-of-system costs render the traditional energy analysis methods obsolete.

Artificial Intelligence (AI) technology is an important tool to address the above challenges. As an important branch in the field of computer science, AI technology aims to realize the self-improvement of computers and the simulation of human intelligence by refining knowledge and experience from data. Since the knowledge extracted from data has a certain generalization ability, AI methods can cope with the uncertainty of the system and enable online decision-making.

Hence, the aim of this thesis is to utilize DRL algorithms to guarantee an efficient and dependable energy management strategy for the MESs. In Chapter 2, with the objective of achieving a low-carbon economic dispatch strategy for the electricity-gas MES, this project takes into account the flexible coordination between the carbon capture system and power-to-gas units. Chapter 3 investigates a two-timescale energy management strategy for the residential MES. This strategy utilizes a multi-agent deep reinforcement learning (MADRL) algorithm to control the internal energy conversion and external energy trading behaviors. Chapter 4 introduces a decentralized energy management strategy for multiple MESs. The proposed framework is a bilevel energy management system. In the bottom layer, a control strategy based on MADRL is used for the multi-energy microgrid (MG) cluster. In the upper layer, energy routers determine the optimal energy trading based on feedback from the bottom layer. Moreover, an energy hub (EH) is a very efficient solution for managing energy in the MES. EVs have been extensively integrated into the power grid in recent years. Hence, taking into account that EHs and EVs pertain to distinct entities, a refined MADRL-based decentralized energy management strategy is suggested to optimize the revenues of the EV entity and decrease the energy expenses of the EH entity. In addition, a specialized neural network is employed to address the intricate uncertainties, hence enhancing the effectiveness of the suggested strategy. Chapter 5

contains information that is related to the topic. Chapter 6 presents the final findings of the thesis.

To conduct efficient simulation of the proposed AI-based energy management strategy, a series of case studies were performed on Python. Specifically, training datasets come from real-world historical datasets, and AI algorithms are programmed based on TensorFlow. Besides, algorithm comparison is also conducted to illustrate the superiority of the proposed method. Simulation results demonstrate that the proposed strategy can (i) reduce energy costs, (ii) deal with uncertainties, (iii) provide real-time energy management strategy, and (iv) realize decentralized energy management for different entities. Besides, the proposed method shifts the computation from online to offline, which greatly reduces the computation of online execution and facilitates later applications.

DANSK RESUME

Med udviklingen af den globale økonomi bliver problemerne med energimangel og miljøforurening mere og mere fremtrædende, hvilket fremmer udviklingen af ny energiproduktion. Vedvarende energiproduktion gennem distribueret adgang til elnettet er en vigtig måde for ny energiproduktion at få adgang til elnettet på. Produktionen af vedvarende energi har imidlertid en stærk tilfældighed og volatilitet, og adgangen til et stort antal vedvarende energikilder medfører dybtgående ændringer i driften af elnettet. Derudover medfører den stokastiske karakter af energipriser og svingende adfærd af kontrollerbare belastninger, herunder elektriske køretøjer (EV), risiko for driften af energisystemet.

En anden igangværende energiovergang er integrationen af forskellige energiformer. Multi-energy system (MES) har potentiale til at give betydelige bæredygtige, effektive, økonomiske og modstandsdygtige fordele. Periodisk elproduktion, usikre og heterogene belastningskrav og systembalanceomkostninger gør de traditionelle energianalysemetoder forældede.

Kunstig intelligens (AI) teknologi er et vigtigt redskab til at løse ovenstående udfordringer. Som en vigtig gren inden for datalogi, AI teknologi sigter mod at realisere selv-forbedring af computere og simulering af menneskelig intelligens ved at raffinere viden og erfaring fra data. Da viden udvundet fra data har en vis generaliseringsevne, kan AI-metoder klare usikkerheden i systemets kildebelastning og muliggøre online beslutningstagning.

Formålet med denne afhandling er derfor at anvende AI metoder til at sikre en effektiv og pålidelig energistyringsstrategi for MES. Med henblik herpå starter dette projekt med en lav CO2-økonomisk afsendelsesstrategi for elektricitet-gas MES, hvor den fleksible koordinering mellem kulstofopsamling og el-til-gas-enheder overvejes. Det relative indhold er præsenteret i kapitel 2. For det andet, for at løse spørgsmålene om centraliseret energistyring, er en decentraliseret energistyringsstrategi udviklet i en bolig MES, hvor multi-agent dyb forstærkning læring (MADRL) metode anvendes til at regulere den interne energikonvertering og eksterne energihandelsadfærd. Tilsvarende indhold præsenteres i kapitel 3. I kapitel 4, for at undersøge en decentraliseret energistyringsstrategi for flere MES, foreslås en bilag energistyringsstrategi, hvor der foreslås en MADRL-baseret styringsstrategi for den nederste lag multi-energy mikrogrid klynge, og de øverste lag energi routere bestemmer den optimale energihandel baseret på bundlagsinformation feedback. Endvidere er energihub (EH) en effektiv løsning til at levere energistyring til MES. I de senere år er elektriske køretøjer også blevet tilsluttet nettet i stor skala. I betragtning af, at EH'er og EV'er tilhører forskellige enheder, foreslås der derfor en forbedret MADRL-baseret decentraliseret energistyringsstrategi for at maksimere overskuddet for EV-enheden og minimere energiomkostningerne for EH-enheden. Desuden bruges et specifikt neuralt netværk til at tackle de komplekse usikkerheder, så

ydeevnen af den foreslåede metode forbedres. Det relative indhold er præsenteret i kapitel 5. Endelig er afhandlingens konklusioner introduceret i kapitel 6.

For at udføre effektiv simulering af den foreslåede AI-baserede energistyringsstrategi blev der udført en række casestudier på Python. Specielt kommer træningsdatasæt fra historiske datasæt i virkeligheden, og AI-algoritmer programmeres baseret på TensorFlow. Desuden udføres algoritme sammenligning også for at illustrere overlegenheden af den foreslåede metode. Simuleringsresultater viser, at den foreslåede strategi kan (i) reducere energiomkostningerne, (ii) håndtere usikkerheder, (iii) levere energi i realtid og (iV) realisere decentraliseret energistyring for forskellige enheder. Desuden skifter den foreslåede metode beregningen fra online til offline, hvilket i høj grad reducerer beregningen af online udførelse og letter senere applikationer.

PREFACE

This Ph.D. project is a summary of outcomes from the project "*Deep Reinforcement Learning-based Smart Energy Management Strategy for an Integrated Energy System with Wind Energy*", conducted at the Renewable Energy Research Group at AAU Energy, Aalborg University, Denmark.

Above all, I want to sincerely thank Prof. Zhe Chen for his consistent support and guidance during my doctoral studies. The individual's proficiency, forbearance, and commitment have played a crucial role in influencing my research and academic development. The insightful feedback and constructive criticism from Prof. Chen have motivated me to pursue excellence throughout my entire Ph.D. research. Prof. Chen is not only my mentor, but also my role model and guide. I have benefited greatly from his dedication and academic qualities. Thank you for his patience, guidance and encouragement when I encountered difficulties and setbacks, his support gives me confidence to overcome all difficulties. Professor's rigorous attitude and demanding requirements have made me understand the seriousness and hardship of doing academic research. Thanks for giving me the space for free development and encouraging me to explore new research directions and think about the depth of the problem.

Secondly, I want to express my heartfelt appreciation for the incredible relationships that have sustained me all through three years. Thanks to my lover Xinyu Cao and my lucky little Loopy, your understanding is my motivation to keep going. Thanks to my families, Xuewei Wu, Kuangpu Liu, Hang Ren. Their support, encouragement, understanding and accompany have been my pillars of strength.

Lastly, but significantly, I would want to express my profound appreciation for my beloved homeland. The scholarships and research grants provided not only eased my financial burden but also allowed me to focus on my academic pursuits. I sincerely hope that more and more people will come to China to enjoy the Chinese profound historical and cultural heritage and magnificent scenery.

Bin Zhang

Aalborg University, January 18, 2024

CONTENTS

LIST OF ACRONYMS

IEA	International Energy Agency
MES	Multi-Energy System
DER	Distributed Energy Resources
DL	Deep Learning
DRL	Deep Reinforcement Learning
MDP	Markov Decision Process
MADRL	Multi-Agent Deep Reinforcement Learning
CCS	Carbon Capture and Storage
CCPP	Carbon Capture Power Plants
ES	Electrical Storage
WE	Water Electrolyzer
FC	Fuel Cell
GB	Gas Boiler
HB	Hydrogen Boiler
GT	Gas Turbine
HT	Hydrogen Tank
MG	Microgrid
MEMG	Multi-Energy Microgrid
EI	Energy Internet
ER	Energy Router
ADMM	Alternate Direction Method of Multipliers
EV	Electric Vehicle
EVAGG	Electric Vehicle Aggregator
P2G	Power-to-Gas
RES	Renewable Energy Sources
PER	Prioritized Experience Replay

AC	Actor-Critic
SAC	Soft Actor-Critic
DQN	Deep Q Network
SA	Scenario Analysis
RMES	Residential Multi-Energy System
MAQ	Multi-Agent Q Learning
MADDPG	Multi-Agent Deep Deterministic Policy Gradient
DG	Distributed Generator
CHP	Combined Heating and Power Plant
TD3	Twin Delayed Deep Deterministic Policy Gradient
MAATD3	Multi-Agent Attention TD3
MLP	Multilayer Perceptron
PV	Photovoltaic
SoC	State-of-Charge
OPF	Optimal Power Flow
EH	Energy Hub
WT	Wind Turbine
ESS	Energy Storage System
DNN	Deep Neural Network
RNN	Recurrent Neural Network
PSO	Particle Swarm Optimization
IEDHS	Integrated Electricity and District Heating System
DHN	District Heating Network
PDN	Power Distribution Network
LSTM	Long Short-Term Memory
MASAC	Multi-Agent Soft Actor-Critic
RMES	Residential Multiple Energy System
MILP	Mixed-Integer Linear Programming

TABLES AND FIGURES

Part I. Report

CHAPTER 1. INTRODUCTION

1.1. BACKGROUND

Energy, particularly electricity, is the foundation of social and economic development [1]. In the past, the majority of electricity was generated using fossil fuels, which not only caused the world energy crisis but also worsened environmental pollution through their overexploitation [2]. The International Energy Agency (IEA) predicts that CO_2 emissions will increase by 130% by 2050, leading to a global average temperature rise of 6°C [3]. According to the IEA [4], the world's cumulative photovoltaic (PV) installed capacity was approximately 800 GW. According to the Global Wind Energy Council [5], the global cumulative wind power installed capacity was 743 GW at the end of 2020. GWEC also forecasted that an additional 469 GW of wind power capacity would be added between 2021 and 2025, which would bring the total installed capacity to over 1.2 TW by the end of 2025. According to statistics from the Danish Energy Agency, Figure 1-1 shows the gross energy consumption in Denmark as of 2022 [6]. It is notable that RES accounted for a significant portion, with wind and solar power alone contributing to nearly 60% of the total electricity production.

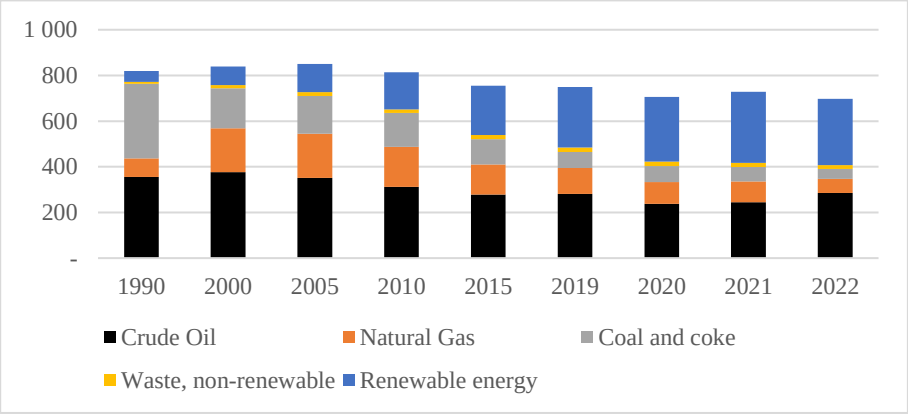


Figure 1-1 Danish gross energy consumption as of 2021. Source: [6]

However, due to the randomness and intermittency of renewable energy, the issue of renewable energy integration has become increasingly prominent. Relying solely on exploiting the existing potential of power systems makes it difficult to overcome the challenges of renewable energy integration. Currently, different energy systems operate in isolation with limited coordination, severely affecting the flexibility of power system operations and failing to fully tap into the potential of these systems. Therefore, developing theories and methods for multi-energy system (MES)

integration is an effective way to address the renewable energy integration challenge [8].

A generalized MES refers to a large-scale system encompassing various energy systems. This system involves different stages, including energy development, conversion, storage, transportation, scheduling, control, management, and utilization. Different types of energy have complex coupling relationship. Natural-gas and heating networks can also be converted into electricity in various ways. Additionally, the integration of distributed energy resources (DERs) further enriches the MES, the structure of which is shown in Figure 1-2. Effective energy management strategies for MESs can optimizes the use of diverse energy sources, leading to improved overall efficiency and reduced operational costs. By coordinating different energy carriers such as electricity, heat, and gas, these strategies enhance system flexibility and reliability, facilitating the integration of RESs [9]. Additionally, effective management strategies can support dynamic demand response and reduce environmental impact, contributing to more sustainable and resilient energy systems [10].

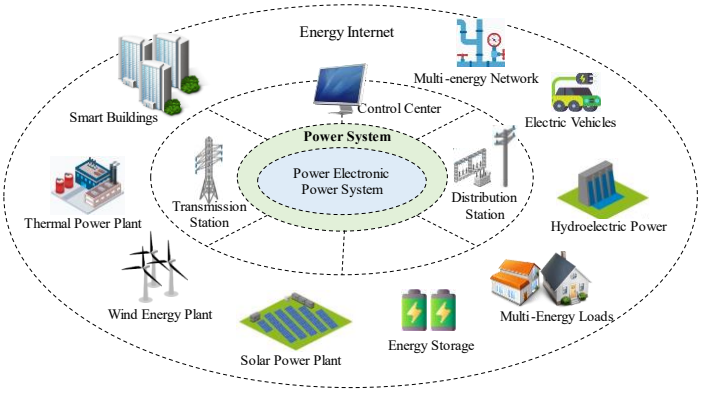


Figure 1-2 A typical framework of the MES [J1].

1.2. STATE OF THE ART

1.2.1. DEEP REINFORCEMENT LEARNING

Deep reinforcement learning (DRL) combines the feature representation capabilities of deep learning (DL) with the decision-making abilities of reinforcement learning (RL). DRL efficiently solves sequential problems by breaking them down into multiple subproblems and solving them step by step [11]. During the training phase, the DRL agent embeds knowledge extracted from historical data into a neural network, enabling online decision-making after deployment. Additionally, with proper design,

DRL can achieve control without relying on physical models, thereby reducing control deviations caused by inaccuracies in the physical model [12][13].

DRL's framework is shown in Figure 1-3. In DRL, the interaction between the agent and the environment can be described as a Markov Decision Process (MDP). In this framework, the agent observes the current state of the environment, selects an action based on the states, and then receives a reward from the environment. The agent's objective is to learn a policy that maximizes the cumulative reward over time by optimizing its actions across different states. This iterative process allows the agent to adapt and improve its strategy to solve complex optimization problems effectively.

Common DRL algorithms are deep Q-Network (DQN), deep deterministic policy gradient (DDPG), twin delayed deep deterministic policy gradient (TD3) and soft actor-critic (SAC). DQN is known for its simplicity and effectiveness in discrete action spaces, but it struggles with stability and exploration, especially in continuous action environments. DDPG addresses continuous action spaces by combining actor-critic methods and deterministic policies, but it often suffers from instability and overestimation of action values. TD3 improves upon DDPG by using two Q-networks to reduce overestimation bias and applying delayed policy updates for better stability, although it can still be sample-inefficient. SAC, on the other hand, introduces entropy regularization to encourage exploration, making it more robust and stable in complex environments, but it tends to be computationally expensive due to its soft policy updates. Each of these algorithms has its strengths and weaknesses, depending on the environment's state and action spaces.

DRL has been widely used in power system optimization, such as voltage control [14], demand response [15], and energy storage management [16]. Furthermore, curriculum learning -based DRL, is effective in solving complex environments, e.g., quantum control [17], by gradually increasing the difficulty of tasks during training. Combining transfer learning with DRL allows an agent to leverage knowledge learned from one task or environment to improve its performance on a related task or in a different environment, which has been applied in navigation [18] and trading strategies [19]. Combining DRL with graph learning to extract topological features is an advanced approach that improves decision-making ability of DRL, e.g., distribution network voltage control [20]. In addition, federated learning-based DRL to address demand response problems offers a powerful approach to optimize energy consumption without compromising data privacy [21]. Besides, quantum RL shows superior performance in computer simulations than classical RL, e.g., better performance on a large search space, faster learning speed and better balances between exploration and exploitation [22]. DRL can also integrate with a large language model to realize real-time optimal power flow, in which some unquantifiable linguistic stipulations can be directly modeled as objective [23]. Besides, integrating physics-informed rules into state-of-the-art DRL algorithms enhances the explainability of the obtained results. This approach has already been applied in power systems [24].

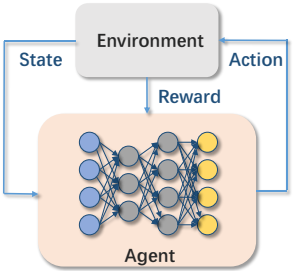


Figure 1-3 Framework of DRL.

In a multi-agent environment, each agent's strategy evolves during training, causing non-stationarity that challenges single-agent DRL algorithms. Multi-agent DRL (MADRL) algorithms using centralized training and decentralized execution address this by leveraging the structural differences between actor and critic networks [25]. The framework of the MADRL is shown in Figure 1-4. The critic network, which uses global information from all agents, helps guide the learning of cooperative strategies. The actor network, using only local information, is used during execution. This design maintains environmental stability for each agent and improves the stability of the training process. MADRL has been used to address multi-robot coordination control problems [26]. In addition, MADRL can be combined with graph learning to enhance algorithm performance by learning the topology and node features of the network, e.g., voltage control [27]. MADRL can also be applied for adaptive traffic signal control in the guidance of transfer learning [28], which can improve the scalability of each agent. In MADRL, multiple agents may have cooperative or competitive relationships. Through transfer learning, the strategies learned by certain agents can be used as the foundation for other agents to learn from.

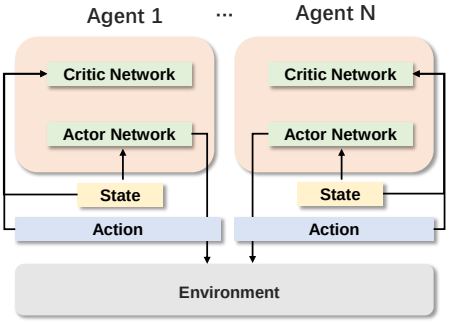


Figure 1-4 Framework of MADRL.

1.2.2. MES ENERGY MANAGEMENT

1) Electricity-Gas MES Energy Management

Integrating electricity and natural gas into a MES offers greater flexibility, efficiency, reliability, and cost-effectiveness [29]. In addition, there is a rising trend in the integration of carbon capture and storage (CCS) technology into coal-fired power plants, leading to the development of carbon capture power plants (CCPP) [30]. Power-to-gas (P2G) technology allows the conversion of excess wind power into synthetic natural gas [31]. Several studies [32]-[34] worked on the availability of fuel, changes in natural gas prices, the influence of wind energy on the costs of power operations, and the development of innovative integrated structures for power and cooling cogeneration systems that offer improved thermal efficiency. Most of these studies tend to focus on single-plant perspectives to investigate the low-carbon energy management strategy, often overlooking the coordination between the CCS and P2G units [35]. The coordination between CCS and P2G can effectively promote the integration of renewable energy, improve the operational flexibility, and reduce generation costs.

In addition, investigating the optimal energy management strategy for the electricity-gas MES is complicated by numerous variables, including fluctuating electricity and gas consumption, uncertain wind power generation, and fluctuating energy prices. Complications also arise from the complex coupling relationships of energy flow models, the lack of network topology, and the intricacies of a non-convex multi-objective function. Traditionally, energy management strategies for MESs have heavily depended on programming methods. The methods can be classified into three primary categories: dynamic programming [36], linear programming [37], and non-linear programming [38]. However, programming methods require a considerable amount of time when handling complex systems or uncertainties, which restricts their scalability in complex systems. Specifically, these methods often require numerous iterations to find the optimal strategy for a given state. Given the computational time required for online deployment, programming-based methods are not well-suited for addressing this problem [39]. Furthermore, to address multiple uncertainties and enable real-time decision-making, DRL-based energy management strategies have been proposed to optimize the operation of electricity-gas MES systems [40].

2) Residential MES Energy Management

Globally, residential energy consumption typically accounts for 30% to 40% of total energy use [41]. This includes heating, cooling, hot water, lighting, and appliances. A residential MES (RMES) includes both heating and electricity supply for residential users. The idea of using energy carriers has arisen as a promising framework for future energy networks [42]. Multi-energy carriers focus on optimizing the interaction between various energy sources and demands. Users may easily modify their energy usage and transition between various energy sources using this method [43]. An effective residential energy management strategy can lead to economic benefits [44], carbon emission reduction [45], enhancing renewable energy usage [46], maintaining residential comfort [47], and balancing load curves [48]. Few studies focus on

residential energy management strategies aimed at addressing both energy trading and energy conversion. This is because the time scales for these two objectives may differ significantly, and their integration introduces additional uncertainties, making the optimization problem more complex.

Energy components in the multi-energy carriers include electrical storage (ES), water electrolyzers (WE), fuel cell (FC), gas boiler (GB), hydrogen boiler (HB), and hydrogen tank (HT). ES devices can also be used to make up for energy shortages by storing extra electrical energy. WE converts electrical energy into hydrogen. This hydrogen can then be used by FC units to produce electricity and heat [49], by HB for heating, or stored in HT for future needs. GB serves as an auxiliary heat source, using purchased natural gas to meet heating requirements.

Hydrogen fuel, noted for its versatility in generating electricity, heating, and powering electric vehicles, plays a significant role in reducing carbon emissions. With ongoing corporate investments and the development of hydrogen infrastructure, hydrogen fuel costs are anticipated to decrease [50], promoting its adoption in residential areas. However, its use in residential buildings is still emerging. The WE is designed to convert electrical energy into hydrogen, which can be stored indefinitely, a significant advantage over other energy storage devices like batteries that require frequent recharging [51]. The investment cost for WEs is decreasing, and their efficiency is improving over time [52].

Centralized residential energy management strategies optimize energy costs in various residential loads and controlling intelligent home appliances for cost minimization [48]. However, centralized energy management strategies face challenges like single-point communication failures and operation maintenance costs [53]. In tackling these challenges, distributed energy management strategies offer an alternative choice, where each subsystem computes its outcomes with minimal communication with others [54]. Different extended frameworks for distributed exchange are available. One direction considers interactions in domestic energy management as a generalized Nash game, while another uses a distributed model predictive controller to regulate collective power consumption [55]. However, model predictive controllers rely on accurate modeling of residential energy systems, which is challenging due to the complex operational modes and energy coupling relationships involved. In addition, the distributed methods suffer from communication delays, e.g., consensus-based methods [56]. Furthermore, data-driven MADRL-based residential energy management strategies have been proposed to optimize residential energy costs. Since MADRL-based strategies only require local information during online deployment, they significantly reduce communication demands [57].

3) Energy Management of Multiple MESs

The shift toward DERs in the modern energy system is steering the electricity sector away from traditional centralized models. In this context, microgrids (MGs) are becoming increasingly vital for enhancing renewable energy usage. More advanced than MGs, multi-energy MGs (MEMGs) are critical for achieving optimal energy solutions by facilitating coordination among different energy sectors like electricity, gas, and heating [58],[59]. MEMGs offer a framework to handle the dynamic interactions and interdependencies among different energy components.

The energy Internet (EI), which is defined by its reliance on renewable energy, decentralized networks, and peer-to-peer connections, is becoming increasingly popular as an attractive option [60]. In decentralized network, each unit makes decisions about its energy consumption or production locally, often based on its specific needs or constraints. It enables the growth and use of many energy sources in a distributed way, providing significant benefits for the environment, economy, and resilience. Energy routers (ERs) are essential components within the EI architecture, similar to internet routers. They facilitate the transmission of both information and energy across MGs, which is necessary for actual EI scenarios [61]. Yet, the decentralized structure of numerous energy networks and the unpredictability of DERs provide difficulties in efficiently controlling units due to its complex physical and communication framework.

Centralized energy management strategies involve a central controller communicating with MGs for global information and decision-making, which is a top-bottom framework [62]. Despite its widespread use, this approach has several drawbacks: high connectivity costs, vulnerability to single-point failures, and significant computational burdens, especially as more DERs integrate into the system [63]. The centralized energy management strategy also struggles with high DER penetration and the need for customized energy interactions, limiting the flexibility of transactions between consumers and markets [64]. There is a growing interest in implementing bottom-up energy management schemes, which offer a more practical strategy for monitoring multiple agent systems (MAS) at both the MG and energy router (ER) levels [45]. These schemes take into account individual MGs' unique consumer energy demands and operational costs, aiming to facilitate future energy planning and cost reduction. For instance, research has demonstrated significant reductions in the levelized cost of electricity through the utilization of bottom-up approaches [46][47]. Instead of starting with a high-level and centralized perspective, bottom-up methods begin by optimizing energy usage at the local level—such as households, devices, or DERs—and then coordinating these local optimizations to achieve broader energy management goals.

However, these bottom-up energy management strategies encounter challenges due to their reliance on traditional mathematical models, which necessitate precise parameter estimation and can be computationally demanding, rendering them unsuitable for real-time energy management [48]. Current research has examined various distributed

techniques, such as game theory, the alternate direction method of multipliers (ADMM), consensus theory, and event-trigger processes, with the purpose of coordinating MGs [49][50]. These techniques aim to address concerns with energy trade and congestion management among MGs. However, their reliance on particular optimization models can lead to complications and possible problems with convergence owing to nonconvexity. Additionally, they have difficulties in dealing with the uncertainties related to renewable energy and complicated energy conversions [51].

4) Energy Management for the MES including EVs

District heating systems involve the centralized production and distribution of thermal energy through pipelines [52],[53]. The integration of electricity and district heating systems has been furthered by the electrification of heating devices and the emergence of energy hubs (EHs). EHs serve as versatile multi-energy carriers, encompassing energy production, conversion, and storage, and play a flexible role in system operations and market trading [54]. Additionally, the rise of electric vehicle (EVs) driven by climate change concerns, air quality improvements, and advancements in battery technology is reshaping transportation. Global EV sales have skyrocketed from 12,000 in 2012 to a record 6.6 million in 2021 [69]. The integration of EVs into the power grid is becoming increasingly important, with projections suggesting a significant increase in electricity demand [70]. Research indicates that integrating EVs into MESs can enhance operational flexibility and reduce costs. For instance, research [71] observed an 8.81% cost decrease by integrating EVs into a MES. Research [72] examined the optimized scheduling of a zero-carbon MES for the next day, using EVs to meet the electricity and cooling requirements. Literatures [73][74] further investigated the economic and emission scheduling in local MESs that include plug-in EVs. They also studied the optimized planning of MESs that include transportation, natural gas, and active distribution networks.

The field of EH optimization is expanding, addressing intricate energy interconnections using methods such as stochastic programming [75]. This study involves the control of the influence of ES on operational expenses and the formulation of approaches for dependable and effective energy administration [76]. A significant amount of academics is now prioritizing the centralized coordination of EHs and EV Aggregators (EVAGGs) in energy management [77]-[84]. Centralized methods involve a central entity handling all decision-making processes. While these methods can minimize energy purchase costs and optimize various objectives, they face challenges like dependency on perfect communication conditions, privacy concerns for prosumers, and potential delays in response times, hindering real-time scheduling [85][86]. In addition, model-based algorithms require accurate modeling of uncertainties and find it difficult to make real-time decisions during online deployment. It’s significant to investigate a model-free decentralized energy management strategy for the MES and EVAGG entities.

To avoid reliance on precise modeling of the MESs and uncertainties, model-free DRL methods have been applied into the optimal energy management of EHs and EVAGGs. For example, Liu et al. [87] and Qiu et al. [88] proposed DQN and DDPG algorithm to minimize the operation costs in a smart EH, respectively. However, DQN and DDPG-based strategies are centralized and do not account for the privacy of different entities. Additionally, DQN is specifically suited for problems with discrete action and state spaces, and DDPG is sensitive to hyperparameter settings. In addition, these studies did not use specialized neural networks to handle uncertainty, nor did they incorporate the system's safety constraints into the algorithm training.

1.3. THE OBJECTIVES OF THIS THESIS

To tackle the aforementioned challenges, the objectives of this thesis are summarized as follows:

1) DRL-based low-carbon economic energy management strategy for the electricity-gas MES.

This thesis develops a data-driven energy management strategy for the electricity-gas MES. A DRL algorithm is applied to find the optimal low-carbon energy management strategy. The coordination between P2G and CCS units is investigated.

2) MADRL-based two-timescale energy management strategy for the residential MES.

This thesis applies an MADRL algorithm to investigate the two-timescale energy management strategy for the residential MES, where an hourly-ahead energy trading agent and a 15-min-ahead energy conversion agent are set. The learned strategy can flexibly adjust unit operations in response to varying load profiles and energy prices.

3) MADRL-based bottom-up energy management strategy for multiple MESs.

This thesis develops a bottom-up energy management framework for the EI network. The bottom layer is an MG cluster composed of multiple MESs, and an MADRL algorithm is applied to learn the optimized operation strategies for the MESs. The upper layer is an ER cluster responsible for energy allocation across MESs.

4) MADRL-based decentralized energy management strategy for MESs and EVAGG.

This thesis presents a DRL-based energy management strategy for the grid-connected MG with EVs. Furthermore, this thesis develops a decentralized energy management strategy for the EH and EVGAA entities, where each entity can make decisions based on local measurements.

The relationships between the above four parts are shown in the Figure 1-5.

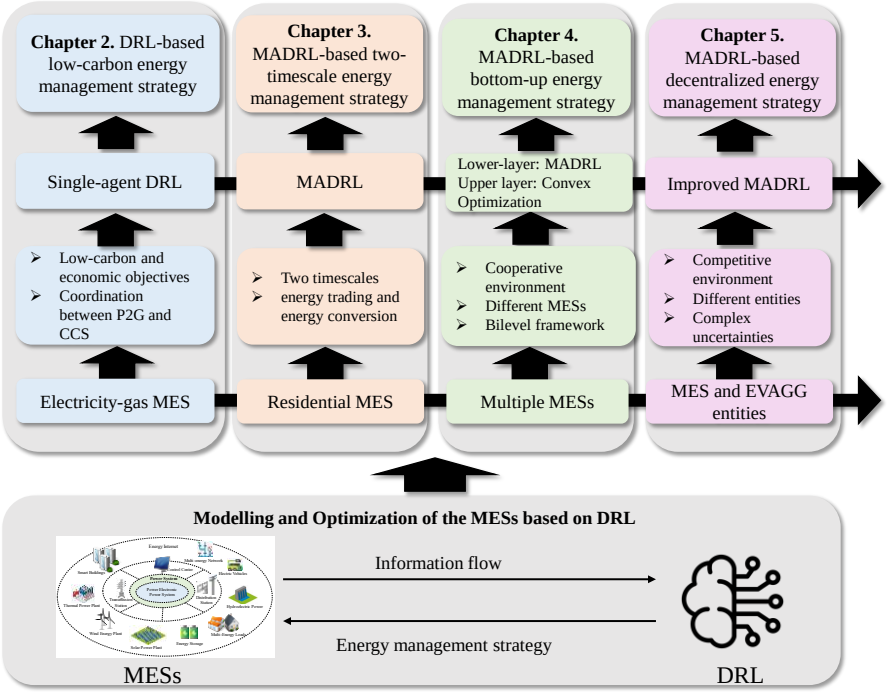


Figure 1-5 The contents of the thesis

As discussed earlier, modern MESs are complex and broad, which includes integrated energy networks, residential MESs, EI network including multiple MESs, and the MES interacting with other entities (e.g., EVAGG). Therefore, this thesis focuses on developing characterized energy management strategies for different types of MESs.

Specifically, Chapter 2 focuses on hourly-level low-carbon economic operation in an electricity-gas MES, which is a centralized management strategy. In Chapter 3, we transition to a residential MES, exploring a two-timescale energy management strategy to reduce energy costs while using MADRL to create individual strategies for an energy trading agent and an energy conversion agent. In Chapter 4, we further consider energy management strategies for an EI composed of multiple MESs, proposing a bottom-up management framework in a collaborative environment. Finally, in Chapter 5, we examine a competitive environment with EVAGG entities and investigate how to develop decentralized energy management strategies to maximize the EVAGGs' profits and minimize the energy cost of MESs. The system modeling of different MESs is provided in each chapter.

Furthermore, the methodology also follows a progressive approach. It starts with the single-agent DRL algorithm, moves to the MADRL algorithm, and finally involves improving the MADRL algorithm to enhance training effectiveness. Since solving energy management problems with DRL or MADRL requires transforming the original problem into an MDP or Markov game, the description of MDP or Markov game is essential. Although this may lead to structural repetition, each description is different, as it is specific to different energy management problems.

1.4. THESIS OUTLINE

The thesis is written based on the publications in the Ph.D. project, and is presented in the form of a collection of papers. The contents of this thesis are divided into two sections: a Report and Selected Publications.

Figure 1-6 outlines the structure of the thesis. Chapter 1 presents the research background, objectives and contributions. Chapter 2 investigates the low-carbon economic energy management strategy for the electricity-gas MES, where the coordination of P2G and CCS units is studied. In Chapter 3, two-timescale energy management strategy for the residential MES is elaborated to minimize operation cost. Chapter 4 exhibits a bottom-up energy management strategy for multiple MESs under the framework of EI. In Chapter 5, a decentralized energy management strategy for MES and EVAGG entities is investigated to maximize their profits. The conclusion is shown in Chapter 6.

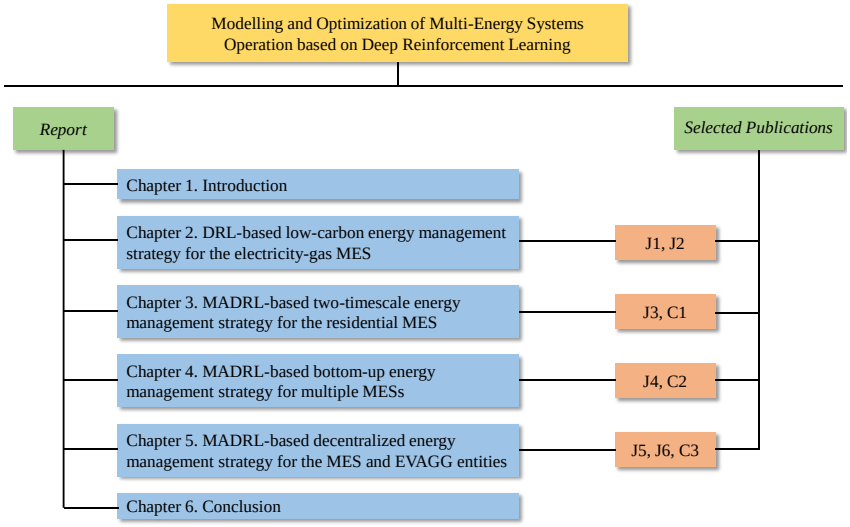


Figure 1-6 The outline of this thesis.

1.5. LIST OF PUBLICATIONS

Journal papers

J1. **B. Zhang**, X. Wu, A. Ghias and Z. Chen, "Coordinated Carbon Capture Systems and Power-to-Gas Dynamic Economic Energy Dispatch Strategy for Electricity-Gas Coupled Systems considering System Uncertainty: An Improved Soft Actor-Critic Approach," *Energy*, vol. 271, no. 126965, May 2023.

J2. **B. Zhang**, W. Hu, X. Xu, Z. Zhang and Z. Chen, "Hybrid Data-Driven Method for Low-Carbon Economic Energy Management Strategy in Electricity-Gas Coupled Energy Systems based on Transformer Network and Deep Reinforcement Learning," *Energy*, vol. 273, no. 127183, Mar. 2023.

J3. **B. Zhang**, X. Xu, W. Hu, and Z. Chen, "Two-Timescale Autonomous Energy Management Model based on Multi-Agent Deep Reinforcement Learning Approach for Residential Multicarrier Energy System", *Applied Energy*, vol. 351, no. 121777, Dec. 2023.

J4. **B. Zhang**, W. Hu, A. Ghias, X. Xu, and Z. Chen, "Multi-Agent Deep Reinforcement Learning based Distributed Control Architecture for Interconnected Multi-Energy MG Energy Management and Optimization," *Energy Conversion and Management*, vol. 277, no. 116647, Feb. 2023.

J5. **B. Zhang**, W. Hu, X. Xu, T. Li, Z. Zhang and Z. Chen, "Physical-Model-Free Intelligent Energy Management for a Grid-Connected Hybrid Wind-Microturbine-PV-EV Energy System via Deep Reinforcement Learning Approach", *Renewable Energy*, vol. 200, pp. 433-448, 2022.

J6. **B. Zhang**, W. Hu, D. Cao, A. Ghias, and Z. Chen, "Novel Data-Driven Decentralized Coordination Model for Electric Vehicle Aggregator and Energy Hub Entities in Multi-Energy System Using an Improved Multi-Agent DRL Approach," *Applied Energy*, vol. 339, no. 120902, Jun. 2023.

Conference papers

C1. **B. Zhang**, Z. Chen, and A. Ghias. "Deep Reinforcement Learning -based Energy Management Strategy for a MG with Flexible Loads," , *2023 the 7th International Conference on Power Energy Systems and Applications* (ICoPESA 2023)

C2. **B. Zhang**, Z. Chen, and A. Ghias. "A Data-Driven Approach towards Fast Economic Dispatch in Integrated Electricity and Natural Gas System", *2022 the 3rd International Conference on Power Engineering (ICPE 2022)*.

C3. **B. Zhang**, Z. Chen, X. Wu, D. Cao, and W. Hu. "A MATD3 -based Voltage Control Strategy for Distribution Networks Considering Active and Reactive Power Adjustment Costs", *2022 IEEE International Conference on Power Systems and Electrical Technology* (PSET 2022).

CHAPTER 2. A LOW-CARBON ENERGY MANAGEMENT STRATEGY FOR THE ELECTRICITY-GAS MES

The contents of Chapter 2 are based on the following two papers:

J1: B. Zhang, X. Wu, A. Ghias and Z. Chen, "Coordinated Carbon Capture Systems and Power-to-Gas Dynamic Economic Energy Dispatch Strategy for Electricity-Gas Coupled Systems considering System Uncertainty: An Improved Soft Actor-Critic Approach," *Energy*, vol. 271, no. 126965, May 2023.

J2: B. Zhang, W. Hu, X. Xu, Z. Zhang and Z. Chen, "Hybrid Data-Driven Method for Low-Carbon Economic Energy Management Strategy in Electricity-Gas Coupled Energy Systems based on Transformer Network and Deep Reinforcement Learning," *Energy*, vol. 273, no. 127183, Mar. 2023.

2.1. INTRODUCTION

The high penetration of wind power resulting high randomness and uncertainty pose significant challenges to investigate low-carbon economic operation strategy for MESs. Traditional model-based strategies rely on accurate modeling of uncertainties, which is often difficult to achieve. Therefore, a data-driven low-carbon operation strategy based on DRL is proposed in this chapter. In Section 2.2, the model of electricity-gas MES including objective function, constraints and coordination of P2G and CCS is established. Section 2.3 formulates the investigated problem as MDP, and presents an improved SAC algorithm with prioritized experience replay (PER). The effectiveness of the proposed strategy is verified in the simulation in Section 2.4. Conclusion is given in Section 2.5.

2.2. SYSTEM DESCRIPTION

Figure 2-1 illustrates the structure of the energy management of the electricity-gas MES, including cyber space and physical space [89],[90]. In the cyber space, energy information from the electricity-gas MES is collected before making energy management strategy. In the physical space, the electricity-gas MES is coupled by multiple components, including GT and P2G. Natural gas demand is met through purchases from the gas well and P2G. Electrical load is supplied by the main grid, wind power, CCS, GT, and coal-fired units. The detailed mathematical models can be found in Section 2 in [J1].

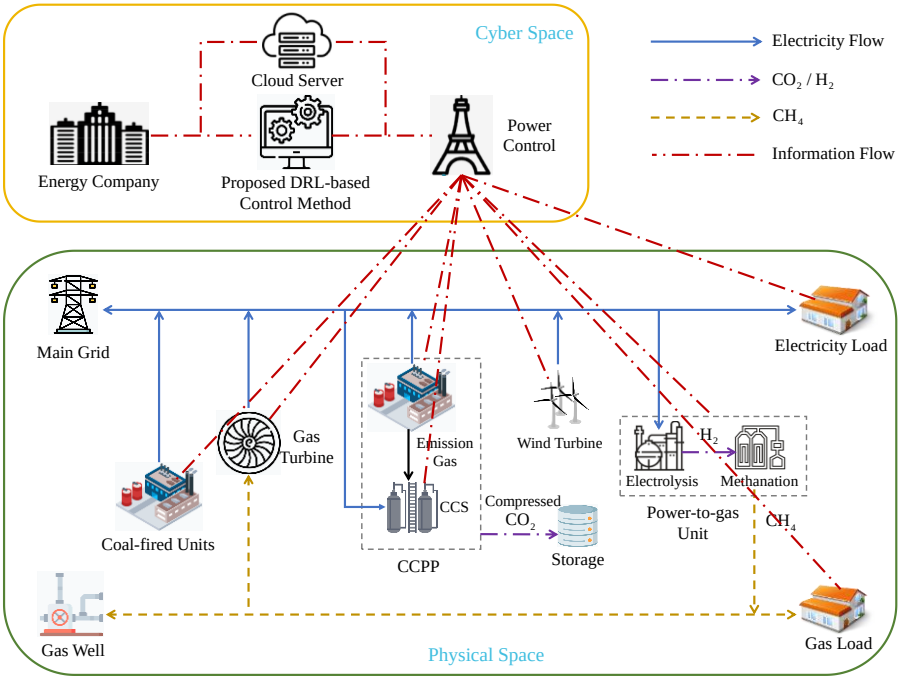


Figure 2-1 Structure diagram of the electricity-gas MES with P2G and CCPP units [J1].

2.2.1. CCPP OPERATION

CCPP model consists of a fossil fuel unit and a CCS unit. The fossil fuel unit provides electricity for both the load and the CCS. The electricity consumption of the CCS includes a fixed consumption and consumption related to CO_2 processing. The CCS captures CO_2 emitted by the fossil fuel unit, with a capture rate of 90%.

2.2.2. P2G OPERATION

The structure of P2G operation is illustrated in Figure 2-2. P2G is a coupling unit between electricity and gas systems, functioning as both a supply of natural gas and an electrical load. P2G operation encompasses two main processes: electrolysis and methanation [69]. H_2 storage is utilized to provide operation flexibility. The electricity consumption of P2G is used for water electrolysis, and the heat generated during the methanation process can be recycled by the CCS, thereby reducing the energy consumption of the CCS.

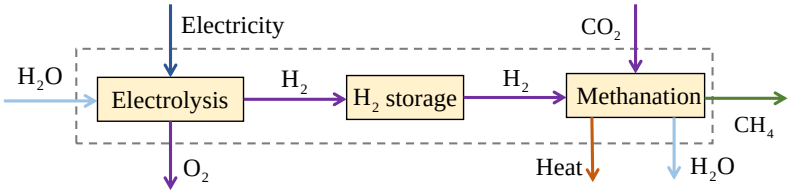


Figure 2-2 Operation scheme of the P2G facility [J1].

2.2.3. COORDINATION BETWEEN CCS AND P2G UNITS

Figure 2-3 details the coordination operation between CCS and P2G units. The coordination involves carbon capture process and P2G operation. In order to improve operation flexibility, CO_2 storage, H_2 storage and GT units are used. The H_2 storage stores excess H_2, which is then utilized by the GT during periods of high electricity prices to reduce energy expenses.

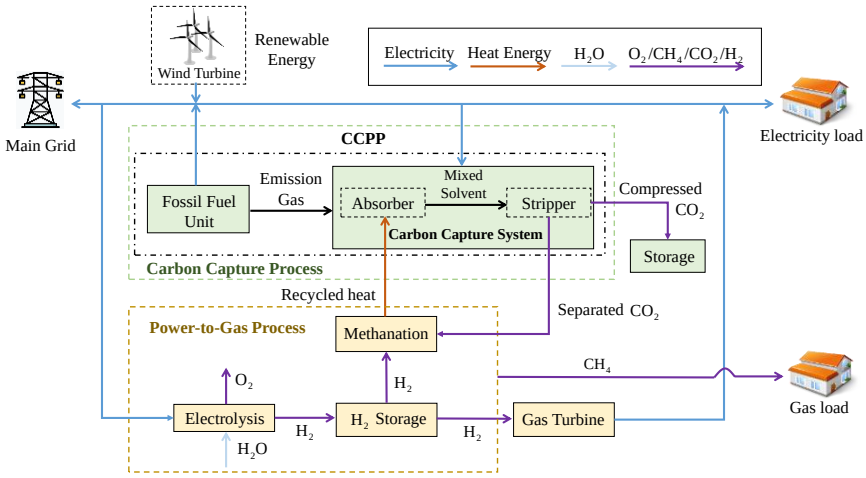


Figure 2-3 Coordination model of the P2G and CCS [J1].

The objective function of the low-carbon economic operation is to reduce the overall cost during [0, T], which is expressed as follows:

$$F = \min \sum_{t=0}^{T} C_t^{oc} + C_t^{cp} + C_t^{wp} \tag{2.1}$$

where $\{C^{oc}, C^{cp}, C^{wp}\}$ are the operation cost, the CO_2 processing cost and the penalty cost of wind curtailment. The constraints include unit operation constraints, energy

storage constraints, and gas and electricity balance constraints. T represents the time period, and since the goal is to optimize the daily operation cost, T is set to 24.

2.3. METHOD INTRODUCTION

In this section, the studied low-carbon economic operation problem is first formulated as MDP. Then, an improved SAC algorithm is presented. Finally, the DRL-based energy management method is proposed.

2.3.1. MDP FORMULATION

The MDP consists of four elements, including state set (S), action set (A), reward function (R), and state transition function (P).

1) State: The states $s_t \in S$ at time slot t are wind power output, electricity loads and gas loads.

2) Action: The actions $a_t \in A$ at time slot t are electricity output of CCPP and gas-fired generators, the electricity used to capture CO_2, the electricity output of GT and the CH_4 generation of the P2G.

3) Reward function: The reward is defined as the negative form of the objective function. The Reward $r_t \in R$ at time slot t is expressed as follows:

$$r_t = -(C_t^{oc} + C_t^{cp} + C_t^{wp}) \tag{2.2}$$

4) Transition probability: The state transition probability P represents the probability of the instant state moves to the next state. The state transition probabilities of wind power and load demands cannot be determined, but DRL can learn the relationship between states and actions through interactions with the environment.

5) System problem: The system operation optimization problem is indeed a stochastic optimization problem, as the dynamics of the MES (such as state transitions and rewards) involve inherent uncertainty. This uncertainty is captured using an MDP model, where the solution for the MDP is to find the optimal policy $\pi^*(a_t \mid s_t)$ to maximize the cumulative reward $R_t = \sum_{i=t}^{T} \gamma^{i-t} r_i$, where γ denotes the discount factor within $[0,1]$. The MDP framework allows us to model and solve this stochastic optimization problem by learning the optimal actions to take in each state to achieve the highest expected cumulative reward.

2.3.2. IMPROVED SAC ALGORITHM

The SAC algorithm based on actor-critic (AC) structure is used to solve the MDP problem. AC structure consists an actor network and a critic network, which is presented Figure 2-4. The actor parameterized by φ takes the state as input and outputs the action based on the policy π. The critic network parameterized by θ outputs the Q value $Q(s_t, a_t)$ which is used to direct the actor chooses action that has higher reward. Furthermore, SAC uses two critic networks to solve overestimation of Q-values, and employs target networks to improve training stability [91],[92].

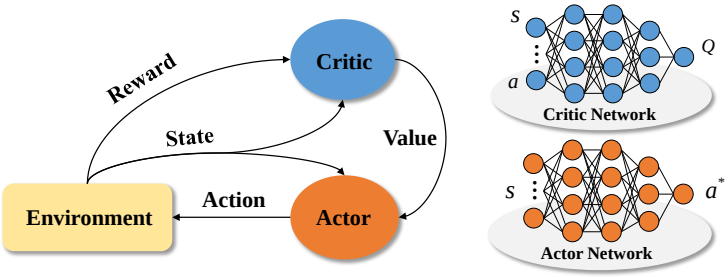


Figure 2-4 Structure of the actor-critic network [J1].

The maximum entropy is used to improve exploration and keep the algorithm from being stuck in local optima, which is expressed as follows:

$$J(\pi) = \sum_{t=0}^{T} E_{\pi}[r(s_t, a_t) + \alpha H(\pi(\cdot \mid s_t))] \tag{2.3}$$

where $H(\pi(\cdot \mid s_t))] = -\log \pi(a_t \mid s_t)$ is the entropy item, and α is the entropy coefficient. The critic and actor networks are updated based on gradient decent.

To improve training efficiency and convergence, PER mechanism is used. PER assigns higher weights to important samples obtained from the experience replay buffer, increases the sampling probability of those samples. The probability is expressed as follows:

$$p_j = \frac{\delta_j^{\lambda}}{\sum_k \delta_k^{\lambda}} \tag{2.4}$$

where λ denotes the priority control coefficient, and δ_j^{λ} is the weight of sample j.

The pseudocode of the PER-SAC algorithm is shown in Algorithm 1, and details about the PER-SAC can be found in Section 3 in [J1]:

Algorithm 1. Pseudocode of PER–SAC algorithm

// Start training
1. Initialize critic network, actor network and their target networks, respectively.
2. Initialize experience replay buffer.
3. **For** *each episode,* **do**
   // generate training data.
4. **For** *each time step,* **do**
5. Obtain action a_t at a given state s_t under policy π_φ.
6. Take action a_t and obtain reward r_t, and environment moves to next state s_t.
7. Store $\{s_t, a_t, r_t, s_{t+1}\}$ in experience buffer.
8. End **For**
   // Train neural networks
9. Sample from experience buffer based on probability p_j provided by PER.
10. **For** *each update step*, **do**
11. Update the critic networks:

$$\theta_i \leftarrow \theta_i - \lambda_Q \nabla_{\theta_i} J_Q(\theta_i) \text{ for } i \in \{1,2\}.$$

12. Update the actor network:

$$\varphi \leftarrow \varphi - \lambda_\pi \nabla_\varphi J_\pi(\varphi).$$

13. Update the entropy coefficient:

$$\alpha \leftarrow \alpha - \lambda_\alpha \nabla_\alpha J(\alpha).$$

14. Update each target network.

$$\overline{\theta}_i \leftarrow \tau\theta_i + (1-\tau)\overline{\theta}_i, \ \overline{\varphi} \leftarrow \tau\varphi + (1-\tau)\overline{\varphi}$$

15. End **For**
16. END

2.3.3. PER-SAC -BASED ENERGY MANAGEMENT STRATEGY

The framework of the proposed PER-SAC -based energy management strategy is shown in Figure 2-5. The environment is the electricity-gas MES, and the system operator is the DRL agent that determines the operation of controllable units. In the offline training, PER-SAC algorithm continuously updates network parameters to maximize cumulative rewards until it eventually converges. When deployed online, the actor network fixes its parameters and outputs the real-time decisions. Details about algorithm updating can be found in Section 3.2.2 in [J1].

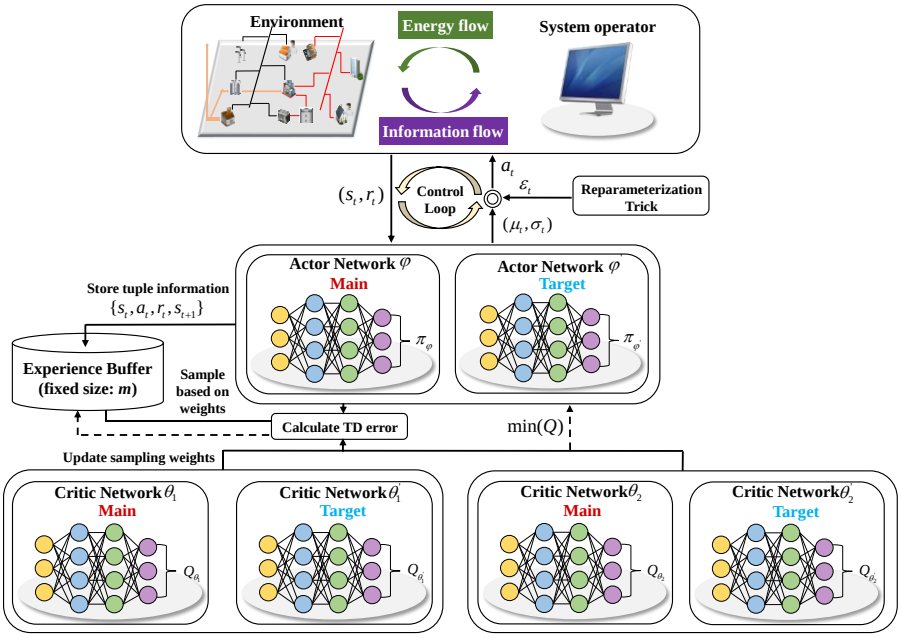


Figure 2-5 Framework of the proposed PER-SAC energy management strategy [J1].

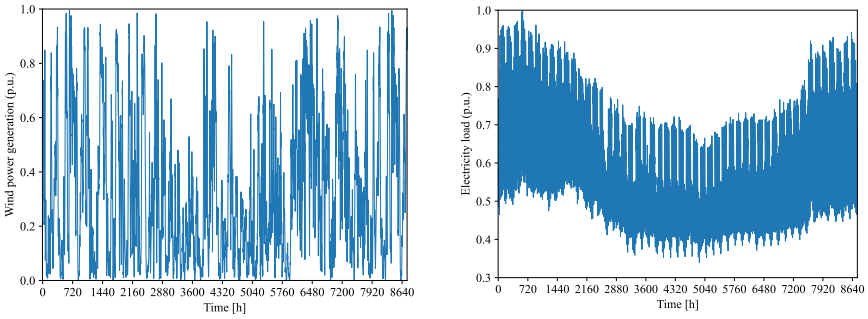
2.4. NUMERICAL SIMULATION

This simulation validates the effectiveness of the proposed energy management strategy through the use of real-world historical data.

2.4.1. CASE SETUP

The system example is an urban industrial park. With additional parameter details given in [J1], Figure 2-6 shows wind power generation and electricity load. Each time step is one hour, and each episode is one day (24 time steps). Five cases are set: Case 1: MES without CCS and P2G; Case 2: MES with P2G; Case 3: MES with CCS; Case 4: MES with CCS and P2G operating independently; Case 5: Same as Case 4, but with coordination between CCS and P2G units.

Due to the lack of coordination from P2G, Cases 3 and 4 face a challenge as all captured CO_2 needs to be transported and stored, with no efficient accommodation. Furthermore, in Case 4, the heat produced by the methanation reaction remains unused by CCS. In addition, in Cases 2 and 4, the production of methane (CH_4) involves using CO_2 directly obtained from atmosphere, rather than using a CO_2 storage device.



(a) (b)

Figure 2-6 The training dataset utilized to train the PER-SAC algorithm: (a) Wind power generation, (b) Electricity load [J1].

2.4.2. ALGORITHM TRAINING

The algorithm parameter settings are detailed in [J1]. The convergence of episodic average rewards of Case 5 is presented in Figure 2-7. The benchmarks are SAC algorithm and DQN algorithm. During the training process, DRL agent continuously adjusts weights of neural networks until the episodic reward reaches a stable state, indicating that an optimal strategy has been attained. As seen, the PER-SAC algorithm demonstrates a steadier and quicker average return than SAC. Moreover, due to the DQN algorithm's limitation to discrete action spaces, it achieves a lower average return.

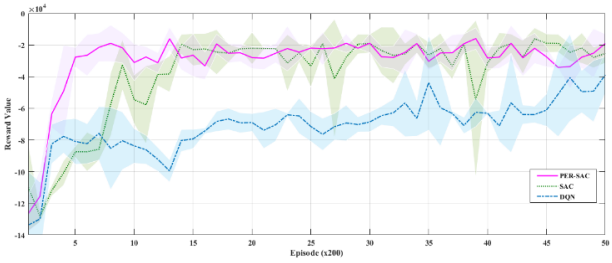


Figure 2-7 The convergence of cumulative rewards per episode in the PER-SAC, SAC and DQN algorithm [J1].

2.4.3. RESULTS ANALYSIS

A test day which is not included in the training dataset is used to test the well-trained low-carbon energy management strategy. The test data is shown in Figure 2-8. The peak values of the gas load, wind generation and electrical load are 1840 kcf , 210 MW and 580 MW, respectively.

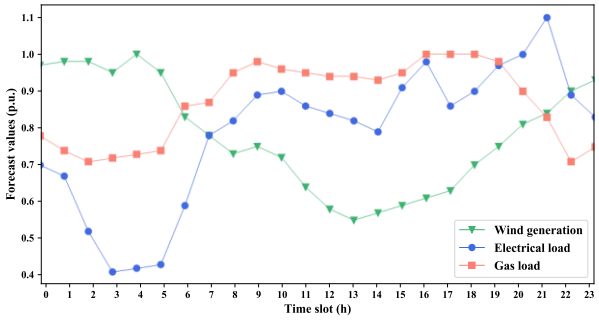


Figure 2-8 Load demand and wind power generation on a test day [J1].

2.4.3.1 Analysis of Wind Power Utilization Results

Wind power utilization results of 5 cases are shown in Figure 2-9. In Case 1, highest wind curtailment is observed during the early hours (1 to 6) due to the absence of P2G facility and CCS units for utilizing excess wind energy. While Cases 2 and 3 show a reduction in wind curtailment compared to Case 1, the complete absorption of surplus wind power is still not achieved, even with the inclusion of P2G facilities and CCS in Cases 4 and 5.

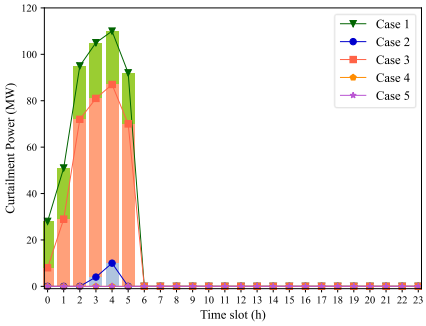


Figure 2-9 Wind power curtailment in Cases 1-5 [J1].

2.4.3.2 Operation of CCPP, P2G and GT units

Figure 2-10 displays the operation of CCPP. Between hours 1 and 6, there is a noticeable decrease in the net power of CCPP for Cases 3-5 compared to Cases 1-2, primarily due to the CCS. Therefore, approximately 20 MW of wind power is promptly utilized to provide electricity.

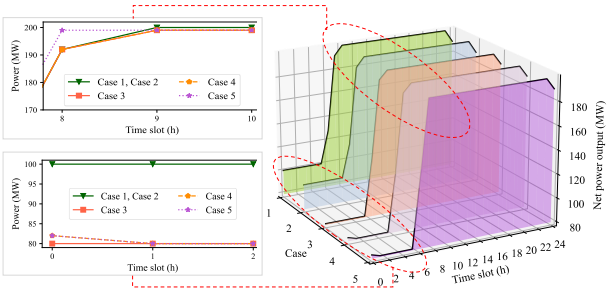


Figure 2-10 Operation results of CCPP [J1].

Furthermore, the P2G unit operation is shown in Figure 2-11. P2G operates only during periods of high wind power generation. Since both Case 4 and Case 5 include CCS, the CCPP will supply more electricity, thereby reducing the input power of P2G to lower operation costs.

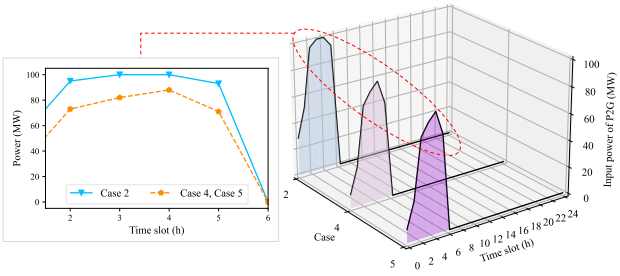


Figure 2-11 Operation of P2G unit in Case 2, 4 and 5 [J1].

Figure 2-12 displays the operation of GT and waste heat utilization in Cases 2, 4, and 5. In Cases 2 and 4, where there's no coordination of P2G and CCS, the produced H_2 is directly used for CH_4 synthesis. However, in Case 5, the presence of H_2 and CO_2 storage devices allows for the decoupling of the CH_4 synthesis process. Consequently, Case 5 offers two applications for the produced H_2 – methane synthesis or electricity generation through GTs. Notably, during peak electricity demand periods, specifically hours 15-16 and 20-21, some of the H_2 is utilized by GTs to produce additional electricity. To meet the rising gas demands at hours 11, 16, and 19, the Sabatier reaction is employed to lower gas supply costs. Furthermore, the heat from the reaction is recycled in the CCS to capture CO_2, reducing the generation costs for the CCPP.

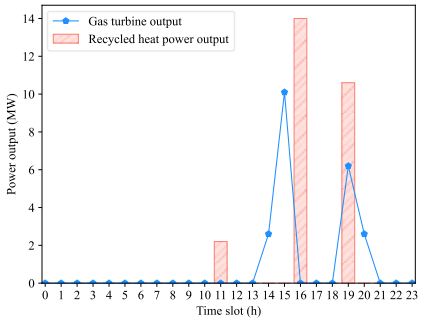


Figure 2-12 Operation of GT and the utilization of waste heat in Case 5 [J1].

2.4.4. ALGORITHM PERFORMANCE

To validate the generalization and robustness of the proposed strategy, a comparison analysis of the proposed strategy and scenario analysis (SA) over different forecast errors is conducted. The description of the SA can be found in Section 4.3 in [J1]. The prediction errors are generated following the normal distribution. The wind power profiles for the real-scenario and two prediction scenarios are presented in Figure 2-13. Cost comparison results for different scenarios over 14 consecutive days are presented in Figure 2-14. As seen, the operation cost and cumulative cost achieved by the proposed strategy are closed to the optimal results. The satisfactory results stem from the fact that DRL learns near-optimal policies from a large amount of historical data. As a result, DRL performs well in environments where the training and testing sets exhibit similar random characteristics. In contrast, SA algorithms rely on precise modeling of uncertainties, making their outcomes highly sensitive to prediction errors in the deterministic model.

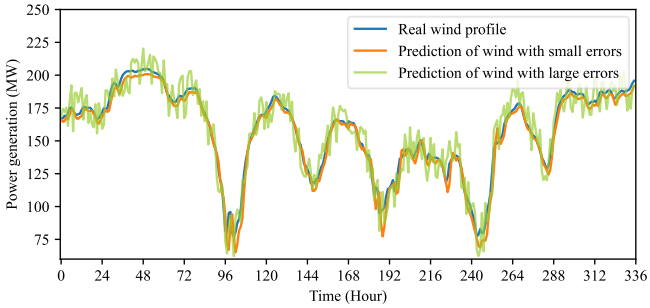
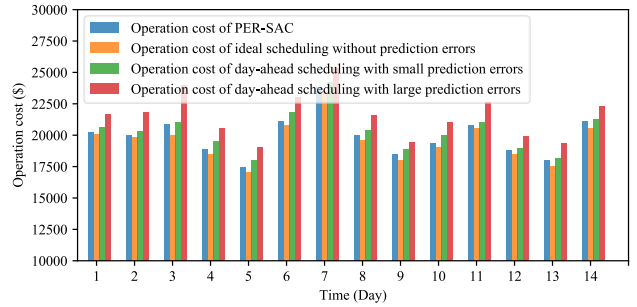
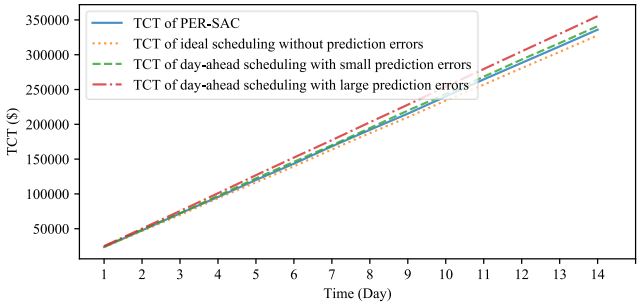


Figure 2-13 Wind power profiles for the real-scenario and two prediction scenarios [J1].



(a)



(b)

Figure 2-14 Cost for 14 test days: (a) operation cost; (b) cumulative cost.

The comparison results of different algorithm are presented in Table 2-1 and Table 2-2. Case 1 is the benchmark case. The total cost provided by SA method is 0.68% lower than that of PER-SAC algorithm. SA relies on the accurate uncertainty modeling and requires too much calculation time, but PER-SAC only requires the forward pass of the well-trained actor network during online operation, enabling real-time and continuous decision-making. Since DQN requires discretizing the action space, the cost results are unsatisfactory.

Table 2-1 Comparative analysis of total cost under different forecast accuracy [J1]

Case	Method	Cost ($/d)	Improvement (%)
Case 1	-	37,198.62	0
Case 5	PER–SAC	23,514.89	36.78
	DQN	36,213.65	2.65
	SA	23,264.23	37.46

Table 2-2 Computation performance of different algorithms [J1]

Method	Offline training time (s)	Online operation time (s)
PER-SAC	341.237	0.039
DQN	221.748	0.048
SA	-	1563.851

2.5. CONCLUSION

In this chapter, a DRL-based low-carbon economic energy management strategy for the electricity-gas MES is investigated. To solve the uncertainties, the studied problem is first formulated as MDP, and solved by the PER-SAC algorithm. Case study shows that the controlled units can flexibly adjust their operations to increase wind power utilization and reduce operation costs. In comparison with other algorithms, the proposed strategy effectively reduces operation costs and can provide real-time operation strategy.

CHAPTER 3. MADRL-BASED TWO-TIMESCALE ENERGY MANAGEMENT STRATEGY FOR THE RESIDENTIAL MES

The contents of Chapter 3 are based on the following two papers:

J3: B. Zhang, X. Xu, W. Hu, and Z. Chen, "Two-Timescale Autonomous Energy Management Model based on Multi-Agent Deep Reinforcement Learning Approach for Residential Multicarrier Energy System", *Applied Energy*, vol. 351, no. 121777, Dec. 2023.

C1: B. Zhang, Z. Chen, and A. Ghias. "Deep Reinforcement Learning -based Energy Management Strategy for a MG with Flexible Loads", *2023 the 7th International Conference on Power Energy Systems and Applications (ICoPESA 2023)*.

3.1. INTRODUCTION

Residential energy use accounts for a significant portion of total energy consumption. It's significant to investigate an effective residential energy management strategy to minimize energy costs. Residential energy management includes internal short-term energy conversion and external long-term energy trading. However, multiple uncertainties including long-term and short-term uncertainties make the energy management complex introducing high-dimension stochastic constraints. It's difficult to solve it by using single-agent DRL algorithm. Therefore, this chapter proposes an MADRL-based two-timescale residential energy management strategy, considering the hourly-ahead energy trade and the 15-minute-ahead energy operation. Section 3.2 describes the two-timescale energy management problem. In Section 3.3, MADRL algorithm is introduced, and case study including deterministic and stochastic studies is conducted in Section 3.4. Conclusion is given in Section 3.5.

3.2. MODEL DESCRIPTION

The architecture of the RMES is shown in Figure 3-1. Electricity and natural gas can be purchased from the electricity grid and gas network through the energy trading system. It can also sell any excess electricity that it produces. Rooftop PV panels are used to generate electricity. ES can make up for energy shortages by storing extra electricity. WE converts electricity into hydrogen. FC uses hydrogen to produce electricity and heat, and HB uses hydrogen for heating. Hydrogen can be stored in HT

for future needs. GB serves as an auxiliary heat source, using purchased natural gas to meet heating requirements. The feasibility of these units has already been demonstrated.

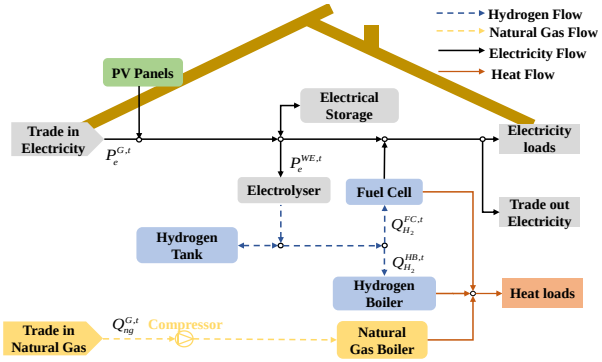


Figure 3-1 The architecture of the RMES [J3].

3.2.1. TWO-TIMESCALE ENERGY MANAGEMENT FRAMEWORK

The two-timescale energy management is illustrated in Figure 3-2. At each hour, the amounts of natural gas bought from the external gas network and electricity traded with the external grid are determined and remain unchanged until the next hour. Within each hour, the 15-minute operations of the coupling units are determined.

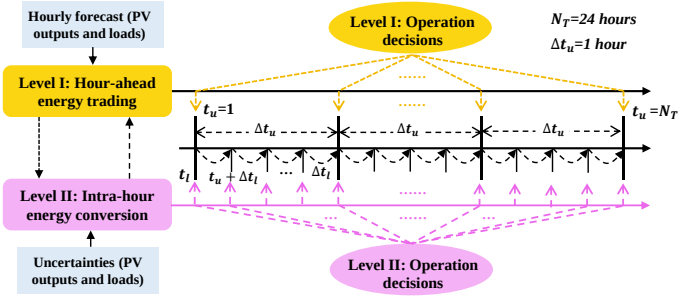


Figure 3-2 Two-timescale energy management framework [J3].

3.2.2. OBJECTIVE FUNCTION

This objective is to minimize the energy cost, which is outlined as follows:

$$F = \min \sum_{t=1}^{24} (\lambda_e^t P_e^{G,t} + \lambda_{ng}^t Q_{ng}^{G,t}) \tag{3.1}$$

where λ_e^t and λ_{ng}^t denote the prices of gas and electricity at time *t*; $P_e^{G,t}$ is the traded electricity; $P_e^{G,t} > 0$ refers to the action of purchasing electricity from the external power grid; and $P_e^{G,t} \leq 0$ refers to selling electricity on the wholesale market; $Q_{ng}^{G,t}$ represents the quantity of gas being exchanged.

3.2.3. COUPLING UNITS

The mathematical models of coupling units involved in the RMES are listed as follows:

$$P_e^{FC,t} = \eta_e^{FC} \times Q_{H_2}^{FC,t} \tag{3.2}$$

$$Q_h^{FC}(t) = \eta_h^{FC} \times Q_{H_2}^{FC,t} \tag{3.3}$$

$$Q_{H_2}^{WE}(t) = \eta^{WE} \times P_e^{WE,t} \tag{3.4}$$

$$Q_h^{GB,t} = \eta_{ng}^{GB} \times Q_{ng}^{G,t} \tag{3.5}$$

$$Q_h^{HB,t} = \eta_{H_2}^{HB} \times Q_{H_2}^{HB,t} \tag{3.6}$$

where $Q_{H_2}^{FC,t}$ represents the hydrogen input, $P_e^{FC,t}$ represents the electrical output, and η_e^{FC} represents the electricity conversion coefficient for the FC at time t; $Q_h^{FC}(t)$ and η_h^{FC} signify the heat output and heat conversion coefficient of the FC, respectively, at time t. For the WE at time t, $P_e^{WE,t}$ is the electricity input, $Q_{H_2}^{WE}(t)$ is the hydrogen output, and η^{WE} is the conversion coefficient. Eqs. (3.5)-(3.6) describe the conversion functions for the GB and HB. Here, $Q_h^{GB,t}$ and η_{ng}^{GB} refer to the heat generated and the heat conversion coefficient of the GB at time t. For the HB at time t, $Q_{H_2}^{HB,t}$, $Q_h^{HB,t}$ and $\eta_{H_2}^{HB}$ represent the hydrogen input, heat outflow, and hydrogen conversion coefficient, respectively.

The mathematical models of ES charging/discharging are defined as follows:

$$E_e^{t+1} = E_e^t + P_e^{ES,t}(I_{(P_e^{ES,t}>0)}\eta_{e,ch}^{ES} - \frac{I_{(P_e^{ES,t}\leq 0)}}{\eta_{e,dis}^{ES}})\Delta t \tag{3.7}$$

$$E_{H_2}^{t+1} = E_{H_2}^{t} + Q_{H_2}^{HT,t}\left(I_{(Q_{H_2}^{HT,t}>0)}\eta_{H_2,in}^{HT} - \frac{I_{(Q_{H_2}^{HT,t}\le 0)}}{\eta_{H_2,out}^{HT}}\right) \tag{3.8}$$

$$P_e^{ES\min} \le P_e^{ES,t} \le P_e^{ES\max} \tag{3.9}$$

$$Q_{H_2}^{HT\min} \le Q_{H_2}^{HT,t} \le Q_{H_2}^{HT\max} \tag{3.10}$$

$$0 \le E_e^t \le E_e^B \tag{3.11}$$

$$0 \le E_{H_2}^t \le E_{H_2}^B \tag{3.12}$$

Eqs (3.7)-(3.8) illustrate the evolution of energy levels in the ES and HT. $P_e^{ES,t}$ and $P_{H_2}^{HT,t}$ indicate the charging/discharging power and hydrogen at time t, respectively. $\eta_{e,ch}^{ES}$ and $\eta_{e,dis}^{ES}$ represent the charging coefficient and discharging coefficient of the ES, respectively. Similarly, $Q_{H_2}^{HT,t}$, $\eta_{H_2,in}^{HT}$ and $\eta_{H_2,out}^{HT}$ represent the hydrogen input, hydrogen outflow, and the coefficients for the inflow and outflow of the HT, respectively. Eqs. (3.9)-(3.10) establish the constraints on the rate at which electricity can be charged and discharged, as well as the limitations on the flow of hydrogen in and out at time t. Similarly, equations (3.11) and (3.12) represent the constraints of the capacity of ES and HT at time t, where E_e^B and $E_{H_2}^B$ represent the maximum capacity of the ES and the HT.

The energy balance among electricity, hydrogen and heat at time t is expressed below, and more details can be found in [J3]:

$$P_e^{ES,t}\Delta t + P_e^{WE,t}\Delta t + P_e^{L,t} = P_e^{PV,t} + P_e^{G,t} + P_e^{FC,t}\Delta t \tag{3-13}$$

$$Q_{H_2}^{HT,t}\Delta t + Q_{H_2}^{FC,t}\Delta t + Q_{H_2}^{HB,t}\Delta t = Q_{H_2}^{WE,t}\Delta t \tag{3-14}$$

$$Q_h^{L,t} = Q_h^{GB,t}\Delta t + Q_h^{FC,t}\Delta t + Q_h^{HB,t}\Delta t \tag{3-15}$$

3.3. METHOD INTRODUCTION

This section presents the proposed MADRL -based residential energy management strategy.

3.3.1. MARKOV GAME FORMULATION

The MADRL application for decision-making in energy trading and conversion is facilitated through Markov game formulation, which integrates states, actions, and rewards [91].

1) Environment and agent: The environment is RES. Two agents including an energy trading agent and energy conversion agents are set. The energy conversion agents are FC, WE and HB, respectively.

2) State: The states of the energy trading agent are load demands and energy tariffs. The states of the energy conversion agent are load demands and the excess capacity of the HT and ES.

3) Action: The energy trading agent's actions include the amount of gas and electricity purchased. The energy conversion agent's actions are outputs of the FC, HB, and WE.

4) Reward: The reward function of the energy trading agent at time t is expressed as follows:

$$r_{et}^{t} = -C_{et}^{eco,t} - C_{e,et}^{pen,t} - C_{th,et}^{pen,t} \tag{3.16}$$

where $C_{et}^{eco,t}$ is the economic cost, and { $C_{e,et}^{pen,t}$, $C_{th,et}^{pen,t}$ } represent penalties for a shortage in power and heat supply, respectively. The reward function of the energy conversion agent at time t is defined as follows:

$$r_{ec}^{i,t} = r_{agent,ec}^{i,t} + r_{sys,ec}^{t}, i \in (FC, WE, HB) \tag{3.17}$$

The system-level reward function is used to solve constraint violations, as presented in:

$$r_{sys,ec}^{t} = (r_{sys,ec}^{dif,t} + r_{sys,ec}^{ES,t} + r_{sys,ec}^{HT,t}) / 3 \tag{3.18}$$

where $r_{sys,ec}^{dif,t}$ describes energy imbalance of supply and demand, and $r_{sys,ec}^{ES,t}$ and $r_{sys,ec}^{HT,t}$ are constraints violations of ES and HT, respectively. Details can be found in [J3].

3.3.2. MADRL TRAINING AND EXECUTION

Figure 3-3 shows the structure of the proposed MADRL-based two-timescale residential energy management strategy. During the training phase, each agent needs both its specific local measurements and the contribution χ_i from other agents for estimating the Q-value. In the execution phase, the trained actor networks with set weights are used for real-time execution, while the critic networks remain inactive.

Each hour begins with the upper-level agent implementing its policy $a_{et}^{t_u}$ as derived from its actor network $\mu^{\theta_{et}^*}$, following its specific observation $s_{et}^{t_u}$. Throughout the hour, each energy conversion agents executes its optimal policy $a_{1:N-1}^{t_l}$ as provided by its actor network $\mu^{\theta_{1:N-1}^*}$, based on local observations [92]. The pseudocodes of the proposed strategy during the training stage and execution stage are provided in Algorithm 1 and Algorithm 2, respectively. Details about the proposed algorithms are given in Section 3 in [J3].

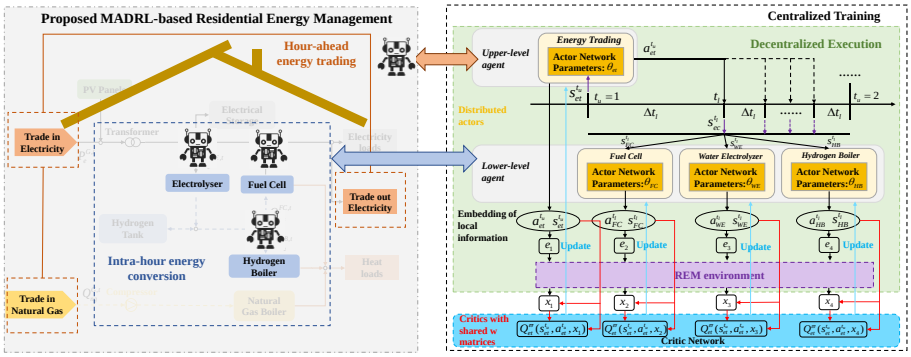


Figure 3-3 Structure of the MADRL-based two-timescale energy management strategy [J3].

Algorithm 1 Offline training of the proposed strategy

1. Initialize RMES environment
2. **for** *each episode* **do**
   - 2.1. Obtain initial $s_{et}^{t_u}$ for hourly-ahead trading agent
   - 2.2. Choose action $a_{et}^{t_u}$ based on the trading
   - 2.3. Execute action $a_{et}^{t_u}$ and obtain the reward $r_{et}^{t_u}$
   - 2.3. **for** *each time step* $t_l = 15$ from 1 to $60/\Delta t_l$ **do**:
     - 3.3.1. Choose action $a_{1:N-1}^{t_l}$ for the energy conversion agents based on its observation $s_{1:N-1}^{t_l}$
     - 3.3.2. Perform action $a_{1:N-1}^{t_l}$ and obtain the reward $r_{1:N-1}^{t_l}$ and the next state $s_{1:N-1}^{t_l+1}$
     - 3.3.3. Store experience in replay buffer
     - 3.3.4. Sample a random batch of L transitions from replay buffer
     - 3.3.5. for each sampled tuple
     - 3.3.6. Update the critic and actor weights based on backpropagation
     - 3.3.7. Update the target network weights
   - **end for**
3. **end for**

Algorithm 2 Online execution of the proposed strategy

1. Utilize the well-trained weights from the actor networks of all agents
2. **for** each episode **do**
   - 2.1. Obtain initial states for hourly-ahead trading agent
   - 2.2. Choose the trading agent's $a_{et}^{t_u} = \mu(s_{et}^{t_u}, \theta_{et}^*)$

2.3. Execute the trading agent's action and obtain the reward $r_{et}^{t_u}$

2.4. **for** *each time step* $t_l = 15$ from 1 to $60 / \Delta t_l$ **do**:

2.4.1. Choose energy conversion agents' actions based on its local observation

2.4.2. Execute the energy conversion agent's actions and obtain the reward $r_{ec}^{t_l}$ for the energy conversion agents

end for

end for

3.4. CASE STUDY

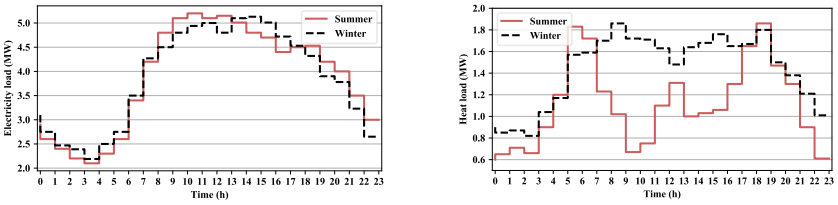
The performance of the MADRL-based two-timescale energy management strategy is assessed in this simulation. Training data, including PV generation and residential household demand at hourly intervals, are sourced from real-world datasets [93]. Information on hourly electricity and natural gas prices is obtained from sources [94], [95].

3.4.1. SIMULATION RESULTS ANALYSIS

First, the MADRL is implemented in the RMES for a deterministic study. Table 3-1 introduces various scenarios for the deterministic study, featuring different seasonal load profiles and electricity and gas price variations. Figure 3-4 displays the summer and winter load curves. Three models of pricing energy are shown in Figure 3-5: The three types of prices are PV-EP (peak-valley price), Ex-EP (extreme price), and RT-EP (real-time pricing). Furthermore, three distinct scenarios are being considered regarding gas prices: a baseline case (Ben-GP) in which the gas price remains constant at \$26/MWh; an extreme case (Ex-GP) where the price is fixed at \$50/MWh; and a peak-valley scenario (PV-GP) in which the gas price is \$20/MWh during non-peak hours (7-18) and \$30/MWh during peak hours.

Table 3-1 Different scenarios specifications [J3].

Scenarios	**Load profile**		**Price curve**	
	Electricity	Heat	Electricity	Gas
Ben	Winter	Winter	RT-EP	Ben-GP
PV-EP	Winter	Winter	**PV-EP**	Ben-GP
Ex-EP	Winter	Winter	**Ex-EP**	Ben-GP
Summer	**Summer**	**Summer**	RT-EP	Ben-GP
Ex-GP	Winter	Winter	RT-EP	**Ex-GP**
PV-GP	Winter	Winter	RT-EP	**PV-GP**



(a) *(b)*

Figure 3-4 Electricity and heat load profiles exhibit seasonal variations: a) Electricity demand; b) heat demand [J3].

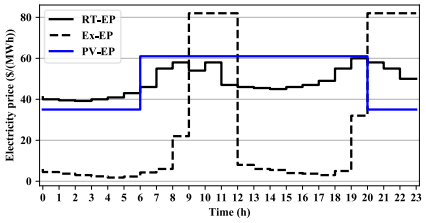
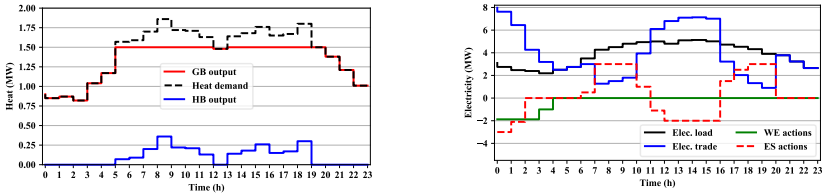


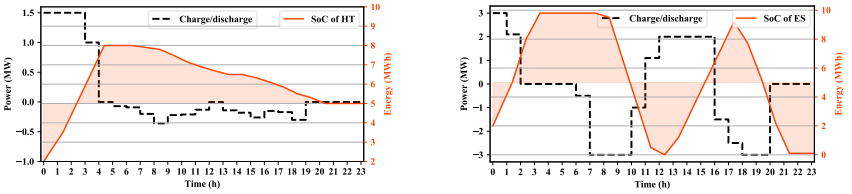
Figure 3-5 Electricity price profiles in different market scenarios [J1].

1) The benchmark scenario: Figure 3-6 illustrates the supply of heat and electricity demands, while Figure 3-7 shows the state of charge (SoC) of ES and HT units. The energy trading agent gives GB priority for heating since gas price is cheaper than electricity. However, during peak heat demand, GB's output falls short, prompting activation of the HB to tap into HT hydrogen. The FC agent remains inactive when demands are met. During the electricity price is low, extra power from the grid charges the ES and produces hydrogen via WE. ES balances electricity during peak electricity prices, while minimal electricity is acquired to meet load demand by the energy trading agent, shutting down the WE agent.



(a) *(b)*

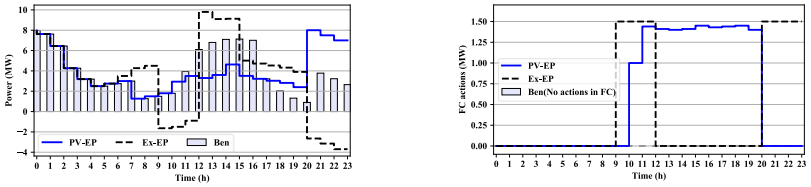
Figure 3-6 Agent actions in the Ben scenario: a) heat supply; b) electricity supply [J3].



(a) *(b)*

Figure 3-7 SoC changes of the HT and ES in case Ben: a) SoC of HT; b) SoC of ES [J3].

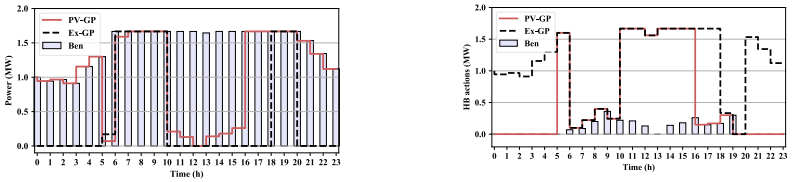
2) Impact of energy prices: Figure 3-8 illustrates the actions of the energy trading and FC agents in the PV-EP and Ex-EP scenarios. In comparison to other scenarios, the Ex-EP scenario experiences a surge in the amount of electricity acquired from the grid during periods when electricity prices are low. The trading agent opts to sell electricity by discharging the ES and working together with the FC agent when power prices in the external electricity market surge to $80/MWh. In the PV-EP scenario, direct purchases are used to meet the electricity demand between the hours of 0–6 and 21–23. Moreover, during high daytime electricity prices and load, the FC begins supplying power in periods 11-20, while the ES is prioritized for power supply during 7-10.



(a) *(b)*

Figure 3-8 The actions of agents against electricity price trends: a) energy trading agent; b) FC agent [J3].

The behaviors of the HB and energy trading agents in the PV-GP and Ex-GP scenarios are contrasted in Figure 3-9. In Ex-GP, gas purchases from the external network are minimal compared to other scenarios, yet occur during periods 7-10 and 19-20 due to elevated electricity prices. The HB agent in Ex-GP predominantly handles heat supply during peak hours, such as 0-6, 11-18, and 21-23, utilizing hydrogen from the WE as it's more cost-effective than buying natural gas, given the lower real-time electricity prices. In PV-GP, the HB agent reduces its output during 17-19 as the hydrogen supply depletes. Consequently, the energy trading agent opts for natural gas purchases, finding it more economical than producing hydrogen with WE when electricity prices exceed those of gas.



(a) *(b)*

Figure 3-9 The actions of HB agents and energy trading under different gas prices: a) energy trading agent; b) HB agent [J3].

3) Impact of load profile: Table 3-2 compares the energy costs of different load profiles. The Ben case incurs an overall energy cost that is 12.45% higher than in the Summer Scenario. In the Ben Scenario, the greater heat demand results in elevated natural gas procurement. Furthermore, the HB and FC agents opt to provide heat during the peak demand hours in the Ben scenario, necessitating extra electricity for hydrogen generation.

Table 3-2 Cost comparison under the Ben and Summer scenarios [J3].

Scenarios	**Cost ($)**	
	Electricity	Natural Gas
Ben	311.62	559.61
Summer	297.89	476.91

3.4.2. ALGORITHM COMPARISION

The algorithm comparison is conducted in the 50 random scenarios, and the average power imbalances are presented in Figure 3-10. The proposed method achieves the smallest amount of power imbalance. The MAQ strategy, which is based on Q-learning and multi-agent system, performs the worst due to the need to discretize both state and action spaces. The MADDPG-I strategy does not consider the system-level reward, with agents only optimizing their own rewards. The MADDPG-C strategy uses the same reward settings as the proposed method but does not account for the contributions of other agents when calculating Q-values. Furthermore, the proposed strategy has the lowest energy cost and is closest to the theoretical optimum among the four strategies. Details about the comparison algorithms are provided in Section 4 in [J3].

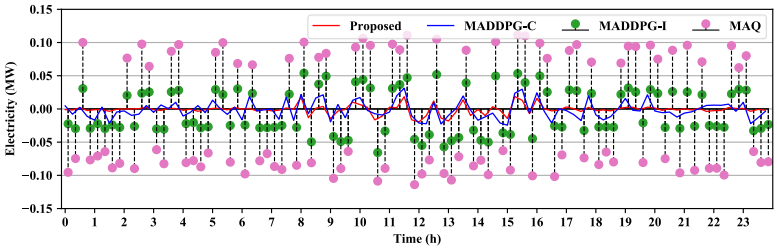


Figure 3-10 Power imbalances of different strategies [J3].

Table 3-3 Energy cost and training time of different strategies [J3]

Method	Energy cost ($)	Total training time (min)
MAQ	541	29
MADDPG-I	493	51
MADDPG-C	432	46
Proposed	421	51
Theoretical benchmark	413	-

3.5. CONCLUSION

In this chapter, a two-timescale residential energy management strategy based on MADRL is proposed to optimize the energy purchase and energy operation costs. Deterministic study on different load profiles, gas and electricity prices validates the effectiveness of the proposed energy management strategy. Furthermore, the superiority and robustness of the proposed strategy is verified compared to other MARL strategies on a stochastic study.

CHAPTER 4. MADRL-BASED BOTTOM-UP ENERGY MANAGEMENT STRATEGY FOR MULTIPLE MESS

The contents of Chapter 4 are based on the following two papers:

J4: B. Zhang, W. Hu, A. Ghias, X. Xu, and Z. Chen, "Multi-Agent Deep Reinforcement Learning based Distributed Control Architecture for Interconnected Multi-Energy MG Energy Management and Optimization," *Energy Conversion and Management*, vol. 277, no. 116647, Feb. 2023.

C2: B. Zhang, Z. Chen, and A. Ghias. "A Data-Driven Approach towards Fast Economic Dispatch in Integrated Electricity and Natural Gas System", *2022 the 3rd International Conference on Power Engineering (ICPE 2022)*.

4.1. INTRODUCTION

Previous contents have only considered energy management strategies for individual MESs and have not addressed cooperation between multiple MESs. The EI concept enabled by ERs facilitates energy-sharing among MESs. However, conventional EI energy management framework is top-down and centralized, which is susceptible to the single-point failure and heavy computational burden. This chapter presents a MADRL-based bottom-up EI framework to solve these issues. Section 4.2 presents the mathematical model of the investigated EI system. The proposed MADRL-based bottom-up energy management strategy is provided in Section 4.3. Case study is conducted in Section 4.4. Conclusion is given in Section 4.5.

4.2. SYSTEM DESCRIPTION

4.2.1. SYSTEM ARCHITECTURE

As depicted in Figure 4-1, the EI system comprises two layers: a bottom layer consisting of several MGs, each linked to a local ER, and an upper layer where the ER network connects with the main grid. To achieve cost minimization, each MGs, based on its local data, calculates and reports its power exchange amount. The ERs at the upper layer utilize the power exchange data from the MGs to assess energy transactions with the main grid and efficiently manage power distribution among themselves.

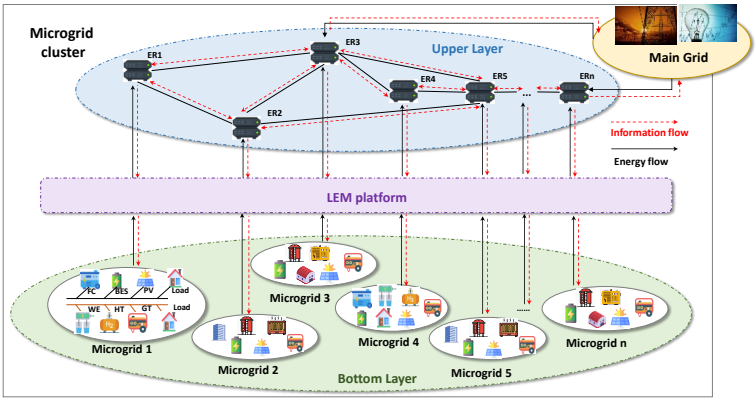
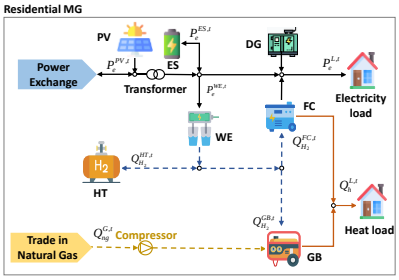
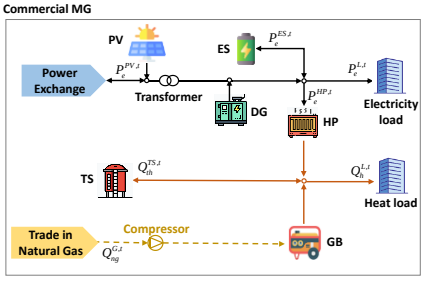


Figure 4-1 The architecture of double-layer EI system [J4].

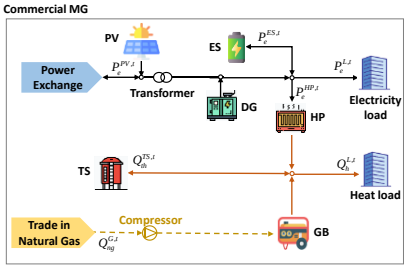
The platform for local energy management is known as local energy management (LEM). Every MG, indicated by MG_i, has a different composition. A residential MES, for instance, can include heat and electricity demands in addition to HT, FC, WE, distributed generator (DG), and ES system. The structures of residential, commercial, and industrial MGs are shown in Figure 4-2. TS is the thermal storage system, HP is the heat pump, and CHP is the combined heating and power plant. The mathematical models of different MGs can be found in Section 2 in [J4].



(a)



(b)



(c)

Figure 4-2 Three different MGs: (a) residential, (b) commercial, and (c) industrial [J4]

4.2.2. MODELS OF BOTTOM LAYER AND UPPER LAYER

4.2.2.1 Bottom Layer

In the bottom layer, each MG is viewed as a separate entity, focusing on its specific objectives over a given time horizon $[0,T]$. These objectives encompass achieving a balance between energy production and consumption while minimizing operation costs:

$$J_{i_1}^{bottom} = \sum_{t=0}^{T}(C_i^{FC,t} + C_i^{DG,t} + C_i^{HT,t} + C_i^{ES,t} + C_i^{WE,t} + C_i^{GB,t} + C_{i,ng}^{G,t}) \tag{4.1}$$

$$J_{i_2}^{bottom} = \sum_{t=0}^{T}(C_i^{HP,t} + C_i^{DG,t} + C_i^{TS,t} + C_i^{ES,t} + C_i^{GB,t} + C_{i,ng}^{G,t}) \tag{4.2}$$

$$J_{i_3}^{bottom} = \sum_{t=0}^{T}(C_i^{CHP,t} + C_i^{TS,t} + C_i^{ES,t} + C_i^{GB,t} + C_{i,ng}^{G,t}) \tag{4.3}$$

Eqs. (4.1) to (4.3) define the cost functions for residential, commercial, and industrial MGs, respectively. The objective function of the MG_i in bottom layer is defined as:

$$\min_{u_i \in \mathcal{U}} \mathbb{E}[J_i^{bottom}], \quad s.t.\ \text{Eqs.}(4.1)-(4.3) \tag{4.4}$$

where the expectation $\mathbb{E}$ describes randomness, and u_i is the control variables. The constraints include energy converter constraints, energy storage constraints and energy balance constraints, which can be found in [J4]. Upon solving the optimization problem presented in Eq. (4.4), the power exchange data is ascertained and then conveyed to the upper-layer ER network.

4.2.2.2 Upper Layer

The upper layer analyzes power exchange data from the bottom layer to determine the optimal power allocation between the ERs and the main grid [83]. The objective of the upper-layer cost J^{upper} is defined as:

$$J^{upper} = \sum_{t=0}^{T}[\sum_{i\in\mathcal{V}}(\varepsilon_e^{s,t} I_{(P_i^{G,t}>0)} - \varepsilon_e^{p,t} I_{(P_i^{G,t}\le 0)})P_i^{G,t} + \sum_{(i,j)\in\xi}\mu_{ij}(P_{i,j}^{ER,t})^2] \tag{4.5}$$

In this context, the initial term denotes the profits from power trading between ER_i and the main grid at time t. Here, $\varepsilon_e^{s,t}$ and $\varepsilon_e^{p,t}$ denote selling electricity price and buying electricity price, respectively, at time t. The second term in Eq. (4.5) relates to the transmission cost incurred when $P_{i,j}^{ER,t}$ is transmitted over the link between ER_i and ER_j, where μ_{ij} is the cost coefficient for this transmission. The constrains can be found in [J4]. The optimization objective of the upper layer is shown:

$$\min_{P_i^{G,t}, P_{i,j}^{ER,t}} \mathbb{E}[J^{upper}], \quad s.t.\ \text{Eq.}(4.5)-(4.7) \tag{4.6}$$

4.3. METHOD INTRODUCTION

A Markov game is used to formulate the investigated problem. A brief introduction is given to the twin delayed deep deterministic policy gradient (TD3) algorithm. Finally, a decentralized energy management strategy is presented through the application of the multi-agent attention twin delayed deep deterministic policy gradient (MAATD3) algorithm.

4.3.1. MARKOV GAME FORMULATION

The bottom-layer optimal energy management problem is formulated as a Markov game, incorporating state, action, reward, and state transition probabilities for the agents.

1) Environment and Agents: The environment is bottom-layer multiple MGs, and each MG is set as an agent.

2) State: The state set of the bottom-layer MG cluster includes all MG state information, represented as $s^t = \{s_1^t, s_2^t, .., s_n^t\} \in \mathcal{S}$. The states of MG_i at the time t are PV generation, load demands, SoC and energy prices.

3) Action: The action set includes all control variables of MGs, denoted as $a^t = \{a_1^t, a_2^t, .., a_n^t\} \in \mathcal{A}$. The actions of the residential MG at time t are the output of WE, FC, GB and DG, the actions of the commercial MG at time t are the output of HP, TS and DG, and the actions of the industrial MG at time t are the output of CHP and TS.

4) Reward function: The reward set consists of all reward values of MGs, represented as $r^t : \{r_1^t, r_2^t, .., r_n^t\} \in \mathcal{R}$. The reward value at time t is the negative form of its cost function, as given by: $r_i^t \in \{-J_{i_1}^{bottom}, -J_{i_2}^{bottom}, -J_{i_3}^{bottom}\}$

5) State transition probability: This function $\mathcal{P}(s^t \mid s^{t-1}, a^t)$ describes the probability of the agent *i* moving to the next state s_i^{t+1} after performing an action a_i^t in the current state s_i^t. SoC transition functions can be available. However, the transition functions for PV generation, loads and energy prices are not specified.

6) System problem: The target of the DRL agent is to find the optimal control policy π^* to maximize the expected total reward across a specific time horizon *T*, as given by: $\mathbf{P1}: \max_{\pi} R_t = \mathbb{E}\left[\sum_{\tau=0}^{T} \gamma^\tau r^{t+\tau+1}\right]$. To address this, the employed DRL algorithm allows for learning from historical data and managing partially observable transition functions.

4.3.2. TD3 ALGORITHM

TD3 algorithm contains fives parts: experience replay buffer, target networks, double networks, “delayed” policy updates, and target policy smoothing. For training a DNN-based approximator, an experience replay buffer is employed. This buffer stores a substantial number of historical experiences, serving as a dataset. TD3 employs dual critic networks, known as "twin" networks, to learn two separate Q-functions. It then uses the lower of the two Q-values to minimize the error function, effectively addressing the overestimation issue. TD3 ensures the stability of the Q-value by updating the policy and target networks less frequently, specifically only after each update of the Q-value function. To prevent the policy from becoming brittle due to inaccurate approximations of the Q-function for certain actions, it incorporates clipped Gaussian noise into the target action. Algorithm details can be found in Section 3 in [J4].

4.3.3. PROPOSED MAATD3 METHOD

In the context of multiagent environment, where each MG in the cluster represents an agent (resulting in *n* agents for *n* MGs), we propose the MAATD3 method to develop effective strategies. MAATD3, which blends the multiagent TD3 framework with an attention mechanism, adopts a structure with centralized training and decentralized execution. This means that while the Q-value is computed centrally, policy execution by the agents is decentralized. Furthermore, the critic network incorporates an attention mechanism that allows it to selectively integrate relevant information from other agents, hence improving the accuracy of its Q-value estimations.

4.3.3.1 Attention-based Q-Value Estimation

In earlier MADRL methods, such as MADDPG, the Q-value estimation for agents necessitates the inclusion of all agents' states $\{s_1^t, s_2^t, .., s_n^t\}$ and actions $\{a_1^t, a_2^t, .., a_n^t\}$ as inputs. If we assume that the dimensions of actions and states are identical, the input dimension for critics becomes $n|s_i + a_i|$, which can lead to a significant

computational challenge in policy learning. This challenge escalates as the number of states, actions, and agents increases. In addition, the concepts of state and action are related to the personal physical characteristics of an agent, such as energy preferences and operational portfolios. This poses a challenge in preserving the confidentiality of these qualities when other agents compute their Q-values.

To address this issue, an attention-based Q-value estimation method is proposed. As depicted in Figure 4-3, the Q function's input variables for agent *i* include only its own state s_i^t and action a_i^t, along with the exogenous contributions x_i from other agents. This method allows for a more efficient and privacy-preserving Q-value estimation.

$$Q_{\theta_i}(s_i^t, a_i^t) = f_i(e_i^t, x_i^t)$$
$$e_i^t = g_i(s_i^t, a_i^t) \tag{4.7}$$

where $f_i(\cdot)$ represents a two-layer multilayer perceptron (MLP), and $g_i(\cdot)$ is a one-layer MLP. The term x_i allows agent *i* to incorporate information from other agents in its decision-making process. Although the initial input for x_i is derived from other agents, a privacy-focused approach involves filtering an implicit feature embedding. Additionally, the Q function's input dimension is significantly decreased to $|e_i^t + x_i^t|$, offering scalability through adjusting the neuron count in the MLP's output layer.

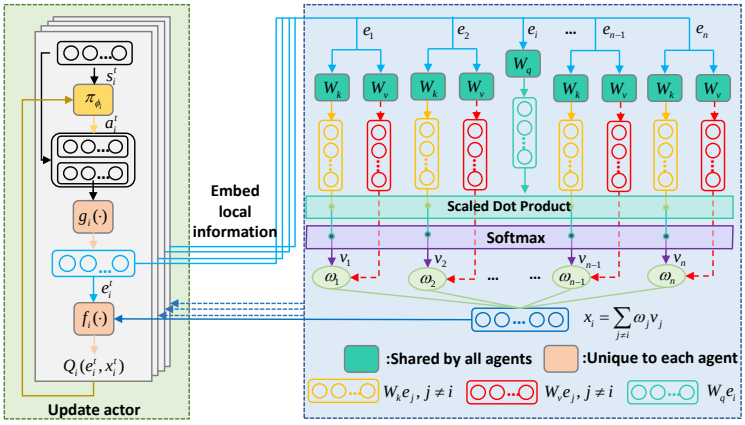


Figure 4-3 The structure of the attention-based Q-value estimation [J4].

4.3.3.2 Training and Execution of MAATD3

Figure 4-4 displays flowchart of the MAATD3 approach. Details about algorithm updating can be found in [J4]. During the training phase, each agent's actor network executes actions independently without sharing information, relying only on local

states as inputs. Conversely, the critic network, employing a centralized approach for Q-value estimation, incorporates the implicit embeddings e_n of all agents, facilitated by an integrated attention mechanism.

During the execution phase, the well-trained actor networks are used to provide decisions for the MG. Throughout each episode, over time intervals $t \in [0, T]$, agent *i* autonomously implements its learned policy. This is based on its specific state s_i^t and through its own actor network $\pi_{\phi_n}^*$, embodying a decentralized approach to decision-making without any information exchange with other agents. The detailed training and execution stages are presented in Algorithm 1 and 2, respectively.

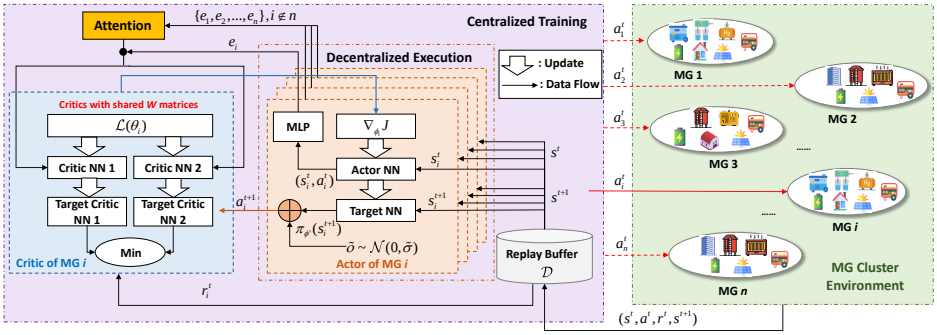


Figure 4-4 The framework of MAATD3 [J4].

Algorithm 1 Training stage of MAATD3

1.Initialize weights of networks and bottom-layer MG cluster environment
2.**For** *each epsiode* **do**
3. Receive initial states $s^0 = \{s_1^0, ..., s_n^0\}$ for all agents
4. **For** each time step **do**
5. Output action $a^t \sim \pi_{\phi_{1:n}}(\cdot \mid s^t)$ for each agent's actor network $\pi_{\phi_{1:n}}$
6. Take actions $a^t = \{a_0^t, ..., a_n^t\}$, return r^t, and the environment moves to the next state s^{t+1}
7. Store experience (s^t, a^t, r^t, s^{t+1}) into the replay buffer
8. **If** experience is stored up **then**
9. Sample a batch experience $\mathcal{B} = \{(s^t, a^t, r^t, s^{t+1})\}$ from replay buffer
10. Update the critic's and actor's weights based on backpropagation
11. Update the target critic's weights via soft updating
12. **End if**
13. **End for**
until convergence

Algorithm 2 Execution stage of MAATD3

1.Obtain weights of the well-trained actors $\pi^*_{\phi_{1:n}}$
2.**For** each episode **do**
3. Obtains initial states $s^0 = \{s_1^0, ..., s_n^0\}$ for all agent n
4. **For** each time step **do**
5. Output actions $a^t = \{a_0^t, ..., a_n^t\}$ for all agents
6. Execute actions $a^t = \{a_0^t, ..., a_n^t\}$ in the environment, return r^t, and the environment moves to the next state s^{t+1}
7. **End for**
8.**End for**

4.3.4. UPPER-LAYER DISPATCH METHOD

Every MG operates autonomously at each time step once it has received the well-trained strategy. Every MG independently calculates the quantity of electricity it trades with its local ER and transmits this information to the upper layer. The upper layer then uses a convex optimization technique to solve the optimal power dispatch problem by combining all of the power exchange data from the MGs.

4.4. NUMERICAL VERIFICATION

In this section, the effectiveness of the proposed bottom-up energy management strategy is validated based on a specific EI network.

4.4.1. SIMULATION SETUP

In the simulation, the bottom-layer MG cluster comprised eight MGs ($n=8$). The configuration of these MGs was as follows: {MG_1, MG_2, MG_3} were categorized as residential MGs, {MG_4, MG_5, MG_6} as commercial MGs, and the remaining MGs were designated as industrial MGs. The primary goal was to minimize total energy costs over a specified time period $[0, 24h]$. The control interval was set at 15 minutes, and each episode consists of 96 time slots ($t = \{1, 2, ..., 96\}$). The training dataset includes renewable generation and load demands at 15-minute intervals. Residential MGs relied on data from residential households [96], commercial MGs used data from a commercial warehouse [97], and industrial MGs used data from a power plant [98]. The historical dataset was partitioned into a training set and a test set, with a ratio of 80% for the training set and 20% for the test set.

4.4.2. OPERATION OF INDIVIDUAL MG

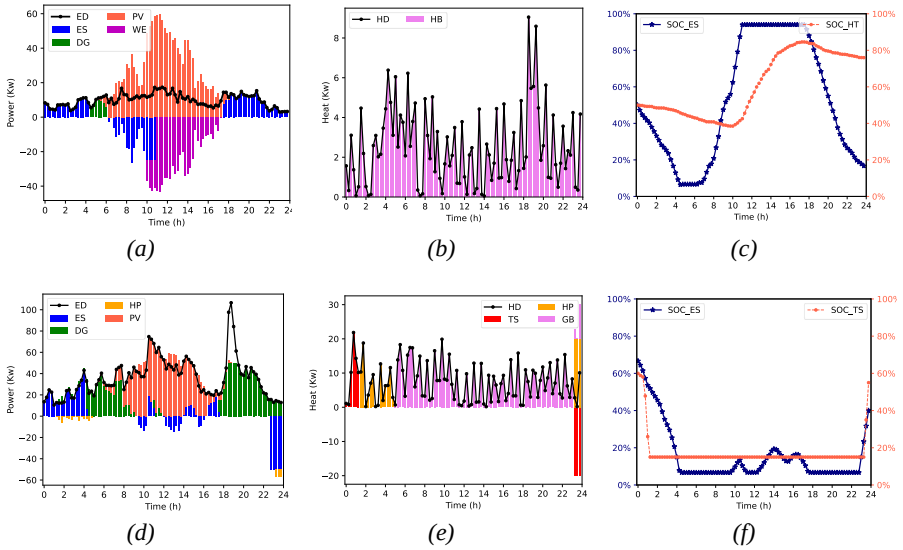
The operation of MG1, MG5, and MG7 were chosen as typical residential, commercial, and industrial MGs, respectively. Figure 4-5 depicts the fluctuations in

the SoC and the satisfaction of electricity and heat requirements for each MG in 15-minute intervals.

The first column of Figure 4-5 shows the electricity demand supply of the MG. It can be seen that to reduce generation costs, ES is prioritized for supplying power. When the SoC of the ES reaches its minimum, the DG and CHP units begin to operate. During periods of high PV generation, the ES stores excess PV energy. In MG_1, the WE unit converts surplus PV into hydrogen, which is stored in the HT. In Figure 4-5(g), due to the limited capacity of the ES, it cannot fully balance the power, forcing MG7 to exchange excess PV.

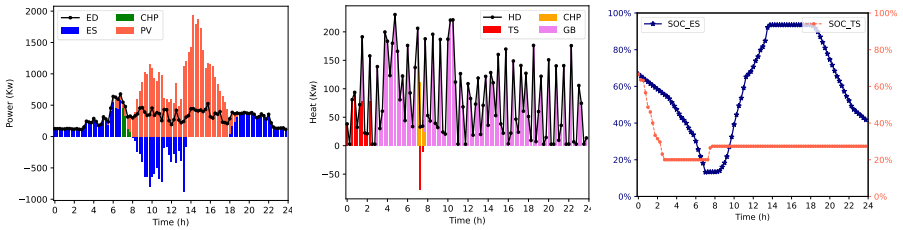
The second column displays the supply of heat load. In MG_1, the heat load is supplied by the hydrogen stored in the HT, without purchasing natural gas from the external gas grid. Given the high generation costs of the FC, it remains offline. In MG_5 and MG_7, natural gas is purchased and used by the GB to provide heating.

The third column illustrates the SOC of the ES, HT, and TS. The energy storage actively participates in MG operation. When PV generation is low, the ES supplies power, and during high PV generation periods, it stores the excess electricity.



(a) *(b)* *(c)*

(d) *(e)* *(f)*



(g) *(h)* *(i)*

Figure 4-5 Figures (a)-(c) presents electricity load upply, heat load supply, and the SOC changes of the ES and HT in the residential MG. Figures (d)-(f) details the commercial MG's operation results. Figures (g)-(i) presents the industrial MG's operation results. ED refers to electricity demand and HD signifies heat demand [J4].

Additionally, Table 4-1 displays the comparative results of three methods with respect to operating cost and computation time. The operation costs of the proposed strategy are 41% less than those of the optimal power flow (OPF) -based strategy and 9.6% less than those of the TD3-based strategy. Furthermore, the computation time of the model-free DRL technique is significantly less than that of the model-based OPF approach because only the forward propagation of the neural network is needed for online testing. The objective of the OPF is to minimize the total generation costs of the MG cluster. The performance of the OPF is often constrained by the need to solve a series of complex, non-convex equations. TD3 lacks mechanisms to efficiently handle multi-agent coordination, which can limit its performance.

Table 4-1 Operation costs and testing time of different methods [J4].

Method	**Operation costs**	**Online testing time (s)**
Proposed	$(1.324 \pm 0.089) \times 10^4$	0.0002
TD3	$(1.451 \pm 0.121) \times 10^4$	0.0002
OPF	$(1.921 \pm 0.328) \times 10^4$	2.21

4.4.3. SIGNIFICANCE OF THE ATTENTION MECHANISM

An assessment is conducted on the impact of the attention mechanism in the proposed strategy. Figure 4-6 compares the training progress of the attention mechanism-utilizing algorithm (MAATD3) with the algorithm that does not use it (MATD3). In this comparison, MAATD3, represented by the red line, demonstrates quicker convergence and attains higher episode rewards compared to the MATD3 algorithm, indicated by the blue line. The training findings demonstrate that the attention mechanism significantly enhances training efficiency and overall learning quality. It achieves this by selectively valuable information from all agents.

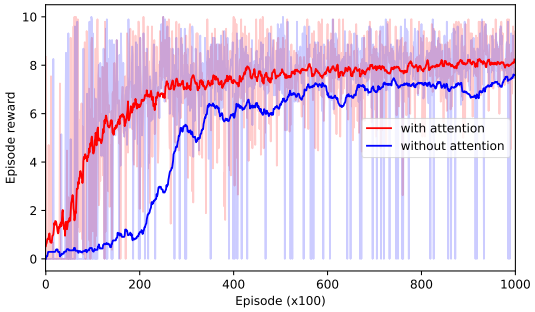


Figure 4-6 The comparison shows how well the suggested strategy trains both with and without the attention mechanism. Shaded areas in the graph indicate the range of immediate rewards, which show notable variability, while the dark lines represent the average rewards over sets of 10 episodes, providing a clearer visualization of the trend [J4].

4.4.4. POWER DISPATCHING ANALYSIS IN THE UPPER LAYER

Figure 4-7 presents an example of power distribution in the upper layer. It demonstrates how the ER network, specifically through ER_4 and ER_8, engages in energy trading with the utility grid. The optimal power dispatching strategy between the ERs and the main grid is determined by the upper controller. In this context, to simulate stochastic electricity prices, the model employs Geometric Brownian Motion, a method often used in modeling stock price processes.

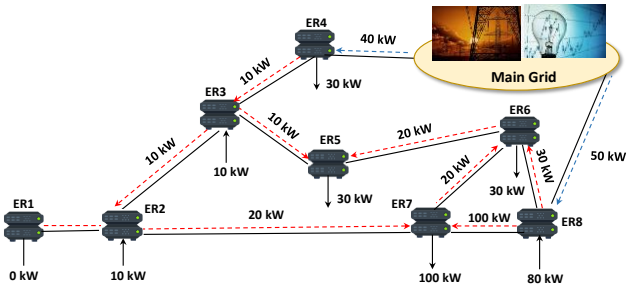
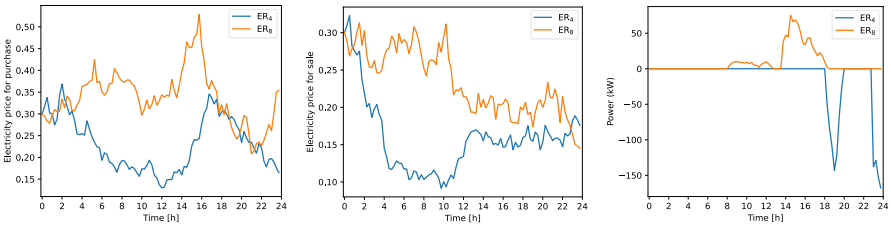


Figure 4-7 The upper layer energy dispatch diagram for the EI scenario, in which every ER_i is connected to an equivalent MG_i. Power distribution and electrical transactions between ERs are represented by blue and red dotted arrows, while power exchange with the MG cluster is shown by black solid arrows [J4].

The power flows between ERs and the main grid are depicted in Figure 4-8, together with the costs associated with purchasing and selling electricity. Notably, power purchases between 08:00 and 18:00 predominantly occur through ER_8, where the selling price is higher compared to ER_4. Conversely, the ER network compensates for power deficits by buying electricity through ER_4, which offers a lower purchase price.



(a) *(b)* *(c)*

Figure 4-8 Geometric Brownian Motion is used to determine the electricity pricing for energy purchase and sale, which are displayed in parts (a) and (b). Part (c) shows how power flow interacts with the main grid [J4].

4.5. CONCLUSION

In this chapter, a bottom-up energy management strategy for the EI network based on MADRL is proposed. The EI network is composed of the bottom-layer MG cluster and upper-layer ER cluster. A model-free MAATD3 algorithm is applied to achieve the optimal energy management strategy for the bottom-layer multiple MES. Each MES only requires local measurements to make the optimized decisions, which preserve its privacy. Besides, the attention mechanism is used to speed up training by selectively utilizing valuable information from other agents. The optimal dispatch in the upper layer is determined through convex optimization. Simulation results validate the effectiveness of the proposed energy management strategy.

CHAPTER 5. MADRL-BASED DECENTRALIZED ENERGY MANAGEMENT STRATEGY FOR THE MES AND EVAGG ENTITIES

The contents of Chapter 5 are based on the following two papers:

J5: B. Zhang, W. Hu, X. Xu, T. Li, Z. Zhang and Z. Chen, "Physical-Model-Free Intelligent Energy Management for a Grid-Connected Hybrid Wind-Microturbine-PV-EV Energy System via Deep Reinforcement Learning Approach", *Renewable Energy*, vol. 200, pp. 433-448, 2022.

J6: B. Zhang, W. Hu, D. Cao, A. Ghias, and Z. Chen, "Novel Data-Driven Decentralized Coordination Model for Electric Vehicle Aggregator and Energy Hub Entities in Multi-Energy System Using an Improved Multi-Agent DRL Approach," *Applied Energy*, vol. 339, no. 120902, Jun. 2023.

C3: B. Zhang, Z. Chen, X. Wu, D. Cao, and W. Hu. "A MATD3 -based Voltage Control Strategy for Distribution Networks Considering Active and Reactive Power Adjustment Costs", *2022 IEEE International Conference on Power Systems and Electrical Technology (PSET 2022)*.

5.1. INTRODUCTION

With the integration of a large number of DERs into MES, the energy interactions between various entities and MES, such as EVAGG, cannot be overlooked. These entities belong to different stakeholders, creating a competitive environment. This chapter focuses on developing decentralized strategies to maximize the profits of EVAGG entities and minimize the energy costs of EHs in the MES. The research is divided into two parts: The first part optimizes energy management strategies in an MES by considering the stochastic behavior of EVs, including charging, departure, and arrival times. The second part formulates decentralized strategies for EVAGGs and EHs within an MES, ensuring privacy protection so that strategies can be made in real-time using only local measurements.

5.2. DRL-BASED ENERGY MANAGEMENT STRATEGY FOR THE MG INCLUDING EVS

5.2.1. INTRODUCTION

Investigating an effective energy management strategy for the renewable-based MG presents significant challenges due to the multiple uncertainties. Additionally, the rising integration of EVs complicates the situation, rendering traditional model-based approaches less effective. This research proposes a model-free DRL-based optimal energy management strategy to minimize operation costs while meeting charging requirements. The flowchart of this research is presented in Figure 5-1. Section 5.1.2 gives the mathematical model of the MG with EVs. Section 5.1.3 presents the framework of the proposed strategy. Case study is conducted in Section 5.1.4 to validate its effectiveness. Conclusion is given in Section 5.1.5.

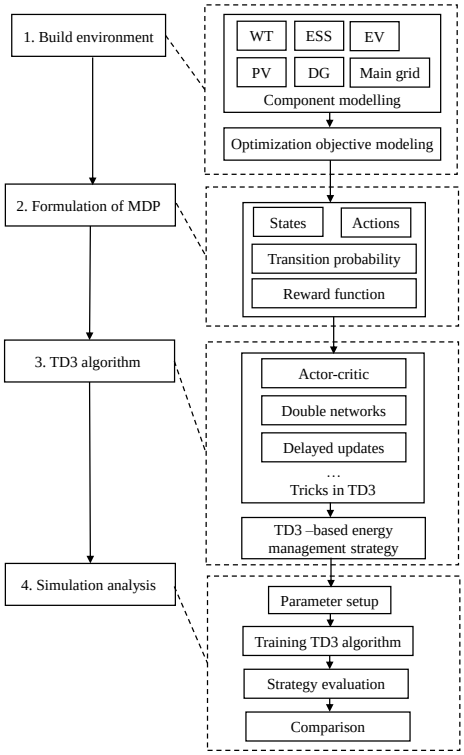


Figure 5-1 Flowchart of this research [J5].

5.2.2. SYSTEM DESCRIPTION

1) Objective function

The grid-connected MG is depicted in Figure 5-2 and comprises DGs, EVs, PV panels, load demands, wind turbine (WT) generators, and a battery energy storage system (ESS).

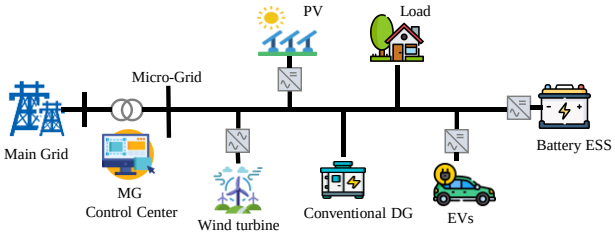


Figure 5-2 Structure of the MG system including EVs [J5].

The scheduling objectives are to minimize DG generation costs, RE curtailment, and transaction costs associated with selling or buying electricity from the main grid.

2) Constraints

Many Constraints need to be met in the scheduling model, such as the maximum power output of each DG, the charging/discharging power and SoC of the battery ESS, the active power exchanged by the main grid, and the charging power and SoC of EVs.

The mathematical models of the objective function and constraints are listed in [J5].

5.2.3. DRL-BASED ENERGY MANAGEMENT STRATEGY

5.2.3.1 MDP Formulation

The MDP formulation is presented as follows:

1) State: The state at time t consists of $s_t = (P_t^{WT}, P_t^{PV}, P_t^{DG}, E_t^{ES}, P_t^{G}, \varepsilon_t, S_t^{EV})^T$: active power output of WT, PV and DG, SoC of ESS, power exchanged with the grid, electricity price and SoC value of EVs, respectively.

2) Action: The action at time t consists of $a_t = (\Delta P_t^{DG}, P_t^{ESS}, P_t^{EV})$: the adjustment amount of DGs' power generation according to the previous time step, the ESS's charging and discharging power, and the EVs' charging and discharging power.

3) Reward function: The reward r_t at time t is regarded as the negative form of objective function, which is defined as follows:

$$r_t = -(C_t^{DG} + C_t^{EV} + C_t^{G}) + C_t^{RE} \tag{5.1}$$

where C_t^{DG} is generation costs of DGs, C_t^{EV} is EV charging costs, C_t^{G} is the transaction cost associated with buying or selling power from the main grid, and C_t^{RE} represents incentive benefits of RE consumption.

4) Transition probability: The SoC values of battery ESS and EVs can be determined by the previous SoC values and charging/discharging power. However, considering the uncertainties of WT and PV generations, the corresponding state transition probability cannot be available.

5) System problem: The DRL task is to find the optimal control policy π^* to maximize the expected total reward across a specific time horizon T: $\mathbf{P1}: \max_{\pi} R_t = \mathbb{E}\left[\sum_{\tau=0}^{T} \gamma^{\tau} r_{t+\tau+1}\right]$.

5.2.3.2 TD3-based Energy Management Strategy

Figure 5-3 depicts a detailed flowchart of the TD3-based energy management strategy [99]. The details of TD3 algorithm can be found in [J5]. The relationship between the TD3 algorithm and the MG system, which serves as the environment, is depicted in this flowchart.

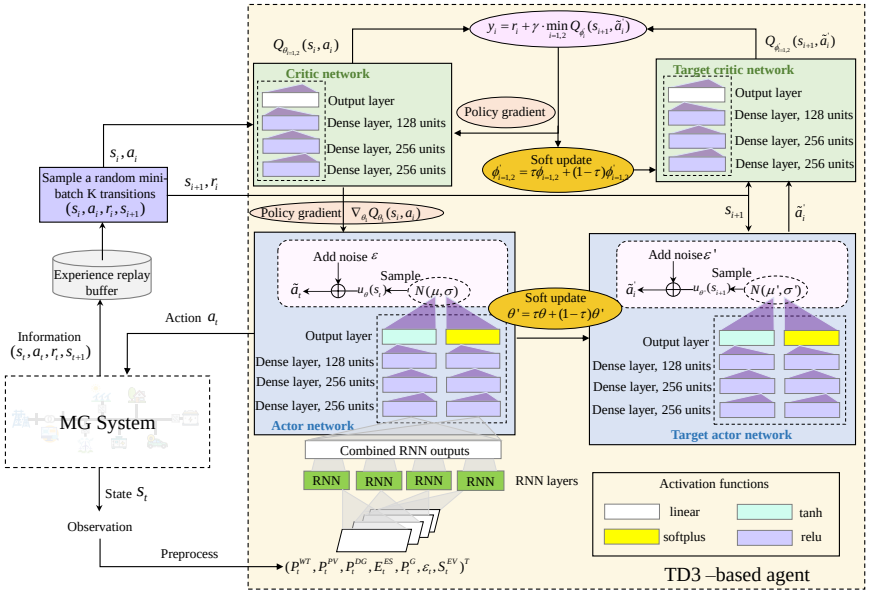


Figure 5-3 Framework of TD3-based scheduling strategy [J5].

Before modifying the DNNs' weights, historical interaction data is stored in the experience replay buffer. The buffer selects a random mini-batch for training after it has accumulated a particular quantity of data. N-element state features are created and preprocessed for every observation time slot. Recurrent neural networks (RNNs) are initially supplied with the preprocessed state feature information. The online actor network then uses the RNNs' output as input to calculate the current action value. An autonomous energy management strategy is constructed for the MG by fine-tuning the network weights using the TD3 algorithm. This network offers a real-time strategy to get optimal operational performance based on the observed data. Details about algorithm updating can be found in Section 3 in [J5].

5.2.4. CASE STUDY

5.2.4.1 Simulation Setup

The application of a benchmark grid-connected MG system to assess the efficacy of the proposed energy management strategy based on the TD3 algorithm is shown in Figure 5-4. Historical annual data sources, such as wind power, solar irradiance, electricity load, and electricity prices, are chosen as training sets [100], [101]. In addition, the behavior of EVs was modeled using a normal distribution, including arrival time, departure time, and initial SoC of EVs.

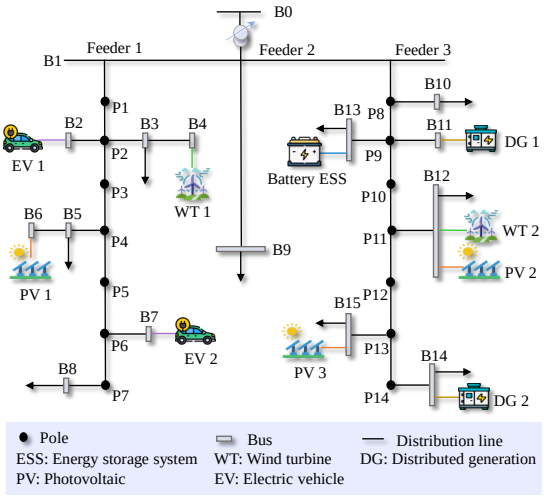


Figure 5-4 Configuration of the benchmark MG network [J5].

5.2.4.2 Training Performance

The cumulative reward changes for each episode during the training process for the TD3 and DDPG algorithms are compared in Figure 5-5. The TD3 algorithm routinely produced better cumulative rewards than the DDPG. In contrast, DDPG exhibited unstable learning behavior and failed to converge effectively. This divergence is attributed to TD3's implementation of key techniques such as delayed policy updates and target policy smoothing, which enhance training stability and efficiency.

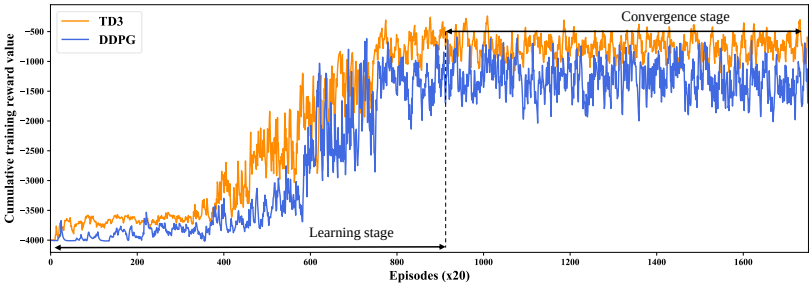


Figure 5-5 Comparison of cumulative reward values of TD3 and DDPG algorithms [J5].

There are three stages to the training process: convergence, training, and exploration. Without changing the DNN parameters, the TD3 agent collects a large amount of interaction data from the experience replay buffer during the exploration phase. The training phase starts when the replay buffer is filled, and the agent keeps modifying the DNN weights to learn an optimized strategy that maximizes cumulative rewards.

Reward values stabilize throughout the convergence phase, signifying that the DRL agent learns the optimal strategy.

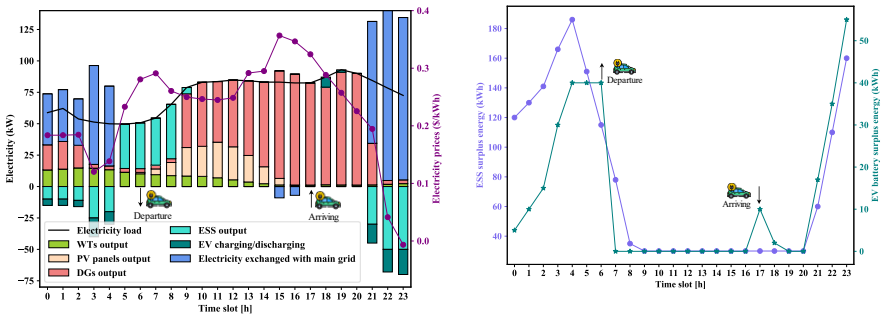
5.2.4.3 Test Results

In order to evaluate the efficacy of the suggested TD3-based scheduling approach, a one-day performance simulation is conducted for every time slot. Furthermore, the simulation test was also conducted based on three consecutive days.

1) Operation results on a test day

RESs, such as PV panels and WTs were initially utilized to meet electricity demands. During the early hours (00:00–04:00), the strategy leaned towards purchasing surplus electricity from the grid to charge the EV and ESS, particularly at 03:00, when wholesale electricity prices were lower. Between 05:00 and 09:00, the ESS was primarily used to supply power until its energy levels neared the minimum threshold.

For the peak hours of 10:00–20:00, with higher electricity prices, the strategy shifted to using DGs for the remaining electricity loads. Increasing DG output was more cost-effective than purchasing excess electricity from the grid. It's interesting to note that the EV released excess electricity upon arrival at 17:00, negating the need to raise DG output. Due to the decreased cost of power, the agent was able to fully charge the EVs and ESS in the last hours of 21:00–23:00, which reduced operating expenses.



(a) *(b)*

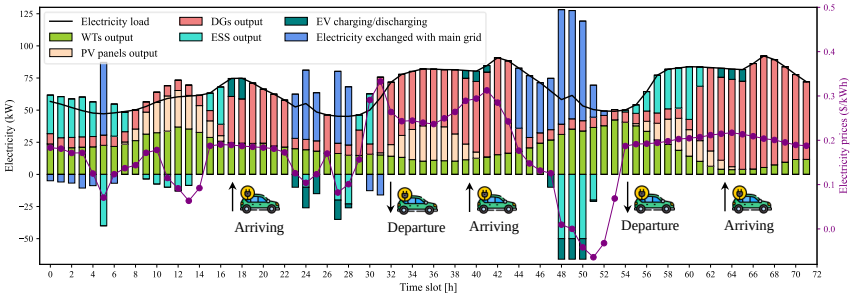
Figure 5-6 (a) Output of units for each time slot; (b) Energy variations in ESS and remaining energy from EVs [J5].

2) Operation results on three consecutive days

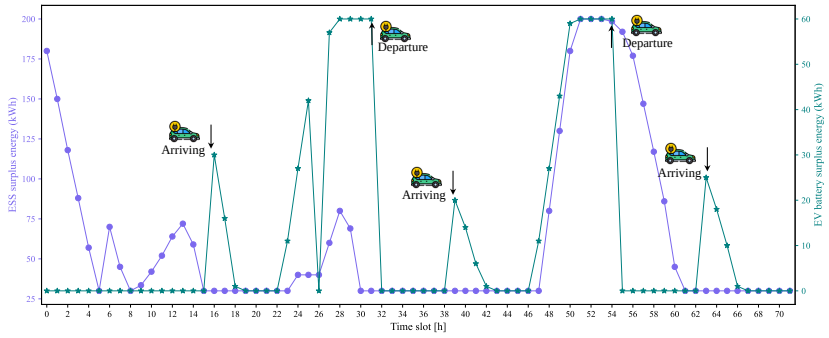
To illustrate the robustness of the TD3-based energy management strategy in real-time optimization, simulation results from three consecutive days are analyzed, as shown in Figure 5-7. Figure 5-7(a) displays the operation of controlled units along with fluctuating wholesale electricity prices. Notably, the strategy minimizes costs by buying extra electricity from the main grid for charging EVs and ESS during times

when electricity prices are low (that is, the 5th, 23–25th, and 47–53rd time slots). On the other hand, during periods of high electricity prices (such the 17–18, 39–41, and 63–65 time slots), the strategy chooses to avoid increasing DG output by using the stored energy in EVs for load support. This decision is contingent upon meeting EV owners' expected energy needs upon departure. Consequently, despite high prices during the 29–32nd time slots, the system preferred buying surplus power over depleting EV batteries. Moreover, during the significantly expensive 63–71st time slots, generating electricity via DGs was favored over grid purchases to curtail costs.

Figure 5-7 (b) presents that the ESS's energy is regulated within predefined operational bounds. It also ensures that the expected battery levels are maintained when EV owners depart, particularly during the 31st and 54th time slots.



(a)



(b)

Figure 5-7 (a) Power output of the controlled units; (b) fluctuations in the ESS and remaining energy from EVs [J5].

3) Comparison results with other benchmark methods

The DRL-based energy management strategy for the MG system was evaluated in comparison to the DDPG and particle swarm optimization (PSO)-based optimization. The average daily operating costs throughout a 30-day test period served as the

primary comparative criterion. The parameters of the TD3 and DDPG algorithms are the same.

The daily expenses spent by each approach during the 30-day test period are displayed in Figure 5-8. Quantitative data are shown in Table 5-1, together with typical expenses and computation durations. TD3 was a more economical option than the other techniques. More specifically, TD3 and DDPG decreased total expenses by 15.27% and 11.52%, respectively, in comparison to the PSO-based stochastic method. The performance of DDPG is sensitive to hyper-parameters setting, leading to unstable training. Because the PSO-based stochastic technique relies on iterative calculations to optimize 200 samples, it took the longest to complete. Even while TD3 required more training time than DDPG due to its more complex neural architecture, it was still within reasonable bounds.

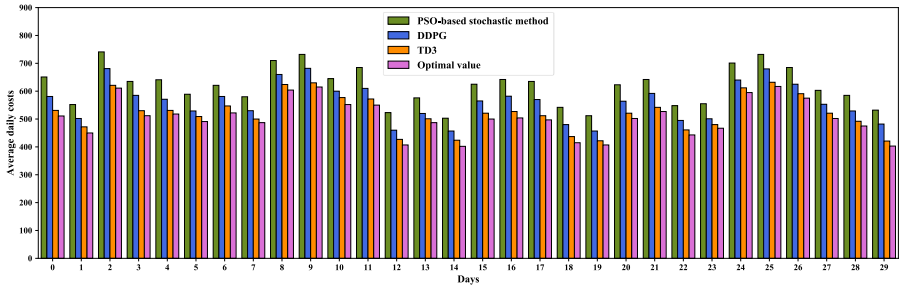


Figure 5-8 Comparison of average daily costs using the TD3, DDPG and PSO [J5].

Table 5-1 Comparison of different examined approaches [J5].

Method	**Average cost ($)**	**Improvement**	**Time consumed (s)**
PSO-based stochastic method	628.65	-	2036.3
DDPG	556.23	11.52%	61.2
TD3	532.63	15.27%	72.6

5.2.5. CONCLUSION

A model-free DRL-based energy management strategy is investigated in order to reduce EV charging expenses and optimize operating profitability. Considering the uncertainties associated with RES, the variability of electricity rates, and the changing charging habits of EVs, the well-trained TD3 agent, by utilizing DNNs, proficiently provides the continuous control of the MG system's components without necessitating prior system modeling knowledge. The effectiveness of the proposed energy management strategy is validated by the simulation results

5.3. MADRL-BASED DECENTRALIZED ENERGY MANAGEMENT STRATEGY FOR MESS AND EVAGG

5.3.1. INTRODUCTION

The purpose of this research is to investigate decentralized energy management strategies for the EVAGG and EH entities in an integrated electricity and district heating system (IEDHS). It encounters several challenges: the ownership diversity of EHs and EVAGGs fosters a competitive environment within the IEDHS. Second, uncertainties like RESs' intermittent nature, electricity prices, and EV users' driving behaviors exist. Lastly, because of the nonlinearity in the thermal and power flow models in the IEDHS, the operational objective offers a multi-objective, nonlinear function that further complicates the problem. Therefore, a data-driven MADRL-based decentralized energy management strategy is studied. In Section 5.2.2, the system model is presented. The proposed decentralized method is given in Section 5.2.3. Case study is conducted in Section 5.2.4. Conclusion is given in Section 5.2.5.

5.3.2. MODELLING OF THE MES AND EVAGG ENTITIES

A comprehensive schematic of the model is shown in Figure 5-9. The EVAGG, which can buy and sell electricity on the wholesale market and to an EH while making sure that EV users' charging needs are satisfied, is shown on the left side of the diagram. The IEDHS is shown with its five main components (the EH entity, the district heating network (DHN), and the power distribution network (PDN)). The electricity subnetwork is responsible for fulfilling electrical demands, while the heating subnetwork meets thermal requirements. The system integrates a CHP unit and a GB, which serve as coupling components linking the electricity and heat networks.

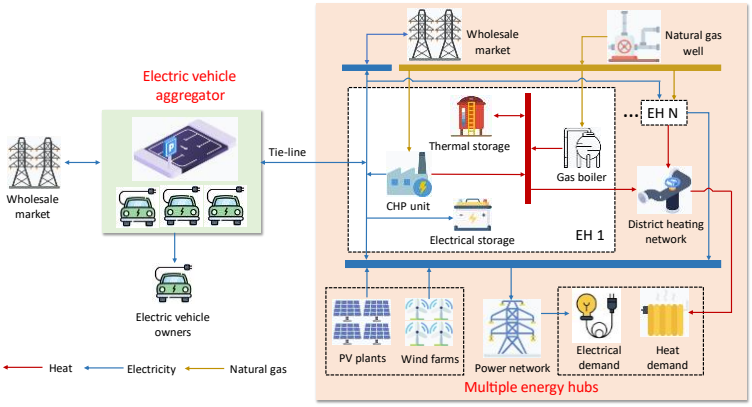


Figure 5-9 Architecture of the studied system including EVAGG and EH entities [J6].

5.3.2.1 PDN Description

In this study, DC power flow equations are used to establish the PDN, including the DC power flow balance, the constraints of generators, and the constraints of the exchanged power. For details can be found in Section 2 in [J6].

5.3.2.2 DHN Description

As depicted in Figure 5-10, the DHN is structured as a dual-layer system, consisting of both supply and return networks [102]. Within the DHN, there are three distinct types of nodes: Firstly, the source nodes, which are responsible for delivering thermal power. Secondly, the load nodes, which utilize this thermal power. And thirdly, intermediate nodes, which serve as conduits for transferring thermal power to neighbor nodes. The process starts with the water flow in the supply network distributing thermal power to each end consumer. The water then recirculates over the return network following the exchange of thermal power at the load nodes. Because of its dual nature, the DHN usually takes into account both thermal and hydraulic models, which represent the interaction between heat transfer and water movement in the system. The DHN consists of hydraulic and thermal models, which are detailed in Section 2 in [J6].

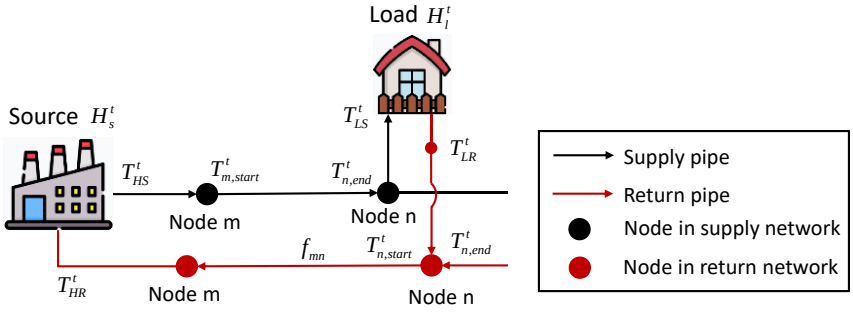


Figure 5-10 Thermal flow model in a simplified DHN [J6].

5.3.2.3 EVAGG Description

The EVAGG model aims to optimize its energy expenses, which are expressed as follows. These include selling power to EV owners at a fixed prices ς_{s-EVO}, interacting with the wholesale market at a locational marginal price ς_u^t, and transacting with the EH entity at contracted hourly price ς_{con}^t [103].

$$F_{EVA}^t = \sum_{m=1}^{M} N_{EV}\mu_m \left\{ \varsigma_{con}^t (P_{EV_m}^{t,EH-EVA} - P_{EV_m}^{t,EVA-EH}) - \varsigma_{s-EVO} P_{EV_m}^{t,s-EVO} + \varsigma_u^t (P_{EV_m}^{t,b-g} - P_{EV_m}^{t,s-g}) \right\} \quad (5.2$$

where μ_m denotes the percent of EV m's type, and N_{EV} is the number of aggregated EVs. $P_{EV_m}^{t,EH-EVA}$ / $P_{EV_m}^{t,EVA-EH}$ indicates that the EVAGG purchases/sells electricity from/to the EH entity. $P_{EV_m}^{t,s-EVO}$ represents that the EVAGG sell electricity to EV owners. $P_{EV_m}^{t,b-g}$ / $P_{EV_m}^{t,s-g}$ indicates that the EVAGG purchases/sells electricity from/to the grid.

5.3.2.4 Energy hub model

The EH model consists of WT, PV, CHP, ESS, and a boiler. The EH model trades electricity with the wholesale market, collaborates with the EVAGG, and ensures the supply of heat and electricity supply. The EH's objective function F_{EH}^t at time t is expressed as below:

$$F_{EH}^t = \sum_{n=1}^{N} \varsigma_{gas}^t Q_{EH_n}^t + \varsigma_u^t (P_{EH_n}^{t,b-g} - P_{EH_n}^{t,s-g}) \tag{5.3}$$

where N is the number of EH models, $Q_{EH_n}^t$ is the gas entering EH_n at time t, $\{P_{EH_n}^{t,b-g}, P_{EH_n}^{t,s-g}\}$ are electrical energy purchased and sold from and to the wholesale market by EH_n at time t, and $\{\varsigma_{gas}^t, \varsigma_u^t\}$ are nature gas price and electricity price at time t.

5.3.3. IMPROVED MADRL ALGORITHM

5.3.3.1 Markov Game formulation

A Markov game is used to formulate the coordination energy management problem between the EVAGG and EH entities, which is expressed as follows:

1) Agents: The EVAGG and each EH are regarded as agents. As a result, a competitive dynamic is shown between the EHs and the EVAGG, although cooperation between the EHs also occurs in this multiagent environment.

2) States: The system states $\mathcal{S}^t = \{s_1^t, s_2^t, ..., s_N^t\}$ contain state information of all agents. The states of EVAGG agent only include its local observation, such as arrival time, departure time, SoC and electricity price. The states of EH agent are power outputs of WT and PV, load demands, electricity price, gas price, and SoC of the ESS.

3) Actions: The EVAGG agent's actions include selling electricity to EV owners, selling/purchasing electricity to/from the market, and selling/purchasing electricity to/from the EHs. The EH agent is responsible for purchasing energy from the wholesale market and operating the boiler, ESS, and CHP.

4) Reward function: The reward value of the EVAGG agent is defined as $r_{EVA}^{t} = -CST_{EVA}^{t}$, and the reward function of the EH agent is $r_{EH}^{t} = -CST_{EH}^{t}$.

5) State transition probability: The transition probability of the storage system can be determined, but the transition probability of uncertainties, such as RESs, load demands and energy prices, is not available.

6) System problem: The objective of agents is to discover the optimal energy management strategy in order to maximize the expected cumulative reward within a certain time period T: $\mathbf{P1}: \max_{\pi} R^{t} = \mathbb{E}\left[\sum_{\tau=0}^{T} \gamma^{\tau} r^{t+\tau+1}\right]$.

5.3.3.2 Long Short-Term Memory Network

A long short-term memory (LSTM) network is applied to predict the uncertainties of RESs, load demands and energy price [104]. Figure 5-11 illustrates the internal configuration of a LSTM unit, which includes a memory cell and three distinct gates: input, forget, and output. The cell state $\boldsymbol{c}^{t}$, which includes a self-connected recurrent edge with a constant weight of one, is instrumental in mitigating issues of vanishing and exploding gradients. Additionally, the roles of the three gates within the LSTM are pivotal. The forget gate $\boldsymbol{f}^{t}$ and input gate i^{t} regulate the flow of data into the cell state $\boldsymbol{c}^{t}$, while the output gate manages the data $\boldsymbol{o}_{t}$ flowing into the next layer $\boldsymbol{h}^{t}$. The input gate processes current input data $\boldsymbol{x}^{t}$ and the previous time step's hidden state $\boldsymbol{h}^{t-1}$ using the tanh function. Since the tanh function yields values between 0 and 1, it allows the input gate to ascertain the influence of the current input on the cell state. Moreover, the forget gate determines the extent to which the previous cell state is preserved in the current cell state.

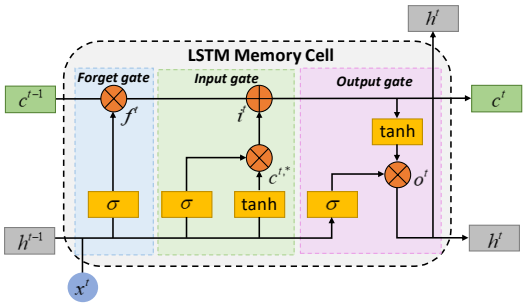


Figure 5-11 The structure of a LSTM neural network [J6].

5.3.3.3 Safe SAC algorithm

SAC algorithm has been discussed in Section 2.3.2. The actions selected and carried out by several agents are independent and multi-dimensional in the setting of MADRL

algorithms. This necessitates the separation of interdependencies within the optimization-based approach. For example, constraints related to balancing thermal and electrical demand and supply, as well as those concerning ESSs, might not be adhered to during the action execution. To accurately represent the physical constraints of the EH operation, a safety index $c^t(s^t, a^t)$ is employed, which is defined below:

$$c^t(s^t, a^t) = -|\Delta H^t| - |\Delta P^t| - |\Delta E_e^t| - |\Delta E_{th}^t| \tag{5.4}$$

where $\{\Delta H^t, \Delta P^t\}$ are imbalance measures of thermal and electrical energy at time t, and $\{\Delta E_{th}^t, \Delta E_e^t\}$ are SoC constraint violations of thermal and electrical storage units at time t. This indicator helps adjust the direction in which the control policy is updated, contributing to a more stable learning process.

5.3.3.4 Implementation of the proposed method

The design of the DNNs and the procedures involved in executing the proposed strategy are summed up in Figure 5-12. The DNNs' structure is shown in Figure 5-12(a). It is observed that the target networks and their corresponding online networks have the same parameters. PV, wind power, and electrical and heat demands are among the datasets given into the LSTM network. The variables used here are normalized. After that, they go via an LSTM layer and produce a vector. A flattening layer converts this vector into a longer feature vector. A sigmoid activation function is used to generate a normalized action value from the actor network, which receives as input a combination of the LSTM outputs and additional state characteristics as SoC and energy prices. The actor network's output layer consists of four neurons for instant action output, while the critic and safety networks, taking a concatenation of state and action vectors as input, output the Q-value and C-value respectively through a single neuron, and the rectified linear units (ReLU) is the activation function [105], [106]. Further details on the algorithm updates can be found in Section 3 in [J6].

The proposed strategy's execution structure, which consists of decentralized online execution and centralized offline training, is shown in Figure 5-12(b). By incorporating information from other agents into the critic network of each agent during offline training, the strategy becomes more resilient to external uncertainties even when only local information is available. The critic network is rendered redundant during online execution, and the actor networks adjust weights to produce real-time strategy. Based on its learned policy, each agent's distinct actor network uses its observed data to make decisions in real time in a totally decentralized manner.

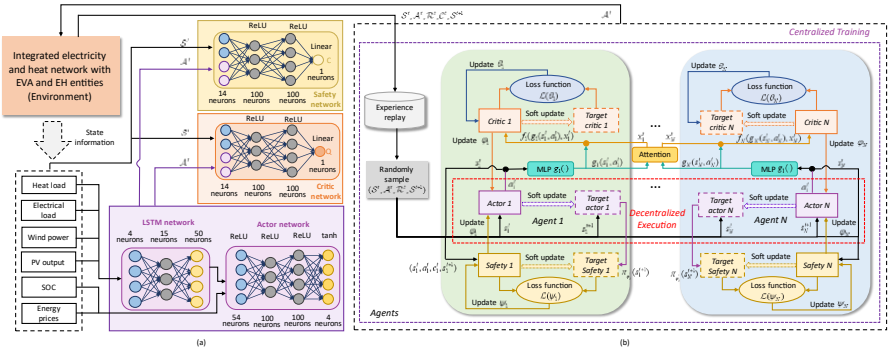


Figure 5-12 Flowchart and structures of the proposed method [J6].

5.3.4. CASE STUDIES

5.3.4.1 Simulation setup

The proposed energy management strategy is applied to an IEDHS. In this context, one episode is comprised of 24 time steps, with each step representing one hour. Figure 5-13 presents the topology of the test IEDHS [107]. The IEDHS includes an IEEE 33-bus PDN, a 4-node DHN, four EH and EVAGG entities. Parameters of the EH are given in [J6]. The Gaussian distribution $N(0.45, 0.01)$ is used to model the SoC of EVs upon arrival at the parking lot. Arrival time and departure time are sampled from a uniform distribution from the sets of $\{6,7,8,9,10,11\}$ and $\{15,16,17,18,19,20\}$, respectively.

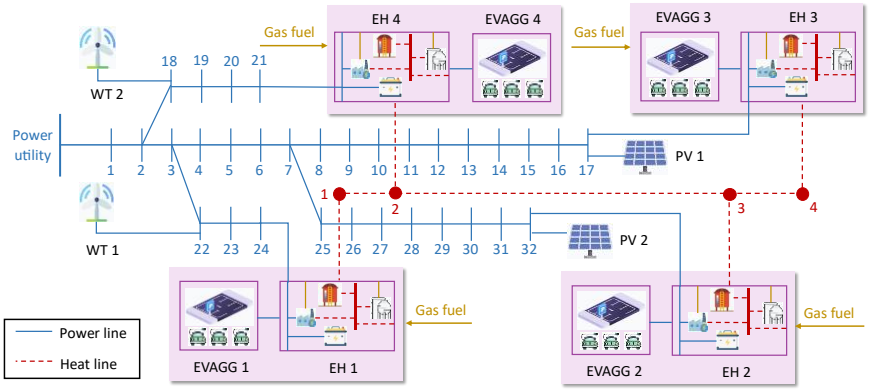


Figure 5-13 Topology of the test system [J6].

One-year historical data is selected as the training data, including PV, wind power, load demand, and energy prices [108][109]. To assess the forecasting accuracy, an unaltered dataset spanning a continuous 30-day period from the same references is employed for testing. Figure 5-14 displays a comparison of the forecasted and actual values. The comparison reveals a close resemblance between the predicted values (represented by an orange line) and the actual values (shown by a blue line), with only minor deviations in a few instances of very high peaks. This similarity validates the effectiveness and accuracy of the LSTM network in making predictions.

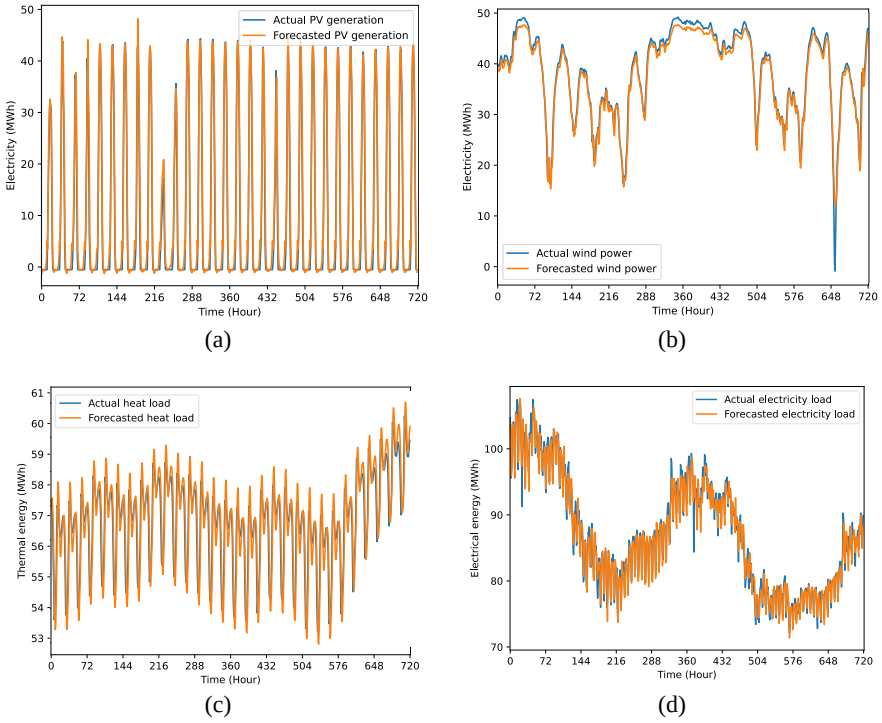


Figure 5-14 Values for predictions derived from a 30-day test dataset: (a) PV generation, (b) wind power, (c) heat load, and (d) electrical load [J6].

5.3.4.2 Training performance

The Concurrent, MASAC and the proposed algorithm are trained for 10,000 episodes. Figure 5-15 and Figure 5-16 show the outcomes of this training, which concentrated on reward convergence and constraint violation. The solid curves in these figures indicate the mean values of the results; the shaded regions correspond to the standard deviations. It is noted that the Concurrent method's training performance is unstable, displaying a significant standard deviation that results in its non-convergent termination. The non-stationary environment that results from agents updating their rules individually in the Concurrent approach causes this instability. Conversely, as

compared to the Concurrent algorithm, the MASAC algorithm, which incorporates an attention mechanism, exhibits smoother learning behavior and a smaller standard deviation. This suggests that the non-stationarity difficulties can be efficiently addressed by centralized training that selectively includes input from other agents.

Out of the three strategies, the proposed algorithm had the lowest standard deviation and the highest cumulative reward. The integration of the safety network and LSTM network is credited with this higher performance, as it greatly improves the quality of the solution.

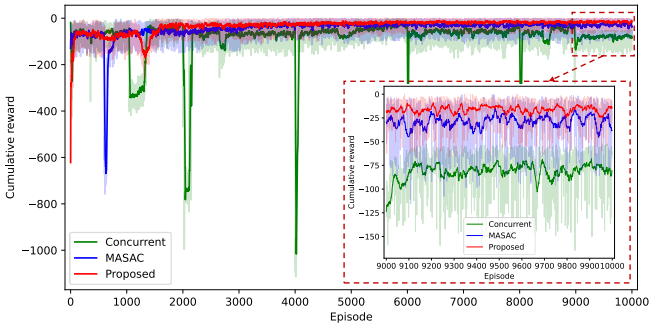


Figure 5-15 Comparison of cumulative rewards of different MADRL methods [J6].

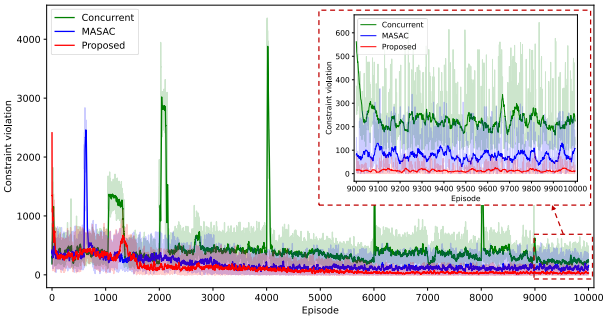


Figure 5-16 Comparison of constraint violations of different MADRL methods [J6].

5.3.4.3 Test results Analysis

Two distinct test scenarios (a summer day and a winter day) are examined. The summer day scenario is marked by lower load demands, high availability of PV power, and limited wind power. In contrast, the winter day features higher load demands, limited PV power, and an abundance of wind power. Figure 5-17(a) and Figure 5-17(b) in the study illustrate the 24-hour demand profiles for heat and electricity for these scenarios, while Figure 5-16(c) displays the electricity and gas price trends for the EH

and EVAGG. Additionally, the specifics of four different EVAGGs, including variables like arrival time, departure time, and initial SoC, are detailed in Table 3.

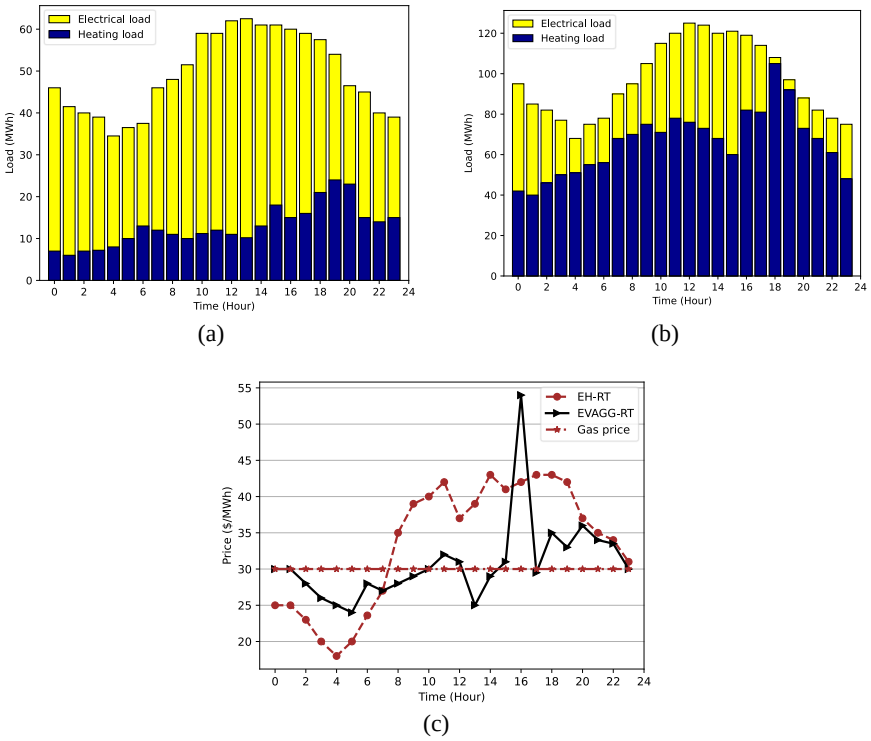


Figure 5-17 (a) The load profiles on the summer day, (b) the load profiles on a winter day, and (c) the trends in electricity prices and gas prices [J6].

Table 5-2 Parameter settings of the EVs [J6].

Parameters	**1**	**2**	**3**	**4**
E_{EV}^{ini}	0.58	0.57	0.22	0.2
AT	5	7	11	8
DT	19	15	18	15

1) EVAGG operation

Figure 5-18 provides the operational activities of four EVAGGs during a typical winter day. The power transactions between the EVAGG and the wholesale market are shown in Figure 5-17(a). The EVAGG starts obtaining power from the upper grid one hour in advance of the EVs pulling into the parking lot. Because of the higher electricity pricing at that hour, the EVAGG only sells electricity back to the grid once,

at hour 16. EVs are typically charged in the hours prior to their departure, as shown in Figure 5-17(b). In order to meet the EH's electricity demands, the EVAGG sells electricity to the latter between the hours of 07:00 and 14:00, as shown in Figure 5-17(c). Because the contracted costs for power from the EH are less than the wholesale market rates, the EVAGG also exhibits a bias for purchasing electricity from the EH. The charging and discharging behaviors of the combined batteries are displayed in Figure 5-17(d). There are three primary components to the discharging process: providing electricity to the EH from 7:00 to 14:00 hours, charging EVs when they depart the parking lot between 15:00 and 19:00 hours. Electrical energy from the EH and the wholesale market is used to charge these batteries.

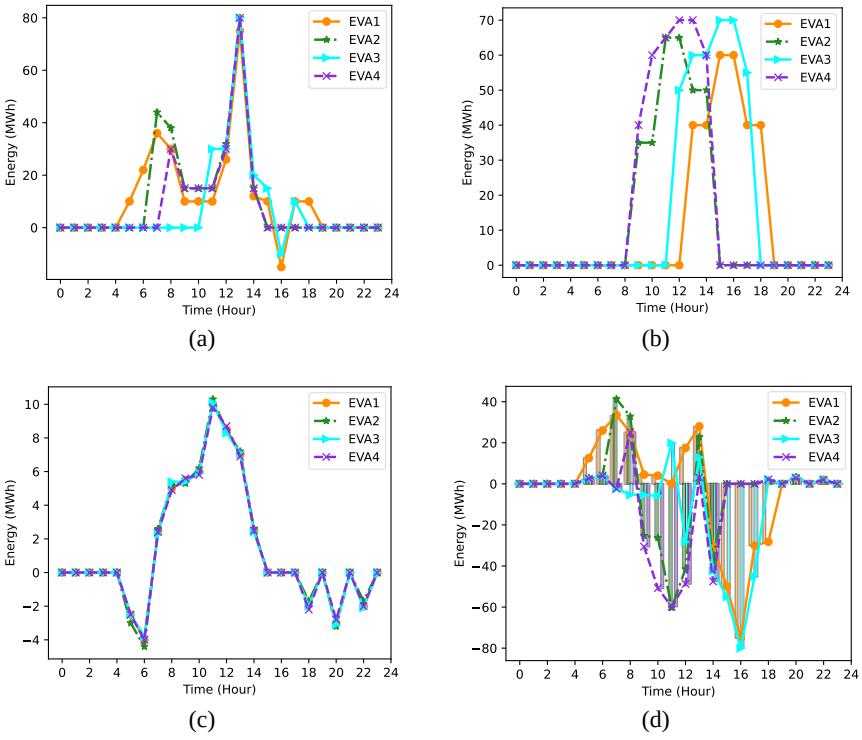
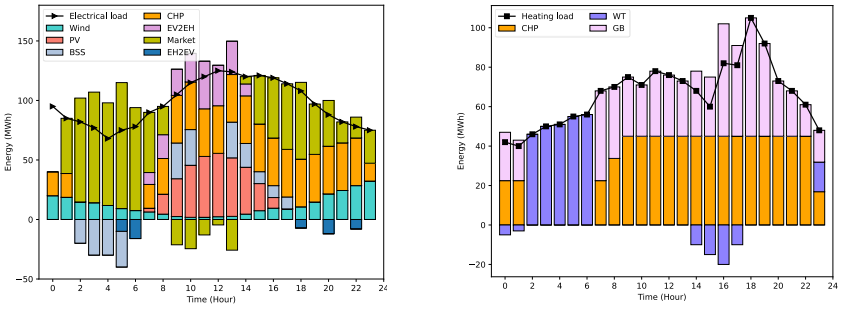


Figure 5-18 Electrical is traded with the electrical market in (a), sold to EV owners in (b), exchanged with EH in (c), and the EVAGGs' charging and discharging operations are handled in (d) [J6].

2) EH operation

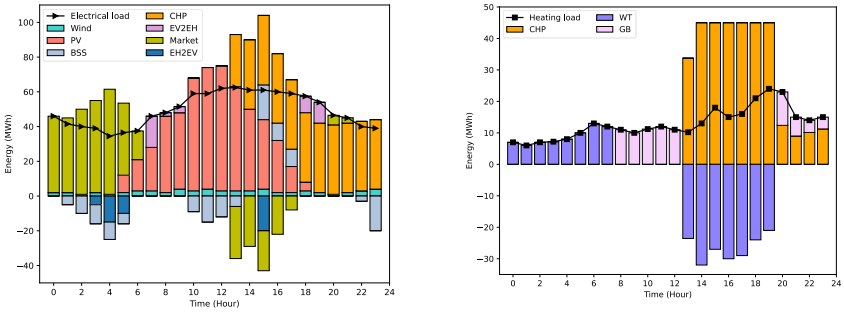
Figure 5-19 demonstrates the supply of heat and electricity demands on a winter day, highlighting the utilization of wind power and PV generation. During off-peak electricity hours (00:00 to 06:00 h), electricity is bought from the wholesale market for various purposes including charging batteries and selling to the EVAGG. During

peak hours (08:00 to 20:00 h), the EH buys electricity from the EVAGG, sells excess to the market, and relies on the CHP and batteries for load supply. Between 14:00 and 23:00 h, the EH purchases electricity from the market due to CHP's maximum output. Natural gas, being cheaper than electricity, predominantly fuels the GB and CHP for heating, with WT being used when gas and electricity prices are close (2-6 h and hour 23). Figure 5-20 outlines the energy management on a summer day. Similar to the winter day, electricity is purchased early in the day for supporting electric loads and charging the BSS. When gas is less expensive than electricity, the strategy uses RESs to satisfy demand, especially when it comes to utilizing GB for heating (09:00–16:00 h). When there is minimal RES availability, a strong demand for electricity, and relatively low gas prices, the CHP units are used. During times of high demand, ESSs have a flexible role. Notably, compared to a winter day, a summer day with more PV generation exhibits greater energy export and less import. Overall, EH's real-time energy requirements are efficiently managed by the learnt strategy, which adjusts to varying seasonal conditions. To meet demand, the strategy makes use of RES, generators, and energy storage devices, proving its efficacy and versatility.



(a) (b)

Figure 5-19 EH operation on a winter day: (a) the strategy employed to satisfy electricity demand; (b) the strategy used to meet heat demand [J6].

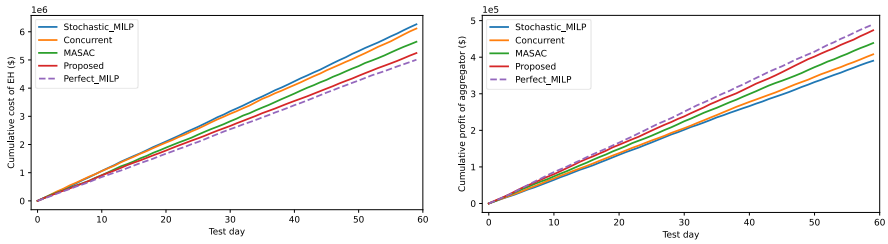


(a) (b)

Figure 5-20 EH operation on a summer day: (a) the strategy employed to satisfy electricity demand; (b) the strategy used to meet heat demand [J6].

5.3.4.4 Algorithm Comparison

The proposed strategy is compared with two baseline model-based strategies, namely stochastic-mixed-integer linear programming (MILP) [110] and perfect-MILP [111], as well as two advanced MADRL algorithm, the Concurrent [112] and MASAC algorithms, to illustrate the enhanced performance considering the LSTM and safety networks. Figure 5-21 shows the total daily cost of EHs and total daily profit of EVAGGs over a test dataset. The average cost, profit, and computation performance of different algorithms are presented in Table 5-3. It can be seen that the optimization results of the proposed method are very close to those of the perfect-MILP algorithm, with the EH cost being 2.01% higher and the EVAGG profit being 2.12% lower than those of perfect-MILP. However, solving the perfect-MILP algorithm requires precise modeling of the system. Additionally, compared to MASAC, the proposed method achieves a 3.06% reduction in EH cost and a 6.82% increase in EVAGG profit. The proposed algorithm, based on the MASAC algorithm, incorporates LSTM and a Safety network, thereby improving performance. The Concurrent method can cause environmental non-stationarity, leading to poor training performance. In terms of computation time, the proposed method takes longer than MASAC due to its more complex neural network structure. However, for online deployment, since only forward propagation of the neural network is required, millisecond-level decision-making can be achieved. Details about the comparison algorithms can be found in Section 4 in [J6].



(a) (b)

Figure 5-21 Results obtained by different algorithms over a test dataset: (a) total daily cost of EH; (b) cumulative daily profit of EVAGG [J6].

Table 5-3 Cost, profit and computation time of different methods [J6].

Algorithms	Average daily cost of EH ($)	Average profit of EVAGG ($)	Training time (min)	Online calculation time (s)
Concurrent	103969	7281	503	0.013
MASAC	98662	8012	114	0.024
Proposed	95733	8598	125	0.029
Stochastic-MILP	105035	7265	-	3.685
Perfect-MILP	93846	8784	-	1.379

5.3.5. SUMMARY

This chapter introduces a novel decentralized energy management strategy based on an improved model-free MADRL. This strategy aims to minimize daily operation costs for EH entities and maximize daily profit for EVAGGs. The uncertainties are predicted via a LSTM network. The coordination between two entities is then modeled as Markov games, tackled using an MADRL algorithm. In this framework, each EH or EVAGG entity is chosen as an agent, respectively. The MADRL strategy incorporates offline centralized training for learning optimal coordinated strategy and decentralized execution, allowing agents to make real-time decisions based on local measurement. Additionally, a safety network is utilized to consider equality constraints, such as balancing demand and supply. The rationality and robustness of the proposed strategy are evaluated in the simulation.

CHAPTER 6. CONCLUSION

6.1. SUMMARY

This thesis investigates data-driven energy management strategies for the MESs to optimize the economic and low-carbon operation considering the uncertainties of RES, loads, energy prices and EVs' charging/discharging behaviors. The purpose of this thesis is to understand complex MES from a data-driven point of view, without assuming too much prior knowledge.

In Chapter 2, a low-carbon economic energy management strategy for the electricity-gas MES based on DRL is investigated. The coordination between P2G and CCS units is considered. The low-carbon economic dispatch problem is formulated as MDP, and solved by an improved SAC algorithm. Simulations demonstrate that the proposed strategy achieves faster convergence and a more stable training process compared to traditional DRL algorithms.

In Chapter 3, a two-timescale energy management strategy based on the MADRL algorithm is investigated to minimize energy costs of the residential MES. The strategy considers internal energy conversion and external energy trading for the residential MES, taking into account the various operational parameters of each MES component. Simulations in deterministic and stochastic scenarios demonstrate the effectiveness and superiority of the proposed strategy.

In Chapter 4, an MADRL-based bottom-up energy management strategy is investigated for multiple MESs, which is composed of the upper-layer ER cluster and bottom-layer MG cluster. An MADRL algorithm learns the optimal operation strategy for the bottom-layer MG cluster to minimize energy costs. The optimal energy allocation is completed in the upper-layer ER cluster. Simulation validates the effectiveness of the proposed energy management strategy.

In Chapter 5, a decentralized energy management strategy for EH and EVAGG entities is investigated to reduce the energy costs of the EH and increase the profit of the EVAGG entity. A LSTM network is used to predict the system uncertainties, and a safety network is used to ensure the operating constraints. Simulation demonstrates the effectiveness and superiority of the proposed strategy. Besides, decentralized execution can safeguard the privacy of various entities.

In summary, this thesis primarily demonstrates the use of DRL to learn an optimal energy management strategy, aiming to optimize the economic costs of MESs. The end-to-end nature of DRL effectively addresses the uncertainties and nonlinearities in MES optimization. In terms of modeling, the focus shifts from centralized energy management of a single MES to multi-agent decentralized energy management,

involving multiple MESs or the MES with EVAGG entities. The DRL algorithms are improved to solve these problems. Additionally, the real-time decision-making capability of DRL highlights its potential for practical application.

6.2. FUTURE WORK

This Ph.D. project aims to propose DRL-based algorithms to solve energy management problems at different levels. However, a number of limitations still exist:

- Hyperparameter determination is a necessary process for DRL, and researchers often need to spend a lot of effort on parameter tuning to get the optimal model performance. Running experiments manually for parameter tuning can be used for small models with small parameter sizes, but when parameter optimization is performed for large models, the manual-only approach becomes impractical. In the future fast parameter tuning algorithms can be used with the help of DL models.

- DRL methods rely on big data, and not all domains have the ability to obtain a large amount of sample data, and the cost of obtaining a large number of training samples is still high in modern power grids due to the presence of physical barriers in the energy layer and information barriers in the information layer. In the future, when the training samples are insufficient, the opposite idea is utilized to generate pseudo-labels with unlabeled data or pseudo-data with labels, forming a sample generating network.

- The MES in the operation process may appear the extreme situation, and it is difficult to ensure the feasibility of the strategy given by the agent. The model knowledge is embedded in the neural network of DRL to construct a data-knowledge fusion-driven DRL algorithm, which improves the robust performance of the DRL control strategy through the embedding of knowledge.

- The proposed strategy has not been tested in real-time on an actual MES but is instead based on historical data. This is mainly due to the large scale of the investigated energy system model, which makes real-world testing challenging. Therefore, in future work, a physics-informed MADRL algorithm could be further developed, incorporating the system's physical constraints into the neural network to enhance the interpretability of the resulting strategy.

References

[1] X. Tian, C. An, "The role of clean energy in achieving decarbonization of electricity generation, transportation, and heating sectors by 2050: A meta-analysis review," *Renewable and Sustainable Energy Reviews*, vol. 182, no. 113404, Aug. 2023.

[2] Adams, S., Klobodu, E., Apio, A, "Renewable and non-renewable energy, regime type and economic growth," *Renewable Energy*, vol. 125, pp. 755-767, Sep. 2018. https://doi.org/10.1016/j.renene.2018.02.135.

[3] M. Aneke and M. Wang, "Energy storage technologies and real life applications – a state of the art review," Applied Energy, vol. 179, pp. 350-377, 2016.

[4] O. Yolcan, "World energy outlook and state of renewable energy: 10-year evaluation," *Innovation and Green Development*, vol. 2(4), no. 100070, Dec. 2023.

[5] GWEC, Wind Power & Green Recover Hub. 2020, https://gwec.net/green-recovery-data-analysis/.

[6] Preliminary Energy Statistics for Denmark for the year 2024, Danish Energy Agency, 2024. [online]. Available: https://ens.dk/en/our-services/projections-and-models/global-report.

[7] F. Calise, F. Cappiello, L. Cimmino et al., "Dynamic simulation and thermoeconomic analysis of a power to gas system," *Renewable and Sustainable Energy Reviews*, vol. 187, no. 113759, Nov. 2023.

[8] J. Rifkin, "The Third Industrial Revolution; How Lateral Power is Transforming Energy, the Economy, and the World", 2011.

[9] Q. Hassan, P. Viktor, T. Musawi et al., "The renewable energy role in the global energy transformations," *Renewable Energy Focus*, vol. 48, no. 100545, Mar. 2024.

[10] Y. Liu, C. Feng, "Promoting renewable energy through national energy legislation," *Energy Economics*, vol. 118, no. 106504, Feb. 2023.

[11] Z. Wang and T. Hong, "Reinforcement learning for building controls: The opportunities and challenges," *Applied Energy*, vol. 269, no. 115036, Jul. 2020.

[12] Z. Zhu, Z. Hu, K. Chan et al., "Reinforcement learning in deregulated energy market: a comprehensive review," *Applied Energy*, vol. 329, no. 120212, Jan. 2023.

[13] D. Qiu, Y. Wang, W. Hua et al., "Reinforcement learning for electric vehicle applications in power systems: a critical review," *Renewable and Sustainable Energy Reviews*, vol. 173, no. 113052, Mar. 2023.

[14] D. Cao, J. Zhao, and W. Hu et al., "Model-free voltage control of active distribution system with PVs using surrogate model-based deep reinforcement learning," *Applied Energy*, vol. 306, no. 117982, Jan. 2022.

[15] B. Wang, Y. Li, and W. Ming et al., "Deep reinforcement learning method for demand response management of interruptible load," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, Jul. 2020.

[16] F. Gorostiza and F. Gonzalez-Longatt, "Deep reinforcement learning-based controller for SOC management of multi-electrical energy storage system," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, Nov. 2020.

[17] H. Ma, D. Dong, and S. Ding et al., "Curriculum-based deep reinforcement learning for quantum control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 11, Nov. 3023.
[18] A. Anwar and A. Raychowdhury, "Autonomous navigation via deep reinforcement learning for resource constraint edge nodes using transfer learning," *IEEE Access*, vol. 8, Feb. 2020.
[19] Z. Zhu, K. Lin, and A. Jain et al., "Transfer learning in deep reinforcement learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 11, Nov. 2023.
[20] Y. Zhao, J. Liu, and X. Liu et al., "Enhancing the tolerance of voltage regulation to cyber contingencies via graph-based deep reinforcement learning," *IEEE Transactions on Power System*, vol. 39, no. 2, Mar. 2024.
[21] S. Bahrami, Y. Chen, and V. Wong, "Deep reinforcement learning for demand response in distribution networks," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, Mar. 2021.
[22] J. Li, D. Dong, and Z. Wei et al., "Quantum reinforcement learning during human decision-making," *Nature Human Behavior*, vol. 4, pp. 294-307, 2020.
[23] Z. Yan and Y. Xu, "Real-time optimal power flow with linguistic stipulations: integrating GPT-agent and deep reinforcement learning," *IEEE Transactions on Power system*, vol. 39, no. 2, Mar. 2024.
[24] B. Huang and J. Wang, "Applications of physics-informed neural networks in power system – A review," *IEEE Transactions on Power Systems*, vol. 38, no. 1, Jan. 2023.
[25] T. Nguyen, N. Nguyen, and S. Nahavandi, "Deep reinforcement learning for multiagent systems: a review of challenges, solutions, and applications," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, Sep. 2020.
[26] Y. Jin, S. Wei, and J. Yuan et al., "Hierarchical and stable multiagent reinforcement learning for cooperative navigation control," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 1, Jan. 2023.
[27] R. Yan, Q. Xing, and Y. Xu et al., "Multi-agent safe graph reinforcement learning for PV inverters-based real-time decentralized vol/var control in zoned distribution networks," *IEEE Transactions on Smart Grid*, vol. 15, no. 1, Jan. 2024.
[28] H. Ge, D. Gao, and L. Sun et al., "Multi-agent transfer reinforcement learning with multi-view encoder for adaptive traffic signal control," *IEEE Transactions on Intelligent Transportation System*, vol. 23, no. 8, Aug. 2022.
[29] X. Liu, X. Li, J. Tian et al., "Low-carbon economic dispatch of integrated electricity-gas energy system considering carbon capture, utilization and storage," *IEEE Access*, pp. 25077-25089, Mar. 2023.
[30] S. He, H. Gao, Z. Chen, J. Liu, L. Zhao, G. Wu and S. Xu, "Low-carbon distribution system planning considering flexible support of zero-carbon energy station," *Energy*, vol. 244, part B, no. 123079, Apr. 2022.
[31] X. Guo, Y. Liao, G. Li et al., "Low-carbon economic dispatch of Photovoltaic-Carbon capture power plant considering deep peak regulation," *Journal of Cleaner Production*, vol. 420, no. 138418, Sep. 2023.

[32] K. Jia, C. Liu, S. Li et al., "Modeling and optimization of a hybrid renewable energy system integrated with gas turbine and energy storage," *Energy Conversion and Management*, vol. 279, no. 116763, Mar. 2023.
[33] W. Fan, L. Ju, Z. Tan et al., "Two-stage distributionally robust optimization model of integrated energy system group considering energy sharing and carbon transfer," *Applied Energy*, vol. 331, no. 120426, Feb. 2023.
[34] Y. Dong, H. Zhang, P. Ma et al., "A hybrid robust-interval optimization approach for integrated energy systems planning under uncertainties," *Energy*, vol. 274, no. 127267, Jul. 2023.
[35] L. Kang, J. Wang, X. Yuan et al., "Research on energy management of integrated energy system coupled with organic Rankine cycle and power to gas," *Energy Conversion and Management*, vol. 287, no. 117117, Jul. 2023.
[36] J. Liu, L. Ma, Q. Wang, "Energy management method of integrated energy system based on collaborative optimization of distributed flexible resources," *Energy*, vol. 264, no. 125981, Feb. 2023.
[37] D. Lei, Z. Zhang, Z. Wang et al., "Long-term, multi-stage low-carbon planning model of electricity-gas-heat integrated energy system considering ladder-type carbon trading mechanism and CCS," *Energy*, vol. 280, no. 128113, Oct. 2023.
[38] X. Wu, B. Liao, Y. Su et al., "Multi-objective and multi-algorithm operation optimization of integrated energy system considering ground source energy and solar energy," *International Journal of Electrical Power & Energy Systems*, vol. 144, no. 108529, Jan. 2023.
[39] S. Li, W. Hu, and D. Cao et al., "Electric vehicle charging management based on deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 3, May 2022.
[40] T. Yang, L. Zhao, and W. Li et al., "Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning," *Energy*, vol. 235, no. 121377, Nov. 2021.
[41] U.S. Energy Information Administration. Share of total U.S. energy consumption by end-use sectors, 2020. https://www.eia.gov/energyexplained/use-of-energy/. Accessed: Sept. 15th, 2021.
[42] S. Zheng, Y. Sun, B. Li, B. Qi, X. Zhang, and F. Li, "Incentive-based integrated demand response for multiple energy carriers under complex uncertainties and double coupling effects," *Applied Energy*, vol. 283, no. 116254, Feb. 2021.
[43] M. Ahrarinouri, M. Rastegar, and A. R. Seifi, "Multiagent reinforcement learning for energy management in residential buildings," *IEEE Transactions on Industrial informatics*, vol. 17(1), pp. 659-666, Jan. 2021.
[44] S. Korjani, F. Casu, A. Damiano, V. Pilloni, and A. Serpi, "An online energy tool for sizing integrated PV-BESS systems for residential prosumers," *Applied Energy*, vol. 313, no. 118765, May 2022.
[45] Y. Deng, F. Luo, Y. Zhang, and Y. Mu, "An efficient energy management framework for residential communities based on demand pattern clustering" *Applied Energy*, vol. 347, no. 121408, Oct. 2023.
[46] M. Salehi and M. Rastegar, "Distributed peer-to-peer transactive residential energy management with cloud energy storage," *Journal of Energy Storage*, vol. 58, no. 106401, Feb. 2023.

[47] R. Nematirad, M. Ardehali, A. Khorsandi et al., "Optimization of residential demand response program cost with consideration for occupants thermal comfort and privacy," *IEEE Access*, vol. 12, pp. 15194-15207, Jan. 2024.
[48] A. Sridhar, S. Honkapuro, F. Ruiz et al., "Toward residential flexibility—Consumer willingness to enroll household loads in demand response," *Applied Energy*, vol. 342, no. 121204, Jul. 2023.
[49] M. J. Sanjari, H. Karami, and H. B. Gooi, "Analytical rule-based approach to online optimal control of smart residential energy system," *IEEE Transactions on Industrial Informatics*, vol. 13(4), pp. 1586-1597, Aug. 2017.
[50] G. Glenk and S. Reichelstein, "Economics of converting renewable power to hydrogen," Nature Energy, vol. 4(3), pp. 216-222, 2019.
[51] A. Mohammadi and M. Mehrpooya, "Techno-economic analysis of hydrogen production by solid oxide electrolyzer coupled with dish collector," *Energy Conversion and Management*, vol. 173, pp. 167-178, Oct. 2018.
[52] M. Felgenhauer and T. Hamacher, "State-of-the-art of commercial electrolyzers and on-site hydrogen generation for logistic vehicles in South Caralina," *International Journal of Hydrogen Energy*, vol. 40(5), pp. 2084-2090, Feb. 2015.
[53] Y. Wang, D. Qiu, and G. Strbac, "Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems," *Applied Energy*, vol. 310, no. 118575, Mar. 2022.
[54] G. K. H. Larsen, N. D. van Foreest, and J. M. A. Scherpen, "Distributed control of the power supply-demand balance," *IEEE Transactions on Smart Grid*, vol. 4(2), pp. 828-836, Jun. 2013.
[55] G. Wang, Y. Zhou, Z. Lin et al., "Robust energy management through aggregation of flexible resources in multi-home micro energy hub," *Applied Energy*, vol. 357, no. 122471, Mar. 2024.
[56] L. Gao, S. Deng, and H. Li et al., "An event-triggered approach for gradient tracking in consensus-based distributed optimization," *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 2, Mar. 2022.
[57] Z. Wang, F. Xiao, and Y. Ran et al., "Scalable energy management approach of residential hybrid energy system using multi-agent deep reinforcement learning," *Applied Energy*, vol. 367, no. 123414, Aug. 2024.
[58] R. Jain, J. Qin, R. Rajagopal, "Data-driven planning of distributed energy resources amidst socio-technical complexities," *Nature Energy*, vol. 2, no. 17112, 2017.
[59] M. Azimian, V. Amir, S. Mohseni, A. C. Brent, N. Bazmohammad, J. M. Guerrero, "Optimal investment planning of bankable multi-carrier MG networks," *Applied Energy*, vol. 328, no. 120121, 2022.
[60] H. Hua, Y. Qin, C. Hao, J. Cao, "Stochastic optimal control for energy Internet: A bottom-up energy management approach," *IEEE Transactions on Industrial Informatics*, vol. 15(3), pp. 1788-1797, 2019.
[61] H. Hua, Y. Qin, C. Hao et al., "Optimal energy management strategies for energy Internet via deep reinforcement learning approach," *Applied Energy*, vol. 239, pp. 598-609, 2019.

[62] J. Leithon, S. Werner, V. Koivunen, "Cost-aware renewable energy management: centralized vs. distributed generation," *Renewable Energy*, vol. 147(1), pp. 1164-1179, 2020.
[63] S. Henni, M. Schaffer, P. Fischer et al., "Bottom-up system modeling of battery storage requirements for integrated renewable energy systems," *Applied Energy*, vol. 333, no. 120531, Mar. 2023.
[64] D. Cao, J. Zhao, W. Hu, F. Ding, Q. Huang, Z. Chen et al., "Data-driven multiagent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs," *IEEE Transactions on Smart Grid*, vol. 12(5), pp. 4137-4150, 2021.
[65] M. Prina, G. Manzolini, D. Moser, B. Nastasi, W. Sparber, "Classification and challenges of bottom-up energy system models – A review," *Renewable and Sustainable Energy Reviews*, vol. 129, no. 109917, 2020.
[66] N. Pereira and M. Ramon, "Energy usage and human behavior modeling for residential bottom-up energy simulation," *Energy and Buildings*, vol. 279, no. 112653, Jan. 2023.
[67] M. Rastegar, M. Fotuhi-Firuzabad, M. Moeini-Aghtai, "Developing a two-level framework for residential energy management," *IEEE Transactions on Smart Grid*, vol. 9(3), pp. 1707-1717, 2018.
[68] X. Yang, Y. Zhang, H. Wu, H. He, "An event-driven ADR approach for residential energy resources in MGs with uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 66(7), pp. 5275–5288, 2019.
[69] X. Lin, L. Wang, H. Xu, M. Yang, X. Cheng, "Event-trigger rolling horizon optimization for congestion management considering peer-to-peer energy trading among MGs," *International Journal of Electrical Power & Energy Systems*, 147:108838, 2023.
[70] H. Zhou, A. Aral, I. Brandic, M. Kantarci, "Multiagent Bayesian deep reinforcement learning for MG energy management under communication failures," *IEEE Internet of Things Journal*, vol. 9(14), pp. 11685-11698, 2022.
[71] H. Zou, S. Mao, Y. Wang, F. Zhang, X. Chen, L. Cheng, "A survey of energy management in interconnected multi-MGs," *IEEE Access*, vol. 7, pp. 72158–72169, 2019.
[72] G. Chehade, and I. Dincer, "Development and analysis of a polygenerational smart energy hub for sustainable communities," *Energy Conversion and Management*, vol. 226, no. 113475, Dec. 2020.
[73] H. Tian, H. Zhao, C. Liu et al., "A dual-driven linear modeling approach for multiple energy flow calculation in electricity–heat system," *Applied Energy*, vol. 314, no. 118872, May 2022.
[74] T. Liu, D. Zhang, and T. Wu, "Standardised modelling and optimisation of a system of interconnected energy hubs considering multiple energies—Electricity, gas, heating, and cooling," E*nergy Conversion and Management*, vol. 205, no. 112410, Feb. 2020.
[75] A. Sagar, A. Kashyap, M. Nasab, "A comprehensive review of the recent development of wireless power transfer technologies for electric vehicle charging systems," *IEEE Access*, vol. 11, pp. 83703-83751, Aug. 2023.
[76] B. Jones, V. Tien, and R. Elliott, "The electric vehicle revolution: Critical material supply chains, trade and development," *The World Economy*, vol. 46(1), pp. 2-26, Oct. 2022.

[77] M. Ata, A. Erenoglu, I. Sengor et al., "Optimal operation of a multi-energy system considering renewable energy sources stochasticity and impacts of electric vehicles," *Energy*, vol. 186, no. 115841, Nov. 2019.
[78] T. Alabi, L. Lu, Z. Yang, "Improved hybrid inexact optimal scheduling of virtual powerplant (VPP) for zero-carbon multi-energy system (ZCMES) incorporating Electric Vehicle (EV) multi-flexible approach," *Journal of Cleaner Production*, vol. 326, no. 129294, Dec. 2021.
[79] W. Yang, J. Guo, A. Vartosh, "Optimal economic-emission planning of multi-energy systems integrated electric vehicles with modified group search optimization," *Applied Energy*, vol. 311, no. 118634, Apr. 2022.
[80] S. Xie, Z. Hu, J. Wang, Y. Chen, "The optimal planning of smart multi-energy systems incorporating transportation, natural gas and active distribution networks," *Applied Energy*, vol. 269, no. 115006, Jul. 2020.
[81] S. Samanta, D. Roy, S. Roy et al., "Techno-economic analysis of a fuel-cell driven integrated energy hub for decarbonising transportation," *Renewable and Sustainable Energy Reviews*, vol. 179, no. 113278, Jun. 2023.
[82] P. Li, W. Sheng, Q. Duan et al., "A Lyapunov optimization -based energy management strategy for energy hub with energy router," *IEEE Transactions on Smart Grid*, vol. 11(6), no. 4860-4870, Nov. 2020.
[83] W. Huang, E. Du, T. Capuder et al., "Reliability and vulnerability assessment of multi-energy systems: an energy hub -based method," *IEEE Transactions on Power Systems*, vol. 36(5), pp. 3948-3959, Sep. 2021.
[84] A. Jordehi, M. Javadi, J. Catalao, "Day-ahead scheduling of energy hubs with parking lots for electric vehicles considering uncertainties," Energy, vol. 229, no. 120709, Aug. 2021.
[85] D. Cao, J. Zhao, W. Hu et al., "Model-free voltage control of active distribution system with PVs using surrogate model-based deep reinforcement learning," *Applied Energy*, vol. 306(A), no. 117982, Jan. 2022.
[86] Y. Tao, J. Qiu, S. Lai et al., "Integrated electricity and hydrogen energy sharing in coupled energy systems," *IEEE Transactions on Smart Grid*, vol. 12(2), pp. 1149-1162, Mar. 2021.
[87] Y. Liu, D. Zhang, H. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power Energy Systems*, vol. 6, no. 3, Sep. 2020.
[88] D. Qiu, Z. Dong, X. Zhang et al., "Safe reinforcement learning for realtime automatic control in a smart energy-hub," *Applied Energy*, vol. 309, no.118403, 2022.
[89] W. Li, T. Qian, Y. Zhang et al., "Distributionally robust chance-constrained planning for regional integrated electricity–heat systems with data centers considering wind power uncertainty," *Applied Energy*, vol. 336, no. 120787, Apr. 2023.
[90] M. Chen, H. Lu, X. Chang et al., "An optimization on an integrated energy system of combined heat and power, carbon capture system and power to gas by considering flexible load," *Energy*, vol. 273, no. 127203, Jun. 2023.
[91] Y. Dong, H. Zhang, C. Wang et al., "Soft actor-critic DRL algorithm for interval optimal dispatch of integrated energy systems with uncertainty in demand response and renewable energy," *Engineering Applications of Artificial Intelligence*, vol. 127, no. 107230, Jan. 2024.

[92] A. Shakya, G. Pillai, and S. Chakrabarty, "Reinforcement learning algorithms: a brief survey," *Expert Systems with Applications*, vol. 231, no. 120495, Nov. 2023.
[93] "Dataport." Pecan Street. 2018. [Online]. Available: https://www.pecanstreet.org/dataport/.
[94] T. Chen, S. Bu, and X. Liu et al., "Peer-to-peer energy trading and energy conversion in interconnected multi-energy MGs using multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13(1), pp. 715-727, Jan. 2022.
[95] "Electricity Price." ISO New England. 2018. [Online]. Available: https://www.iso-ne.com/.
[96] Z. Qin, H. Hua, H, Liang, R. Herzallah, Y. Zhou, J. Cao, "Optimal electricity trading strategy for a household MG," *Proceedings of 16th IEEE International Conference Control Automation 2020*:1308-1313.
[97] "Commercial and Residential Hourly Load Profiles for all TMY3 Locations in the United States," 2011. [Online]. Available: https://openei.org/datasets/.
[98] S. Silwai, C. Mullican, Y. Chen, A. Ghosh, J Dilliott, J. Kleissi, "Open-source multi-year power generation, consumption, and storage data in a MG," *Journal of Renewable and Sustainable Energy*, vol. 13(2), no. 025301, 2021.
[99] G. Ceusters, R. Rodriguez, A. Garcia et al., "Model-predictive control and reinforcement learning in multi-energy system case studies," *Applied Energy*, vol. 303, no. 117634, 2021.
[100] G. Zhang, W. Hu, D. Cao et al., "Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach," *Energy Conversion and Management*, vol. 227, no. 113608, 2021.
[101] S. Li, W. Hu, Di Cao et al., "Electric vehicle charging management based on deep reinforcement learning," *Journal of Modern Power System and Clean Energy*, vol. 10(3), pp. 719-730, 2021.
[102] S. Zhang, W. Gu, H. Lu et al., "Superposition-principle based decoupling method for energy flow calculation in district heating networks," *Applied Energy*, vol. 295, no. 117032, 2021.
[103] H. Lin, Y. Liu, Q. Sun et al., "The impact of electric vehicle penetration and charging patterns on the management of energy hub – A multi-agent system simulation," *Applied Energy*, vol. 230, pp. 189-206, Nov. 2018.
[104] D. Zhu, B. Yang, Y. Liu et al., "Energy management based on multi-agent deep reinforcement learning for a multi-energy industrial park," *Applied Energy*, vol. 311, no. 118636, 2022.
[105] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," arXiv preprint arXiv: 1707.06347v2, 2017.
[106] J. Zhang, J. Yan, D. Infield et al., "Short-term forecasting and uncertainty analysis of wind turbine power based on long short-term memory network and Gaussian mixture model," *Applied Energy*, vol. 241, pp. 229–44, 2019.
[107] N. Liu, L. Tan, H. Sun et al., "Bilevel heat-electricity energy sharing for integrated energy systems with energy hubs and prosumers," *IEEE Transactions on Industrial Informatics*, vol. 18(6), pp. 3754-3765, Sep. 2021.

[108] D. Connolly, K. Hansen, D. Drysdale et al., "Stratego/heat roadmap europe 3: Enhanced heating and cooling plans to quantify the impact of increased energy efficiency in EU member states," Aalborg University, Denmark, 2016.
[109] A. Ashfaq, A. Ianakiev, "Cost-minimized design of a highly renewable heating network for fossil-free future," *Energy*, vol. 152, pp. 613-626, Jun. 2018.
[110] Y. Wang, N. Zhang, Z. Zhuo et al., "Mixed-integer linear programming-based optimal configuration planning for energy hub: Starting from scratch," *Applied Energy*, vol. 210, pp. 1141-1150, 2018.
[111] R. Davood, B. Hassan, "Probabilistic optimization in operation of energy hub with participation of renewable energy resources and demand response," *Energy*, vol. 173, pp. 384-399, Apr. 2019.
[112] Z. Zhang, F. Liu, and T. Liu et al., "A persistent-excitation-free method for system disturbance estimation using concurrent learning," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 70, no. 8, Aug. 2023.