Multi-Agent Deep Reinforcement Learning Algorithms for Distributed Charging Station Management

Li Junda[1], Wang Tianan[2], Zhang Dingyi[3], Wu Quancai[4], Liu Jian[5]*
Yunnan Power Grid Energy Investment Co., Ltd., Kunming, Yunnan Province, 650000, China[1, 2 3, 4]
Sinotrans Logistics Southwest Co., Ltd., Chengdu, Sichuan Province, 610000, China[5]

***Abstract*—With the continued growth of the electric vehicle (EV) fleet, the issue of cross-regional coordinated scheduling for charging infrastructure has become increasingly prominent, facing challenges such as uneven resource allocation and delayed responses. Considering the complex coupling between charging stations and the power system in a smart grid environment, this paper proposes a distributed scheduling strategy based on multi-agent deep reinforcement learning (MADRL) to achieve efficient, coordinated management of charging infrastructure and power resources. The proposed approach constructs a hierarchical decision-making architecture to jointly optimize intra-regional resource allocation and cross-regional power support, modeling the scheduling process as a Markov Decision Process (MDP) and treating regional charging stations, power nodes, and material units as independent agents. Through the multi-agent deep reinforcement learning mechanism, each agent autonomously learns optimal scheduling policies in the presence of uncertain demand and supply fluctuations, thus enabling rapid response and enhancing system robustness. Simulation results demonstrate that the proposed method effectively reduces scheduling costs and improves resource utilization and service quality. This study provides both theoretical support and practical pathways for building intelligent, efficient, and sustainable charging infrastructure.**

Keywords*—Charging station scheduling; cross-regional coordination; multi-agent systems; deep reinforcement learning; Markov decision process; resource optimization; uncertainty response

I. Introduction

Driven by the global transition in energy structures and the rise of sustainable transportation strategies, the adoption of electric vehicles (EVs) is accelerating rapidly, making them a cornerstone of future low-carbon mobility. This trend places increasing demands on the large-scale deployment and intelligent management of charging infrastructure, especially charging stations. As the terminal nodes of the power grid, charging stations not only provide essential energy replenishment for EVs but also significantly influence grid security, traffic efficiency, and user experience through their operational efficiency, scheduling strategies, and resource allocation.

With the ongoing growth in the number of EVs, charging demands have become highly unbalanced across regions and exhibit strong spatiotemporal fluctuations. On one hand, certain areas have an abundance of charging stations but suffer from low utilization rates; on the other, core urban zones face resource shortages, severe queuing, and service bottlenecks due to concentrated demand. Furthermore, the operation of charging station networks is affected by multiple factors, including power supply capacity, maintenance resources, and traffic conditions, making cross-regional resource coordination and scheduling a pressing challenge.

Existing research has made significant progress in optimizing charging station scheduling, covering aspects such as queue management, load balancing, dynamic pricing, and energy storage integration. However, most approaches are still based on single-region, centralized architectures, lacking comprehensive consideration for cross-regional resource flows, heterogeneous system coordination, and dynamic uncertainties. Particularly in multi-region systems, the scheduling problem becomes highly complex and strongly coupled. Traditional optimization algorithms struggle to address challenges such as real-time requirements, large state spaces, and non-convex objectives, thus limiting their effectiveness in large-scale, practical deployments.

To address these challenges, this paper proposes a cross-regional coordinated charging station scheduling method based on multi-agent deep reinforcement learning (MADRL). In the proposed approach, each node in a multi-region charging station network is modeled as an autonomous agent. Through environmental perception and policy learning, agents achieve both local resource optimization and cross-regional coordination. Specifically, the scheduling problem is formulated as a Markov Decision Process (MDP), with state variables including charging demand, power supply, equipment status, and traffic accessibility. Each agent makes decisions independently based on local information and optimizes its policy collaboratively through a shared reward mechanism. This approach offers strosng generalization and adaptability, effectively coping with dynamic changes in charging demand and power supply, and overcomes the performance bottlenecks of traditional methods when faced with uncertainty and high-dimensional state spaces.

The main objective of this study is to build a distributed, intelligent, self-learning, and collaboratively optimized framework for cross-regional charging station resource scheduling. The aim is to maximize resource utilization, minimize operational costs, and optimize service

*Corresponding Author.

responsiveness. Compared with existing methods, the main innovations of this work include:

1) Proposing a coordinated scheduling modeling framework for cross-regional, multi-node, and multi-resource-type systems, adaptable to highly complex operational environments.

2) Introducing a multi-agent deep reinforcement learning mechanism to enable information exchange and policy collaboration among regional agents;

3) Designing a learning structure based on state sharing and reward feedback to enhance responsiveness to charging demand uncertainties and power resource fluctuations;

4) Validating the proposed method through simulation experiments, demonstrating significant advantages in scheduling efficiency, resource utilization, and system robustness.

This research is of both theoretical and practical significance. On one hand, it offers a novel modeling and solution paradigm for charging infrastructure scheduling, broadening the application scope of multi-agent systems in the integration of transportation and energy. On the other hand, it provides effective technical support for regional resource coordination, optimized energy allocation, and improved user charging experience—contributing to the development of smart cities, green transportation, and regional energy collaboration.

The remainder of this paper is organized as follows: Section II reviews the progress in charging station scheduling and the application of multi-agent reinforcement learning in energy management; Section III presents the proposed system modeling methodology and the multi-agent learning framework is given in Section IV; Section V conducts simulation experiments and performance evaluations based on a multi-region charging network; and Section VI concludes the study and outlines future research directions.

II. Literature Review

A. Techniques Based on Acoustic Features Multi-Agent Reinforcement Learning Algorithm Development

According to different decision-making paradigms, multi-agent reinforcement learning (MARL) algorithms for distributed charging scheduling can be categorized into value-based methods and policy-based methods.

For value-based methods, agents focus on learning value functions to derive optimal strategies. In cooperative MARL for distributed charging networks, value-based methods mainly address how to decouple the centralized team value function for distributed execution. Sunehag et al. [1] proposed value-decomposition networks that break down the team value function into a linear sum of individual agent values, where the optimal policy is obtained by greedy selection on the team value. Rashid et al. [2] further enhanced algorithm performance by representing the joint value function as a nonlinear monotonic combination and giving higher weight to joint actions with greater rewards, thus extending the approach to non-monotonic environments. Son et al. [3] introduced a transformation-based factorization that avoids both monotonicity and additivity constraints. However, these methods often rely on regularization for tractable computation, which can hinder performance in complex charging environments. Mahajan et al. [4] addressed inefficient exploration by proposing multi-agent variational exploration, which improves coordination over extended time horizons.

For policy-based methods, when all agents update their strategies simultaneously, the environment becomes non-stationary, making learning more challenging. Thus, most policy-based MARL algorithms adopt the actor–critic (AC) framework. To mitigate partial observability, techniques such as value function decomposition and the use of centralized critics with additional information exchange during training are commonly adopted; both types typically employ the CTDE (Centralized Training, Distributed Execution) paradigm. For value decomposition, Su et al. [5] designed value-decomposition actor-critic methods using a monotonic mapping between the global and local state values, following a simple time-difference advantage gradient that improves sampling efficiency and converges to local optima. Yang et al. [6] implemented a determinant point process-based method for unconstrained value function decomposition.

In centralized critic approaches, Foerster et al. [7] developed counterfactual multi-agent policy gradients, where centralized critics access joint actions and all agent states, while each agent's policy depends only on its own observation history. Pu et al. [8] constructed a decomposed soft actor–critic method with discrete probability policies and counterfactual advantage functions, supporting efficient policy learning and partially resolving credit assignment for both discrete and continuous action spaces. Lowe et al. [9] proposed Multi-Agent Deep Deterministic Policy Gradient (MADDPG), assigning centralized critics to each agent to support different reward functions in competitive environments. Building on this, Wang et al. [10] extended MADDPG to partially observable settings, utilizing recurrent neural networks in both actor and critic to retain observation history. Li et al. [11] integrated minimax optimization into robust multi-agent reinforcement learning, enabling agents to learn robust strategies under adversarial conditions.

These MARL algorithms have been applied to complex scheduling scenarios in distributed charging networks for electric vehicles, enabling agents to collaboratively optimize charging schedules and energy management. However, despite the effectiveness of CTDE in addressing partial observability, issues such as agent privacy and single-point failure risks remain unsolved in large-scale distributed charging systems.

B. Engineering Applications of Multi-Agent Reinforcement Learning

Multi-agent reinforcement learning (MARL) methods have been widely extended to optimal control problems in engineering, such as traffic control [12], autonomous driving [13, 14], and base station communications [15]. As an emerging distributed decision-making technique, MARL has attracted considerable attention for its ability to address nonlinear objectives, which motivates its application in distributed charging scheduling and energy management for smart grids.

In the context of distributed energy (and charging) management, most MARL approaches adopt the CTDE (Centralized Training, Distributed Execution) paradigm to mitigate the adverse effects of partial observability. Zhang et al. [16] proposed a novel deep transfer Q-learning algorithm based on a virtual leader–follower model to maximize the total revenue of all agents while maintaining supply–demand balance in smart grid scenarios. Compared to heuristic optimization methods, deep transfer Q-learning achieves faster convergence, stronger online learning ability, and effective privacy protection for users. Wang et al. [10] introduced a cooperative fuzzy Q-learning approach for microgrid energy management, ensuring stable power supply for independent microgrids while considering user demand uncertainty and achieving rapid acquisition of management strategies.

Reinforcement learning-based strategies have also been designed for distributed energy and load management in competitive and stochastic energy markets, such as microgrid auction-based time-sharing markets, where model-free Q-learning ensures each agent can find its optimal strategy. Zhu et al. [17] proposed an attention mechanism and soft actor–critic-based method for multi-energy-coupled energy management under renewable energy and demand uncertainty, using counterfactual baselines to accelerate policy learning and minimize long-term energy costs while ensuring user needs. Sun et al. [18] developed a multi-agent energy management optimization framework for integrated energy systems considering electricity, natural gas, and carbon trading, using MADDPG to provide each agent with an online trading strategy that considers individual interests for fair market transactions and privacy protection.

To address deployment challenges in large-scale systems, a few studies have explored fully distributed MARL algorithms. Li et al. [19] designed a distributed Q-learning method under a fully decentralized control framework to solve nonconvex economic dispatch problems, though its optimization accuracy is limited due to the lack of state-action value function approximation. In contrast, Liu et al. [20] used nonlinear function approximation for value functions and introduced a diffusion strategy to enable agent collaboration, yet the convergence range of the value function fitting remains limited. Dai et al. [21] combined value-based methods with quadratic function approximation to handle decision-making in continuous action spaces, but the quadratic approximation is most suitable for convex optimization problems. Li et al. [22] proposed a fully distributed reinforcement learning algorithm for nonconvex economic dispatch, but their stateless design for static scheduling problems does not fully exploit the generalization power of deep learning.

For distributed charging and energy management, the above CTDE-based methods are difficult to deploy across widely-distributed charging infrastructure, and still face challenges such as centralized method security, privacy risks, and high communication costs.

C. Research Gaps

This section analyzes the main strengths and limitations of existing multi-agent reinforcement learning (MARL) approaches for distributed scheduling of heterogeneous charging stations, focusing on decision-making types, handling of partial observability, training schemes, and function approximation. For methods applied to distributed charging scheduling, we further compare their optimization objectives, scheduling precision, and generalization capability under uncertainty. Overall, MARL demonstrates outstanding performance in tackling complex nonlinear scheduling objectives and demand uncertainties, offering an effective research paradigm for distributed scheduling in intelligent charging networks. However, current MARL methods still face several challenges when applied to heterogeneous charging station systems:

1) Observability and deployment issues: To address information barriers caused by partial observability, most existing algorithms adopt a centralized training, distributed execution (CTDE) framework. However, this paradigm is difficult to deploy efficiently across widely distributed charging station network nodes, restricting its practicality and scalability.

2) Constraint handling and sparse rewards: Most current methods translate operational constraints into penalty terms within the reward function, which often leads to sparse reward signals. This sparsity can negatively affect learning efficiency and the feasibility of the final scheduling policy.

Therefore, a key motivation of this paper is to theoretically analyze the impact of partial observability on the convergence of MARL, design a fully distributed MARL algorithm suitable for multi-region charging station systems, and introduce an action space mapping mechanism to strictly confine the decision-making process within the feasible domain during training. These innovations aim to provide more efficient and practical technical solutions for intelligent scheduling in large-scale charging networks.

III. Problem Modeling

A. System Structure for Heterogeneous Charging Station Scheduling

This section investigates a coordinated scheduling system for multiple regions and heterogeneous charging stations, as illustrated in Fig. 1. The system comprises various types of charging equipment, including Fast DC Chargers (FDC), AC Slow Chargers (ASC), and bidirectional V2G (Vehicle-to-Grid) Chargers (V2G).

As shown in Fig. 1, within the multi-region microgrid balancing charging supply and demand, each region can be connected to the interregional grid via substations. The system considers two types of regions: source-load coordination regions (e.g., Region A) and dedicated charging regions (e.g., Region D). Source-load coordination regions integrate multiple types of charging stations (FDC, ASC, V2G) as well as end-users (EVs), while dedicated charging regions deploy only a single type or a subset of charging stations.

In the heterogeneous charging station system discussed in this chapter, each group of charging stations of the same type can be regarded as an independent agent. FDC and ASC can autonomously determine their charging power allocation

according to real-time demand, while V2G chargers can flexibly switch between charging and discharging modes, enabling bidirectional energy flow. Operators are able to sense the local charging demand of EVs in real time and transmit the relevant information to the respective charging station agents. For information flow, both intra- and inter-regional communication among charging station agents is realized through remote point-to-point links, facilitating distributed learning and optimal scheduling decisions.

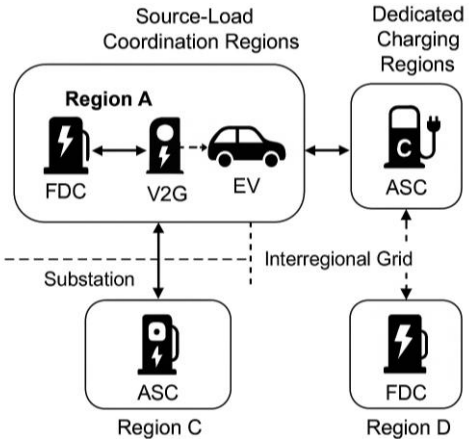


Fig. 1. Schematic diagram of the heterogeneous charging station scheduling system.

The main energy flow in this system is electrical power. The internal structure of the regional charging network is depicted in Fig. 2, where the Charging Station (CS) serves as the core node.

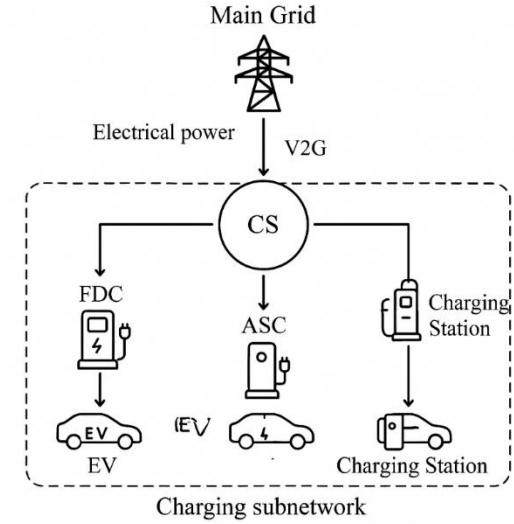


Fig. 2. Integrated heat-electricity network within a region.

The regional charging network can be subdivided into the main grid and the charging subnetwork. Energy flows from the main grid to users (EVs) via different types of charging stations, achieving efficient power supply based on demand. Power demand can be collaboratively met by FDC, ASC, V2G, and charging stations in other regions. V2G chargers can also feed energy from vehicle batteries back to the main grid, enhancing the flexibility of energy regulation. According to the described energy flow and conversion processes, the charging subnetwork mainly serves intra-regional needs, while the main grid supports interregional energy allocation. This setup aligns with the actual requirements of power flow and distributed scheduling.

During the charging scheduling process, priority is given to meeting the charging needs of local users, and any remaining capacity can be used to support grid load balancing. Hierarchical scheduling strategies can effectively improve the utilization efficiency of renewable energy, reduce grid pressure, and advance the goal of green and intelligent mobility.

B. Multi-Domain Integrated Energy System Component Modeling

Based on the structure of the multi-region, multi-type charging station system, the primary components include Fast DC Chargers (FDC), AC Slow Chargers (ASC), bidirectional V2G Chargers (V2G), and Charging Stations (CS). Mathematical models are developed for each component to formulate the distributed charging scheduling problem, as described below:

1) Fast DC charger: FDCs serve as the main high-power charging facilities in the system, primarily catering to EVs with fast-charging needs. The charge/discharge power ratio of FDC can be expressed as:

$$\alpha_{FDC} = \frac{P_{FDC}^{out}}{P_{FDC}^{in}} \tag{1}$$

where P_{FDC}^{out} and P_{FDC}^{in} represent the output and input power of the FDC, respectively. The operational cost function is modeled as a coupled quadratic function:

$$C_{FDC}(P) = aP_{FDC}^2 + bP_{FDC} + c \tag{2}$$

where C_{FDC} denotes the instantaneous operating cost of the FDC, and $a\ b\ c$ are intrinsic parameters. The energy constraint is:

$$0 < P_{FDC} < P_{FDC}^{max} \tag{3}$$

where P_{FDC}^{max} is the FDC's maximum output power.

2) AC Slow charger: ASCs mainly provide charging services for EVs with longer dwell times. The cost function is:

$$C_{ASC}(P) = \alpha P_{ASC}^2 + \beta P_{ASC} + \gamma \tag{4}$$

where P_{ASC} and C_{ASC} denote the output power and instantaneous cost of ASC, and α, β, γ are intrinsic parameters. The power constraint is:

$$P_{ASC}^{min} < P_{ASC} < P_{ASC}^{max} \tag{5}$$

3) Bidirectional V2G charger: V2G chargers enable bidirectional energy flow, allowing energy feedback to the grid while meeting user demand. The charge/discharge ratio is defined as:

$$\alpha_{V2G} = \frac{P_{V2G}^{\text{out}}}{P_{V2G}} \tag{6}$$

where P_{V2G} is the total V2G power, and P_{V2G}^{out} is the power fed back to the grid. The operating cost is:

$$C_{V2G}(P) = gP_{V2G} + h \tag{7}$$

where C_{V2G} is the instantaneous cost, and g, h are intrinsic parameters.

4) Charging station: As the energy dispatch center, CS coordinates energy allocation among regional charging stations. Its dispatch capacity constraint is:

$$0 < P_{CS} < P_{CS}^{max} \tag{8}$$

where P_{CS}^{max} is the maximum dispatch capacity.

C. Distributed Charging Scheduling Problem

In this study, multi-agent deep reinforcement learning (MADRL) is employed to derive optimal economic scheduling and energy feedback policies, aiming to minimize overall operational costs. The distributed energy management problem is formulated as follows:

Objective Function is as follows:

$$C = \sum_{i=1}^{n} C_{FDC,i} + \sum_{i=1}^{m} C_{ASC,i} + \sum_{i=1}^{q} C_{V2G,i} \tag{9}$$

where C is the total operating cost, and n, m, q are the numbers of FDC, ASC, and V2G chargers, respectively.

Capacity Constraints is as follows:

$$\begin{cases} 0 < P_{FDC,i} < P_{FDC,i}^{max}, & \forall i = 1,2,\dots,n \\ P_{ASC,i}^{min} < P_{ASC,i} < P_{ASC,i}^{max}, & \forall i = 1,2,\dots,m \\ 0 < P_{V2G,i} < P_{V2G,i}^{max}, & \forall i = 1,2,\dots,q \end{cases} \tag{10}$$

Power Demand Balance Constraint is as follows:

$$\sum_{i=1}^{c} P_{D,i} = \sum_{i=1}^{n} P_{FDC,i} + \sum_{i=1}^{m} P_{ASC,i} + \sum_{i=1}^{q} \left[(1 - \alpha_{V2G,i})P_{V2G,i}\right] \tag{11}$$

where $P_{D,i}$ is the charging demand of user i, and c is the total number of users.

Energy Feedback Balance Constraint is as follows:

$$\sum_{i=1}^{c_I} P_{D,i}^{\text{feed}} = \sum_{i=1}^{q_I} P_{V2G,i}^{\text{out}}, I = 1,2,\dots,N \tag{12}$$

where $P_{D,i}^{\text{feed}}$ is the feedback energy requirement of user i, q_I is the number of V2G chargers in region I, and N is the total number of regions.

IV. Multi-Agent Deep Reinforcement Learning for Distributed Energy Management

The main method proposed in this chapter addresses the above distributed charging scheduling problem by developing a multi-agent deep reinforcement learning (MADRL) algorithm under a partially observable environment. The following describes the MADRL algorithm framework for heterogeneous charging station systems and discusses relevant technical details.

A. Algorithm Overview

The multi-agent deep reinforcement learning framework developed for heterogeneous charging station systems is illustrated in Fig. 3. In this framework, each FDC, ASC, and V2G charger group acts as an independent agent, employing a static optimization deep learning approach. For each agent i, the POMDP is formalized as $< S_i, A_i, P_i, R_i >$, and the Markov decision process for each charger type is modeled as follows:

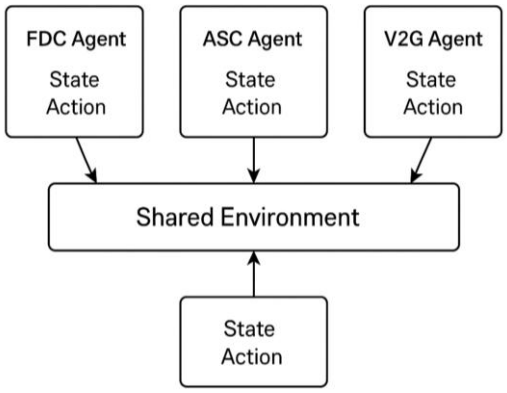


Fig. 3. Multi-agent deep reinforcement learning framework for heterogeneous charging station scheduling.

1) State

a) FDC agent: State vector includes global total charging demand $\sum_{i=1}^{c} P_{D,i}$, local fast charging demand, and total V2G charge/discharge in the region.

b) ASC agent: State vector includes global total charging demand, local slow charging demand, and total V2G charge/discharge in the region.

c) V2G agent: State vector includes global total charging demand, available V2G capacity in the region, and loads of all charger types in the region.

2) Action: For agent i:

a) FDC agent: The output power $P_{FDC,i}$ is the action.

b) ASC agent: The output power $P_{ASC,i}$ is the action.

c) V2G agent: The action vector consists of charging/discharging power $P_{V2G,i}$ (positive for charging, negative for discharging) and the feedback ratio $\alpha_{V2G,i}$.

3) State transition probability: Due to the stochastic nature of charging demand and vehicle arrivals, the state transition distribution for all agents i is assumed to be uniform.

4) Reward: The local reward of each charging station agent is negatively correlated with the objective function above. If an agent fails to satisfy constraints such as capacity or power, its reward is penalized.

This framework considers two deep learning approaches: Target Value Competition Multi-Agent Reinforcement Learning (MADRL-TVC) and Multi-Agent Deep Deterministic Policy Gradient (MADDPG). MADRL-TVC is a

fully distributed method, while MADDPG requires centralized training. Both methods are applied to the distributed charging scheduling problem, with the superior approach selected for the charging station system scheduling policy.

B. Algorithm Procedure

The algorithm used in this chapter is a model-dependent continuous decision-making method, not episode-based, and does not rely on historical sequences, but only on real-time decisions and feedback from the multi-agent system. The implementation process of the MADRL algorithm is shown in Fig. 4, with the core steps as follows:

1) Initialization and state acquisition: Each charging station agent initializes its parameters and obtains the initial state from the environment. State information includes the current load demand of each charger type, user arrival information, available V2G capacity, etc. At each decision step t, each agent selects action $a_{i,t}$ according to its policy.

2) Hierarchical scheduling and action adjustment: To balance different types of demand, a hierarchical control framework is designed for high-priority intra-regional charging needs and cross-regional energy feedback (V2G). Actions are adjusted automatically via distributed binary search, prioritizing high-priority charging demand within each region.

For balancing fast and slow charging demands, the initial search domain for FDC and ASC agents is set as:

$$a_i^{min}(0) = 0, a_i^{max}(0) = P_{FDC,i}^{max} \text{ or } P_{ASC,i}^{max} \tag{13}$$

For V2G feedback, the agent's search domain is:

$$a_i^{min}(0) = -P_{V2G,i}^{max}, a_i^{max}(0) = P_{V2G,i}^{max} \tag{14}$$

The load demand deviation for region I is defined as:

$$\delta_{P,I}(k) = \sum_{i=1}^{c_I} P_{D,i} - \sum_{i=1}^{n_I} P_{FDC,i} - \sum_{i=1}^{m_I} P_{ASC,i} - \sum_{i=1}^{q_I} \alpha_{V2G,i} P_{V2G,i} \tag{15}$$

Running Algorithm 1 with this setup achieves balance between regional charging demand and V2G feedback.

3) Experience collection and network training: Each agent observes the immediate reward r_{t+1} and the next state $s_{i,t+1}$. After each decision, the experience is stored in a replay buffer. At regular intervals, mini-batches are randomly sampled from the buffer to train the deep neural network.

Based on the above approach, charging station agents collaboratively learn optimal energy allocation and economic scheduling policies to minimize the overall system operating cost. The trained model can respond online to varying charging demands, EV arrivals and departures, and random V2G availability, providing real-time scheduling strategies. This method excels at rapid demand response, accommodating renewable energy uncertainty, and improving overall energy utilization efficiency.

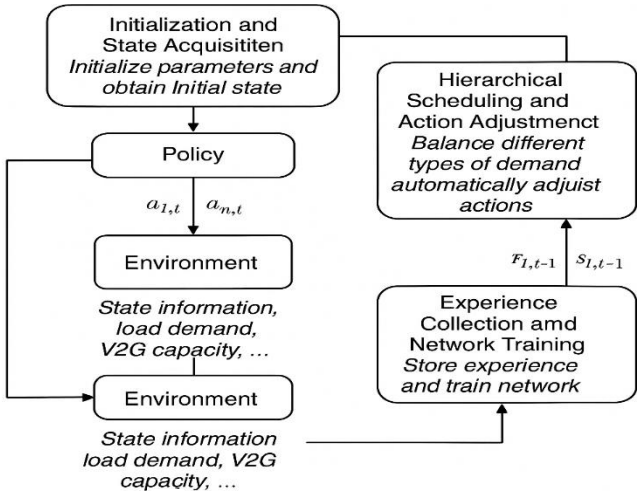


Fig. 4. Multi-agent deep reinforcement learning algorithm flow for multi-domain integrated energy systems.

V. Simulation Validation

In this section, a multi-region, multi-type charging station scheduling scenario is constructed. The proposed hierarchical decision-making framework and multi-agent reinforcement learning method are trained, and the trained models are tested using open-source transportation and charging load datasets. Their responsiveness to load uncertainty and distributed charging demand is analyzed and compared with other benchmark scheduling methods, further validating the performance advantages of the proposed distributed charging scheduling scheme in multi-region scenarios.

A. Simulation Setup

The simulation environment is configured as a multi-type charging station system spanning four regions, as illustrated in Fig. 1. The entire system comprises nine groups of charging stations, including four Fast DC Charger (FDC) groups (deployed in Area A: 1 group, Area C: 1 group, Area D: 2 groups), two bidirectional V2G charger groups (deployed in Area A and Area C), and three AC Slow Charger (ASC) groups (deployed in Area A, Area B, and Area C). Cross-regional power distribution can be achieved via the main grid. Data on vehicle charging demand, EV arrivals and departures, and V2G feedback capabilities for each region are sourced from open-source transportation and charging datasets (such as PEMS, Open Charge Map, etc.). The cost parameters and inequality constraints for each type of charging station are integrated from previous studies or relevant industry standards. All simulation parameters are detailed in Tables I, II, and III. The scheduling interval is set to 15 minutes, and the simulation covers a 24-hour period.

TABLE I. FDC Unit Parameters

NodeRegion	NodeRegion	α_i	β_i	γ_i	$P_{FDC}^{\min}$	$P_{FDC}^{\max}$
1	A	0.0020	10	500	100	600
2	C	0.0025	8	300	50	400
3	D	0.0050	6	100	50	300
4	D	0.0060	5	90	50	200

B. Training Process

Due to the incomparability of immediate rewards across different reward functions, it is challenging to directly track the cumulative rewards during the training process under random states. To evaluate the algorithm's performance during training, 10 random state points are selected and their corresponding rewards are recorded throughout the training. Each observation point includes the power demand and heat demand of Areas A, B, and C, as well as the wind power output from WT unit groups in Areas A and C. To analyze the neural network's fitting performance, the total training loss of the Q-network estimated by the MADRL-TVC method is also recorded during training.

TABLE II. V2G CHARGER AND CS PARAMETERS

NodeRegion	NodeRegion	g_i	h_i	k_{V2G}	P_{V2G}^{max}
5	A	5	15	0.9	90
6	B	4	20	0.9	90

TABLE III. ASC UNIT PARAMETERS

NodeRegion	NodeRegion	a_i	b_i	c_i	P_{CHPP}^{max}	H_{CHPP}^{max}	α_{CHPP}^{min}
7	A	0.0050	6	100	200	100	0.5
8	B	0.0060	5	90	200	90	0.5
9	C	0.0072	3	150	180	80	0.5

Fig. 5 shows the cumulative rewards and total training loss at different observation state points during the training process. It can be observed that the cumulative rewards at all state points exhibit an upward trend. This indicates that the proposed algorithm can improve cumulative rewards through training, thus finding the optimal scheduling strategy for each type of state. The training loss of the Q-network converges rapidly and ultimately becomes very small, demonstrating that the neural network can effectively fit the charging scheduling policy.

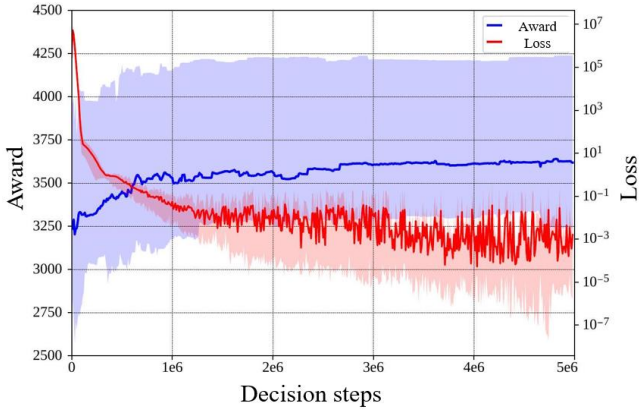


Fig. 5. Cumulative rewards and total training loss during training.

C. Full-Day Energy Management Results Analysis

After extended training under randomly varying states, the proposed algorithm achieves a generalizable, complete model that can provide optimal charging scheduling strategies in real time for arbitrary user demand and V2G capacity availability. Within a one-day scheduling cycle, the MADRL-TVC trained model outperforms MADDPG, so the following presents the optimal strategies under MADRL-TVC scheduling.

The energy allocation of V2G chargers is shown in Fig. 6. For each group of V2G chargers, the energy allocation includes charging power output (P_{V2G}), energy fed back to the grid (F_{V2G}), and energy loss (L_{V2G}). As illustrated in Fig. 6(a) and 6(b), the V2G chargers in Area A and Area C can fully utilize vehicle battery energy for grid feedback during peak load periods, effectively alleviating regional power supply pressure. The V2G energy utilization rate reaches up to 100%.

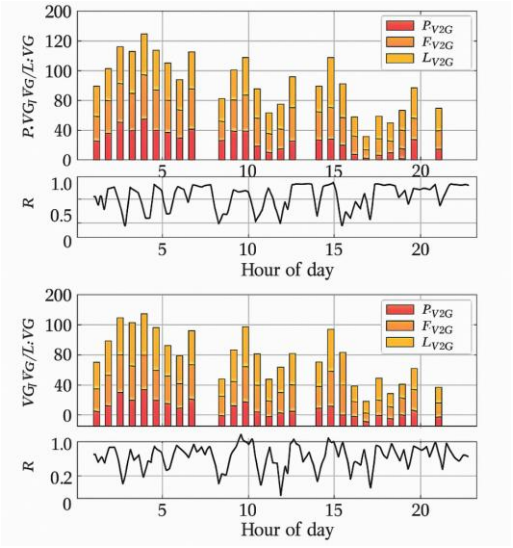
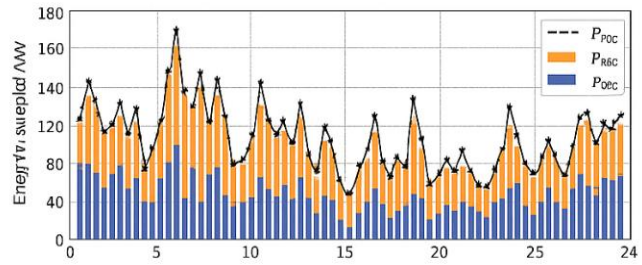


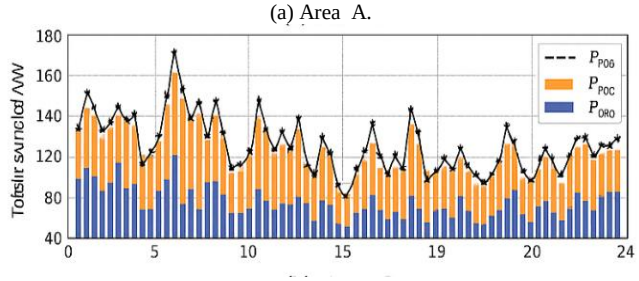
Fig. 6. Regional V2G charger energy allocation and feedback rate.

The scheduling results for charging demand within each region are shown in Fig. 7, displaying how different types of charging stations meet vehicle charging needs in each region for every 15-minute interval throughout the day. P_{FDC}, P_{ASC}, and P_{V2G} represent the energy supplied by each type of charging station, while PD_A, PD_B, and PD_C denote the total demand in each region. The results demonstrate that the charging demands in all regions are precisely met. Due to the stochastic nature of charging demand and vehicle arrivals, the energy allocation structure across regions does not follow a fixed pattern. Nevertheless, the algorithm is able to provide corresponding scheduling strategies in real time, effectively coping with demand and load uncertainty.

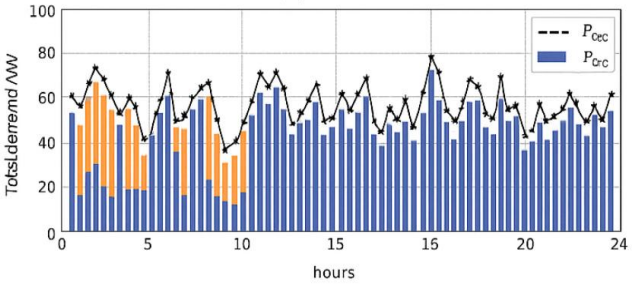
Cross-regional energy scheduling is depicted in Fig. 8, showing the overall charging power allocation at each time interval throughout the day—that is, how each type of charging station and each region jointly respond to the total charging demand. P_{Area_A}, P_{Area_B}, P_{Area_C}, and P_{Area_D} represent the total charging output for each region. The charging power is balanced at all times of the day, with the FDCs (which have higher economic priority) typically handling the main load, while ASCs provide supplementary regulation. V2G chargers mainly contribute energy feedback during peak load periods. The scheduling results indicate that the agents can make distributed collaborative decisions to achieve optimal economic scheduling.



(a) Area A.



(b) Area B



(c) Area C

Fig. 7. Regional charging demand scheduling results.

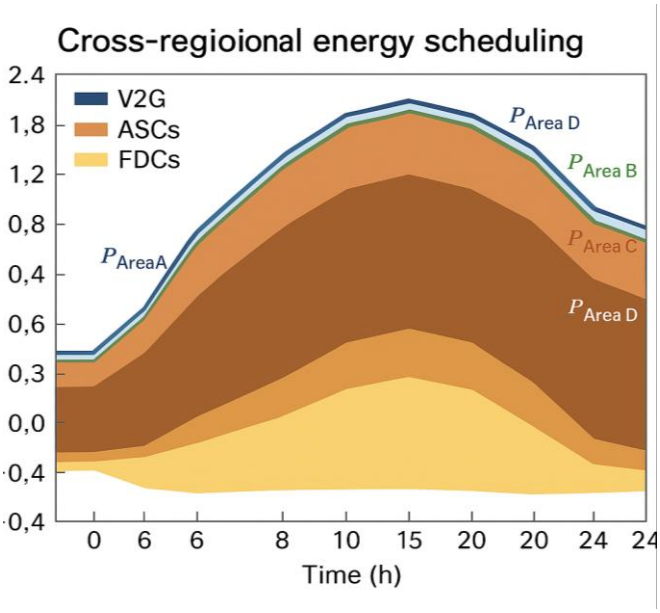


Fig. 8. Overall power scheduling results.

The total system operating cost is shown in Fig. 9, presenting the full-day operating costs for the multi-type charging station system. The results show that the algorithm can effectively reduce overall operating costs and achieve coordinated operation between charging stations and the grid.

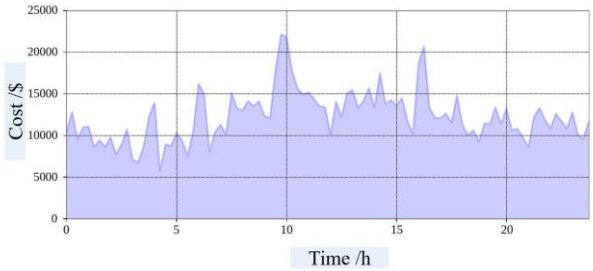


Fig. 9. Full-day operating cost.

D. Comparative Analysis

To evaluate the performance of the distributed charging scheduling method based on multi-agent reinforcement learning (MADRL), it is compared with commonly used centralized optimization methods, including Particle Swarm Optimization (PSO), nonlinear optimization solvers (SCIP, NLOPT), and others. Table IV presents the operating costs of different algorithms at a specific observation state (e.g., total charging demand 800.0, regional demands [100.0, 120.0, 50.0], V2G available power [70.0, 150.0]). Table V lists the average hourly quantified costs for each method over the entire day.

As shown in the tables, compared with centralized stochastic optimization methods such as PSO, the multi-agent deep reinforcement learning approach can effectively reduce system operating costs in distributed charging scheduling. In particular, the MADRL-TVC method achieves performance comparable to mature solvers such as SCIP. Compared with the MADDPG (centralized training, distributed execution) framework, the fully distributed MADRL-TVC displays better convergence in high-dimensional decision spaces, further highlighting the advantages of distributed algorithms.

TABLE IV. OPERATING COST OF DIFFERENT METHODS AT THE OBSERVATION STATE

Algorithm	Solver Type	Optimal Cost	Improvement
PSO	Centralized	12603.94	0
NLOPT	Centralized	12622.88	-0.15%
SCIP	Centralized	12201.77	3.19%
MADDPG	Distributed	12394.07	1.67%
MADRL	Distributed	12202.26	3.19%

TABLE V. AVERAGE QUANTIFIED COST PER HOUR FOR DIFFERENT METHODS

Algorithm	Solver Type	Optimal Cost	Improvement
PSO	Centralized	52102.92	0
NLOPT	Centralized	52977.84	−1.84%
SCIP	Centralized	49129.07	5.56%
MADDPG	Distributed	50968.13	2.02%
MADRL	Distributed	49146.08	5.53.%

The results show that MADRL-TVC not only has advantages in reducing operating costs, but its distributed structure is also better suited to dynamic changes and randomness in charging demand and V2G feedback, offering

stronger generalization and stability. Therefore, the distributed scheduling method based on multi-agent deep reinforcement learning has significant application potential and development prospects in multi-region, multi-type charging station systems.

VI. Conclusion

This Paper proposes a distributed charging scheduling strategy for multi-region, multi-type charging station systems. Different regions and types of charging station groups are modeled as independent agents. Based on user charging demand and V2G feedback capability, collaborative optimization of the multi-agent system enables energy allocation and economic scheduling. For intra-regional fast charging, slow charging, and cross-regional energy feedback, a hierarchical decision-making framework is designed, and the charging scheduling problem is formalized as a Markov Decision Process, with multi-agent deep reinforcement learning employed for collaborative learning. Leveraging the generalization capability of deep reinforcement learning, the proposed method can not only reduce total system operating costs, but also respond in real time to uncertainties in charging demand and energy feedback, effectively improving the economic efficiency and robustness of distributed charging networks.

Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g.” Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks.”

References

[1] Sunehag, P., Lever, G., Gruslys, A., et al. "Value-decomposition networks for cooperative multi-agent learning." arXiv preprint arXiv:1706.05296, 2017.

[2] Rashid, T., Samvelyan, M., de Witt, C. S., et al. "QMIX: Monotonic value function factorisation for deep multi-agent reinforcement learning." Journal of Machine Learning Research, 21(1): 7234–7284, 2020.

[3] Son, K., Kim, D., Kang, W. J., et al. "Qtran: Learning to factorize with transformation for cooperative multi-agent reinforcement learning." In Proceedings of the International Conference on Machine Learning, Long Beach, USA, pp. 5887–5896, 2019.

[4] Mahajan, A., Rashid, T., Whiteson, S., et al. "MAVEN: Multi-agent variational exploration." In Advances in Neural Information Processing Systems, Vancouver, Canada, 2019.

[5] Su, J., Adams, S., Beling, P. "Value-decomposition multi-agent actor-critics." In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Event, pp. 11352–11360, 2021.

[6] Yang, Y., Wen, Y., Wang, J., et al. "Multi-agent determinantal q-learning." In Proceedings of the International Conference on Machine Learning, Virtual Event, pp. 10757–10766, 2020.

[7] Foerster, J., Farquhar, G., Afouras, T., et al. "Counterfactual multi-agent policy gradients." In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, USA, pp. 1–6, 2018.

[8] Pu, Y., Wang, S., Yang, R., et al. "Decomposed soft actor-critic method for cooperative multi-agent reinforcement learning." arXiv preprint arXiv:2104.06655, 2021.

[9] Lowe, R., Wu, Y., Tamar, A., et al. "Multi-agent actor-critic for mixed cooperative-competitive environments." arXiv preprint arXiv:1706.02275, 2017.

[10] Wang, R. E., Everett, M., How, J. P. "R-MADDPG for partially observable environments and limited communication." arXiv preprint arXiv:2002.06684, 2020.

[11] Li, S., Wu, Y., Cui, X., et al. "Robust multi-agent reinforcement learning via minimax deep deterministic policy gradient." In Proceedings of the AAAI Conference on Artificial Intelligence, Hawaii, USA, pp. 4213–4220, 2019.

[12] Wang, Y., Xu, T., Niu, X., et al. "STMARL: A spatio-temporal multi-agent reinforcement learning approach for cooperative traffic light control." IEEE Transactions on Mobile Computing, 21(6): 2228–2242, 2020.

[13] Shalev-Shwartz, S., Shammah, S., Shashua, A. "Safe, multi-agent, reinforcement learning for autonomous driving." arXiv preprint arXiv:1610.03295, 2016.

[14] Chen, S., Dong, J., Ha, P., et al. "Graph neural network and reinforcement learning for multi-agent cooperative control of connected autonomous vehicles." Computer-Aided Civil and Infrastructure Engineering, 36(7): 838–857, 2021.

[15] Gursoy, M. C., Zhong, C., Velipasalar, S. "Deep multi-agent reinforcement learning for cooperative edge caching." In Machine Learning for Future Wireless Communications, pp. 439–457, 2020.

[16] Zhang, X., Bao, T., Yu, T., et al. "Deep transfer Q-learning with virtual leader–follower for supply–demand Stackelberg game of smart grid." Energy, 133: 348–365, 2017.

[17] Zhu, D., Yang, B., Liu, Y., et al. "Energy management based on multi-agent deep reinforcement learning for a multi-energy industrial park." Applied Energy, 311: 1–15, 2022.

[18] Sun, Q., Wang, X., Liu, Z., et al. "Multi-agent energy management optimization for integrated energy systems under the energy and carbon co-trading market." Applied Energy, 324: 1–16, 2022.

[19] Li, F., Qin, J., Zheng, W. X. "Distributed Q-learning-based online optimization algorithm for unit commitment and dispatch in smart grid." IEEE Transactions on Systems, Man, and Cybernetics, 2019: 1–11.

[20] Liu, W., Zhuang, P., Liang, H., et al. "Distributed economic dispatch in microgrids based on cooperative reinforcement learning." IEEE Transactions on Neural Networks, 29(6): 2192–2203, 2018.

[21] Dai, P., Yu, W., Wen, G., et al. "Distributed reinforcement learning algorithm for dynamic economic dispatch with unknown generation cost functions." IEEE Transactions on Industrial Informatics, 16(4): 2258–2267, 2020.

[22] Li, D., Yu, L., Li, N., et al. "Virtual-action-based coordinated reinforcement learning for distributed economic dispatch." IEEE Transactions on Power Systems, 36(6): 5143–5152, 2021.