energies

MDPI

Review

A Review of Smart Grid Evolution and Reinforcement Learning: Applications, Challenges and Future Directions

Na Xu [1], Zhuo Tang [1], Chenyi Si [1,*], Jinshan Bian [2] and Chaoxu Mu [1]

[1] School of Electrical and Information Engineering, Tianjin University, No. 92, Weijin Road, Nankai District, Tianjin 300072, China; naxu@tju.edu.cn (N.X.); tangzhuo@tju.edu.cn (Z.T.)
[2] School of Artificial Intelligence, Anhui University, Qingyuan Campus, 111 Jiulong Road, Hefei 230093, China; js_bian0827@stu.ahu.edu.cn
* Correspondence: sichenyi2021@163.com

Abstract: In the face of the rapid development of smart grid technologies, it is increasingly difficult for traditional power system management methods to support the increasingly complex operation of modern power grids. This study systematically reviews new challenges and research trends in the field of smart grid optimization, focusing on key issues such as power flow optimization, load scheduling, and reactive power compensation. By analyzing the application of reinforcement learning in the smart grid, the impact of distributed new energy's high penetration on the stability of the system is thoroughly discussed, and the advantages and disadvantages of the existing control strategies are systematically reviewed. This study compares the applicability, advantages, and limitations of different reinforcement learning algorithms in practical scenarios, and reveals core challenges such as state space complexity, learning stability, and computational efficiency. On this basis, a multi-agent cooperation optimization direction based on the two-layer reinforcement learning framework is proposed to improve the dynamic coordination ability of the power grid. This study provides a theoretical reference for smart grid optimization through multi-dimensional analysis and research, advancing the application of deep reinforcement learning technology in this field.

Keywords: smart grid; reinforcement learning; optimal power dispatch; reactive power management; distributed control

Academic Editor: Seyed Mahdi Miraftabzadeh

Received: 25 February 2025
Revised: 27 March 2025
Accepted: 1 April 2025
Published: 5 April 2025

Citation: Xu, N.; Tang, Z.; Si, C.; Bian, J.; Mu, C. A Review of Smart Grid Evolution and Reinforcement Learning: Applications, Challenges and Future Directions. *Energies* **2025**, *18*, 1837. https://doi.org/10.3390/en18071837

1. Introduction

The development of a smart grid comprises the deep integration of a traditional power grid and modern information technology, which has been transformed from a traditional power grid with one-way energy transmission to a smart grid with a two-way information interaction, and then to a modern smart grid with a deep integration of distributed resources [1,2]. As the size of a power system expands and the mode of operation diversifies, the complexity and uncertainty of the power grid increases significantly. In particular, the operation of a distribution network is no longer a conventional unidirectional power flow, but becomes a bidirectional interacting system. In this mode, the grid needs to deal with a large number of uncertain factors, including new energy generation fluctuations, load fluctuations, and external disturbances [3]. In addition, the scheduling problem of a power grid becomes more complex with the increasing demands of dynamic loads and multi-objective optimization. How to balance multiple optimization objectives such as voltage stability, power loss, economy, and environmental benefits has become a major challenge in modern grid operations [4,5]. Fault management and voltage stability control are the core issues that need to be addressed to ensure the safe operation of a smart grid. The randomness caused by the high proportion of renewable energy access significantly enhances the uncertainty of grid operation and makes the system more challenging in stability control. The power flow distribution becomes more complicated and the voltage fluctuation becomes more serious [6,7]. Especially in the case of sudden failure, how to realize the rapid voltage recovery becomes an important index to measure the robustness of the system [8,9].

The widespread access of distributed new energy and intelligent power electronic devices also promotes the smart grid to be gradually developed in the direction of a high degree of autonomy, strong interactions and deep intelligence [10]. Although these technologies improve the flexibility and control ability of power grid operation, they also complicate the system structure, making the traditional centralized and distributed control methods face problems such as slow response and localization of optimization targets when dealing with dynamic disturbance [11,12]. In addition, in the scenario of sudden failure, traditional control means are usually static rules and preset strategies, which lack the adaptive optimization ability for complex scenes. Therefore, more intelligent and efficient voltage regulation means are needed to prevent local voltage instability from expanding into large-scale chain failures [13,14]. Active power scheduling and voltage stability control problems involve cooperative optimization decisions for multiple types of devices. Different devices operate with different dynamic response times, resulting in the typical multi-time scale characteristics of smart grid control problems [15–17].

Due to the volatility of renewable energy sources and dynamic changes in load, traditional control methods of power grids have not been able to meet the demand for real-time responses to emergencies and demand fluctuations. Therefore, the control system of a smart distribution grid needs to be capable of fast sensing, accurate prediction, and dynamic adjustment. Specifically, the grid needs to be responsive in real time, able to make quick adjustments to short-term fluctuations and maximize renewable energy utilization through optimized dispatch while ensuring stable grid operation [18,19]. This requires the smart grid to have a high degree of adaptability and fast decision-making capabilities, which can effectively handle the uncertainty caused by new energy fluctuations and load dynamic changes. The main contributions of this study are threefold:

1. This work first reviews the development history of smart grids. By analyzing the evolution of smart grids and related advanced technologies in detail, it expounds the core issues and challenges faced in the process of smart grid optimal scheduling, providing background knowledge in the field of smart grid optimal scheduling and laying the foundation for subsequent discussions.
2. This study reviews the state of the art in deep reinforcement learning for smart grids and analyzes the strengths and weaknesses of existing approaches. In particular, there are challenges in cooperative control of multi-agent systems, convergence of algorithms, and stability. Then, future research directions are discussed in-depth, and key open problems and potential research areas are proposed.
3. A two-layer reinforcement learning optimization framework is introduced to achieve efficient optimization control for smart grids. The upper layer agents are responsible for the coordination and regulation of the global grid, and the lower layer agents perform specific device optimization to enable cooperative optimization of multiple devices. This scheme provides ideas for future research directions.

The remainder of this work is organized as follows: Section 2 presents the history of smart grid development and the current status of research. Section 3 reviews the application of reinforcement learning to the smart grid and presents the challenges of a smart grid. Section 4 proposes a two-level reinforcement learning optimization framework.

Section 5 discusses the emerging trends and challenges in smart grid research. Finally, the work concludes in Section 6.

2. Evolution of Smart Grid Technologies

The traditional power grid is characterized by centralized power generation and one-way energy transmission. Although it has made great achievements in power supply coverage and power security, its scheduling mode lacks flexibility, and its support ability for new energy access and complex load management is weak. In the 1990s, with the rapid development of Information and Communication Technology (ICT), the concept of automated power grids began to take shape. Through the introduction of remote monitoring, data acquisition and control systems, the traditional power grid has achieved basic automatic operation [20–22].

In the 21st century, the rapid development of renewable energy and the increasing demand for distributed energy access have promoted the transformation of the power grid to intelligence. The core feature of a smart grid is two-way information and energy flow, using advanced sensors and communication networks to realize the global interconnection of the power generation side, transmission and distribution network, and end users. Since the 21st century, with the large-scale application of distributed energy (such as photovoltaic power generation, wind power generation), the concept of a microgrid has emerged [23,24]. Microgrids can operate independently in island mode or connect to the main grid to improve regional energy utilization efficiency with distributed generation, energy storage and load shedding at the core. Since 2010, with the rapid development of big data, artificial intelligence and Internet of Things technologies, smart grids have entered the stage of full deployment, and the collaborative optimization of microgrids and smart grids has become a research hotspot. Modern smart grids emphasize the model of distributed autonomy combined with centralized collaboration, through the integration of multiple energy forms and deep intelligent technology, to meet the needs of a high proportion of new energy access and "carbon neutral" goals [25,26].

The smart grid is developing towards more intelligent, distributed and low-carbon environments. Multi-energy complementary systems have become a focus of research, through the integration of power, heat, gas and other energy forms, to achieve a full range of collaborative optimization [27]. At the same time, the deep integration of artificial intelligence, big data analysis and digital twin technology improves the real-time monitoring and prediction ability of systems. On the user side, the introduction of flexible electricity pricing and a demand response mechanism enhances the enthusiasm of users to participate in grid optimization, which lays the foundation for building a more flexible and efficient energy system [28–30].

Some traditional distribution networks have problems, such as low power supply reliability, long power outage time and unstable voltage, especially in areas with a weak power supply, such as old residential areas and urban villages. The digital and intelligent construction of an active distribution network is still in its infancy, and the two-way energy information interaction between users and the smart grid is insufficient, making it difficult to meet the needs of the grid for data sharing and flexible control of a high proportion of distributed energy access [31,32]. Distributed new energy is developing rapidly, but it is difficult for the unidirectional passive network form and single-subject power supply mode of the traditional distribution network to meet the current and future needs of large-scale distributed new energy grid connection, resulting in problems regarding distributed PV network connections in some areas [33]. The distribution network will have a higher new energy acceptance capacity and consumption efficiency, and it will realize the plug and play and local consumption of distributed new energy by optimizing the network structure, upgrading equipment, and innovating the scheduling operation mechanism. The distribution network has changed from the traditional unidirectional radiant network to the bidirectional interactive active network, realizing the flexible access and interaction of each link with regard to the source, grid, load and storage. Moreover, it is gradually becoming an intelligent platform for two-way energy exchange and information interactions with users and distributed energy sources [34,35].

A smart grid is a large-scale power system that integrates modern information technology, communication technology, and power technology. It is highly efficient, flexible, sustainable, and reliable, as shown in Figure 1. It is able to intelligently manage the generation, transmission, distribution and consumption processes, including access to multiple energy forms, such as conventional and renewable energy sources. Big data, cloud computing, artificial intelligence, Internet of Things and other technologies will be widely used in the smart grid to achieve comprehensive perception, real-time monitoring and intelligent control of the distribution network. Through smart meters, smart switches, smart sensors and other devices, real-time collection and analysis of user electricity information are realized, and personalized power services are provided for users [36,37]. Through the collaborative relationship between the distribution network and the microgrid, they jointly participate in the operation and regulation of the grid. The energy storage system can store excess power when there is too much distributed new energy generation, and it can release excess power when the power consumption is at its peak or if new energy generation is insufficient, so as to play the role of peak filling and smoothing power fluctuations. The energy storage system complements distributed new energy and electric vehicle charging facilities to improve the stability and reliability of a smart grid [38].

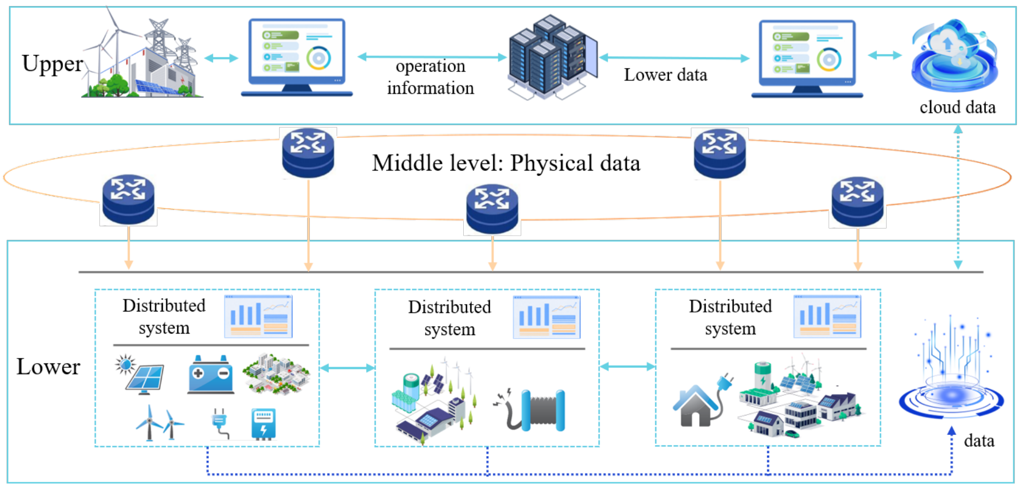

Figure 1. Smart grid system model.

As one of the key technologies in a smart grid, a microgrid plays an increasingly important role in ensuring power reliability, improving energy efficiency and promoting the integration of renewable energy. The microgrid can operate independently and interact with the main grid in both ways to flexibly dispatch distributed energy and energy storage devices, which greatly improves the resilience, flexibility and sustainability of the power system. The research and application of microgrid technologies is growing rapidly and has become an important component of the power system reform and energy transition [39]. It is composed of distributed energy photovoltaic, wind power, battery energy storage, load and control equipment, and can operate while connected to the main grid or independently in the island mode. Microgrids are characterized by flexibility and autonomy, and they are often used to improve energy efficiency and power supply reliability in local areas. The microgrid under the new power system is suitable for a variety of application scenarios [40,41]. In a smart grid, the proportion of distributed renewable energy is increasing year by year. As the integration platform of distributed energy, microgrids can effectively solve the problem of renewable energy absorption. A microgrid can balance the volatility and intermittency of renewable energy through the built-in energy storage system and intelligent scheduling function, reduce the dependence on the main grid, and improve the utilization rate of renewable energy. In some remote areas, isolated islands or special environment power supply and emergency support, such as military bases, hospitals and other areas, microgrids can provide independent power supply to avoid power interruptions caused by power grid failure or unstable remote transmission networks. In addition, a microgrid can quickly switch to the islanding operation mode after natural disasters, such as earthquakes and typhoons, so as to provide emergency power supply for disaster areas and improve the emergency response ability of the power system. At the same time, with the advancement of the power market reform, microgrids can optimize the allocation and dispatch of power resources by participating in the power market. In the electricity market, a microgrid can adjust the generation and consumption strategy according to the change of the price of electricity, and even feed back the excess power to the grid by interacting with the main grid to obtain economic benefits. Microgrids can also participate in demand-side management to balance the supply and demand by adjusting the load demand and optimizing power consumption [42–44].

Under the requirements of the dual development of a smart grid and the effective application of artificial intelligence technology in a power system, microgrid technology faces multiple key issues and challenges. Stability and control techniques for microgrids face the problem of how to maintain the stability and safety of a power system while operating on an island. Distributed generation equipment in a microgrid is prone to voltage frequency instabilities due to its large power generation fluctuations. Therefore, microgrids need to be equipped with efficient control and stabilization techniques to ensure that parameters such as voltage and frequency are within safe ranges. Current common technical means include advanced power electronics, intelligent control algorithms, and fast response energy storage systems [45–47]. Energy storage system integration and optimization play a key role in microgrids, which can balance the volatility of renewable energy and improve the flexibility and stability of the grid. However, the problems of cost, efficiency, and lifetime of energy storage systems are still one of the challenges faced by microgrid technology. How to optimize the configuration of energy storage devices, improve the service life of energy storage systems, and reduce their operating costs are the key directions of future microgrid technology research [48]. The electricity market and policies support the economic benefits of microgrids. As a representative of distributed energy, microgrids face the problems of imperfect power market rules and opaque price mechanisms. In addition, the national policy support for microgrids is uneven, which also affects their promotion and application. Therefore, it is necessary to develop a more perfect electricity market mechanism and policy framework to support the development and application of microgrid technology [49]. As the smartness of microgrids gradually has increased, their network security issues have also attracted increasing attention. Various communication and control devices in a microgrid system are vulnerable to external attacks or failures that may lead to power supply disruption or system collapse. Therefore, the security protection and network security measures of the microgrid are particularly important, and the protection design of the power system needs to be strengthened to ensure the anti-attack ability of a system [50,51].

Importantly, smart grids achieve more efficient power management by optimizing the access, operation and scheduling of microgrids. Microgrids can be used as the core node of distributed energy management in smart grids, providing auxiliary services such as frequency regulation and voltage support to main grid [52]. Smart grids provide wide-area resource optimization and scheduling support for microgrids. Smart grid communication and control technologies such as distributed energy management systems, real-time monitoring and optimization algorithms are the basis for the efficient operation of microgrids. Therefore, when studying and describing smart grids, microgrids are usually treated as an independent subsystem, especially in the context of distributed scheduling and collaborative optimization [53–55]. The improvement of the optimal scheduling approach for microgrids effectively addresses the source–load bilateral uncertainty problem in hybrid renewable energy systems and provides dynamic scheduling strategies to support the design of new grid architectures in the presence of a high proportion of renewable energy. Through the deep integration of load and power generation forecasting technology HRES, the supply and demand fluctuation curve can be accurately matched, so as to enable the flexible adjustment ability of smart grids and the collaborative optimization of smart infrastructure. This technology path not only reduces system operation costs but also drives the deep coupling of smart energy management systems and edge-side distributed control, which lays a key technical foundation for the construction of flexible and digital future smart grids [56].

Being the digital upgrade to traditional power grids, smart grids are promoting the transformation of power systems in a efficient, reliable and sustainable direction through the deep integration of the Internet of Things, artificial intelligence, big data and energy storage technology. Future improvements in automation will reduce operating costs, improve system operation efficiency, and enhance the ability of smart grids/microgrids to cope with emergencies. Multi-energy collaboration and system integration microgrids will gradually realize the collaborative development of various energy forms and form a diversified and collaborative energy system. Different forms of energy, such as photovoltaic, wind, hydrogen and geothermal energy, will jointly participate in the operation of microgrids, and new loads such as energy storage, smart home and electric vehicle charging piles will also become part of microgrids to achieve the optimal allocation of resources and efficient use of energy. Microgrids will not only be an isolated energy system but will be deeply integrated with the main grid to form an interactive distributed energy network. Through the synergy of smart scheduling and market mechanisms, smart grids and microgrids will be able to achieve interconnection with the main grid, improving the security and reliability of power systems while improving the overall energy utilization efficiency.

3. Application of Reinforcement Learning in Smart Grids

3.1. Key Research Directions in Smart Grids

Smart grids can realize real-time scheduling and optimization of power systems, optimize power generation, transmission, distribution and consumption through the real-time analysis of power grid status, and improve the overall operating efficiency. The optimal scheduling problem of smart grids is the key to achieve efficient operation and stable power supply [57,58]. Optimal scheduling requires addressing the challenges of dynamics, multiple objectives, and multiple constraints simultaneously. This dynamism is derived from the fluctuation of new energy outputs and the real-time changes of loads. The multi-objective considers the comprehensive requirements of the economy, robustness and environmental protection. Multiple constraints are reflected in the collaborative optimization of distributed resources, the coordination of time scales and the security guarantee of power grids [59,60]. Traditional optimization methods, such as linear programming (LP), nonlinear programming (NLP), and mixed-integer linear programming (MILP), perform well in small-scale problems. However, they often face problems such as high computational complexity and lack of real-time performance when dealing with high-dimensional nonlinear problems of modern smart grids.

International studies have proposed a variety of solutions to the optimal scheduling problem in smart grids, focusing on the application of artificial intelligence and distributed optimization methods. China is accelerating UHV and intelligent distribution grids with the "dual carbon" goal as its main focus. Europe and the United States are strengthening renewable energy integration and grid resilience through policies and investments. Japan and India are focusing on user-side management and reliability improvements. Despite facing challenges, such as fluctuations of a high proportion of renewable energy connected to the grid, insufficient cross-regional coordination, high investment costs, and user privacy concerns, future smart grids will develop in the direction of integrating the energy Internet, blockchain decentralized transactions, and edge computing autonomous decision making, becoming the core infrastructure to support carbon neutrality goals and zero-carbon power systems. The U.S. distributed energy resource management system enables the efficient consumption of new energy by optimizing the real-time control of distributed energy and microgrids. Europe has extensive experience in the demand response and electricity market optimization, and it has greatly improved the efficiency of grid operations through real-time electricity pricing and user-side load regulation mechanisms. In addition, AI techniques have been widely used in research in European and American countries, including the application of deep learning in load forecasting, as well as in the exploration of reinforcement learning in dynamic scheduling problems. Japan and South Korea have made significant progress in the power sector through big data analytics and smart sensor technologies, improving the transparency and predictive capabilities of grid operations [61–63].

By implementing the "Ubiquitous Power Internet of Things" strategy, whole-chain data from the power generation side to the user side are integrated in an intelligent platform, which provides strong data support for optimal dispatching. Several works have introduced modern smart algorithms such as reinforcement learning and multi-agent cooperative optimization to solve optimal scheduling problems under complex constraints and dynamic environments. Meanwhile, industrial parks and urban demonstration zones are promoting multi-energy collaborative optimization scheduling based on microgrids, which provide new ideas for the efficient utilization of a high proportion of new energy [64–66].

The scheduling module designed based on deep reinforcement learning and a multi-agent approach is shown in Figure 2, where all agents are integrated in the distributed collaborative optimization scheduling center of a smart grid. Agents include the main grid day-ahead scheduling agent, the main grid active power scheduling agent, the reactive power compensation agent, the microgrid group cooperative optimization agent, the microgrid multi-energy cooperative agent, and the real-time dynamic voltage adjustment agent. The dispatch center collects real-time grid operation data, and different types of agents provide optimization strategies based on the current grid operation requirements to assist with the efficient and stable operation of the grid.

In recent years, most of the research has focused on power market mechanisms, user-side optimization, distributed control and UHV transmission, large-scale centralized new energy management, and industrial demonstration applications. In the future, with the further development of artificial intelligence, Internet of Things and digital twin technology, the optimal scheduling problem of smart grids will gradually move towards the direction of global coordination, real-time optimization and multi-objective robustness, providing strong support for the global energy transition and "carbon neutrality" goals [67–69].

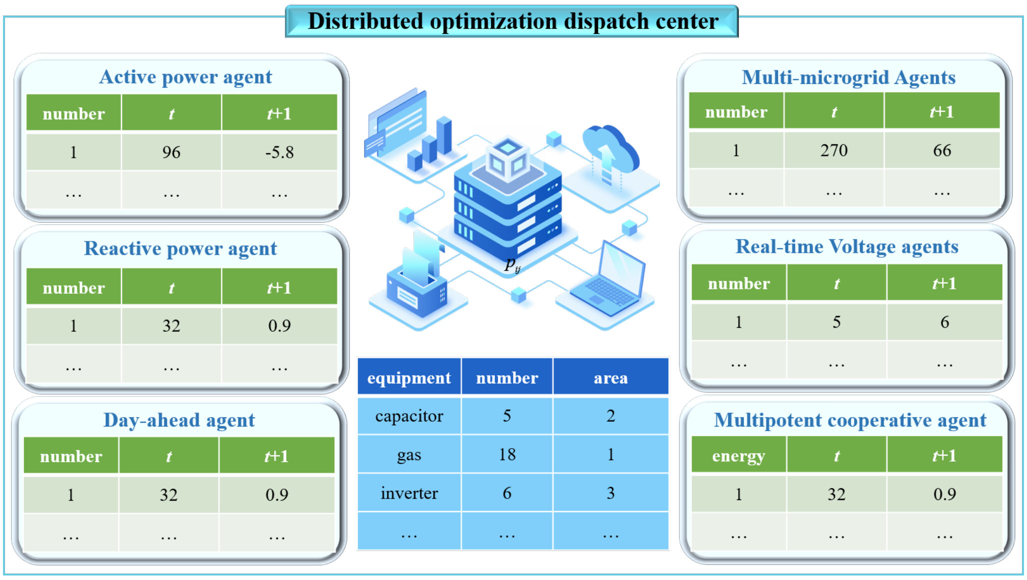

Figure 2. Distributed collaborative optimization dispatch center for smart grid.

3.2. Evolution of Reinforcement Learning in Smart Grid

Reinforcement learning has been gradually introduced into power systems to solve dynamic optimization problems since it was proposed in the 1980s. Its application in power systems has progressed through the development process from a theoretical exploration to a combined application with practical application scenarios [70,71]. In the early development phase, the application of reinforcement learning to power grids was mainly focused on a theoretical validation, with classical Q-Learning methods being applied to simple power system problems. The goal was to explore the potential of reinforcement learning in the optimal operation of power grids, with typical applications such as pricing optimization and load forecasting in electricity markets. Limited by the algorithmic ability of tabular Q-Learning, reinforcement learning faces problems such as the difficulty of state space expansion and the lack of real-time performance when dealing with complex power systems. This area of research lays a theoretical foundation for the application of reinforcement learning in power system [72–74].

Multi-agent reinforcement learning has emerged as a research hotspot to cope with the computational complexity caused by the scale expansion of dynamical systems. Single-agent reinforcement learning is beginning to be applied to deal with optimization tasks, such as reactive power compensation regulation and load management, to achieve improvements in local voltage stability and operating efficiency by optimizing the operation strategy of a single device [75,76]. At the same time, multi-agent reinforcement learning is introduced into the distributed environment to solve the problem of collaboration between multiple nodes. Distributed optimization among nodes has been achieved by designing interaction mechanisms between agents, such as cooperative control of distributed power generation systems and operational optimization of microgrid islanding patterns. At this stage, reinforcement learning research initially showed its potential in distributed power systems [77–79].

The integration of agent, environment, and training algorithm results in a complete training module is shown in Figure 3. In this module, the main program is responsible for coordinating the work of each component, including environment initialization, setting up agents and training algorithms, and managing the training process. The environment module outputs state data and associated grid parameters to the agent, who makes decisions based on the policy network module. The training algorithm guides the agent to optimize the policy through the reward signal. Through continuous iterative training, the reward value of the dynamical agent converges to the optimal value. After saving the agent model, it is embedded in the distributed cooperative optimization dispatch center of a smart grid to assist the smart grid to make stable and efficient operation decisions and optimize its operation.

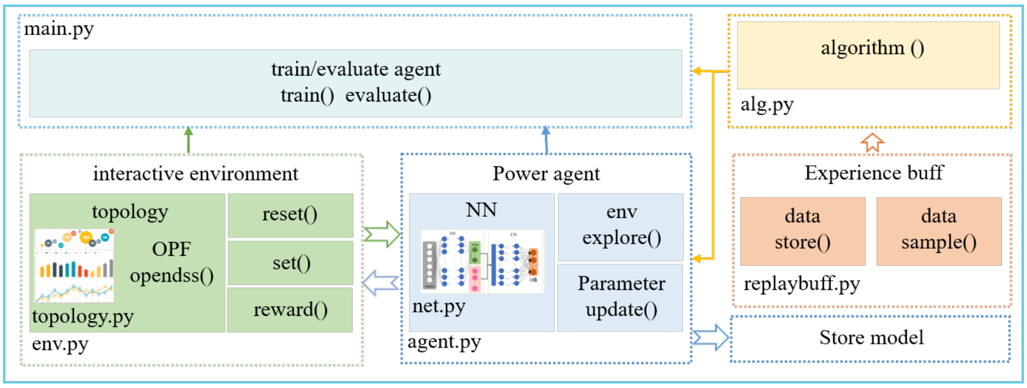

Figure 3. Agent–smart grid environment interaction process.

The introduction of the Deep Q-network (DQN) and its variants, Deep Deterministic Policy Gradient (DDPG) and Proximal Policy Optimization (PPO), provides a powerful solution to problems with high-dimensional state and action spaces [80]. Deep reinforcement learning is beginning to show advantages in more complex power system scenarios, such as dynamic optimal scheduling [81], distributed control [82], and voltage stability control [83]. The use of deep reinforcement learning has led to breakthroughs in distributed energy cooperative optimization and the demand response. In addition, reinforcement learning is further combined with methods, such as the Markov decision process [84] and distributed optimization [85], further improving the robustness and applicability of the algorithm and providing a new solution for the actual power system operation.

The goal of the agent is to choose the appropriate policy in each training round to maximize the final reward.

$$V(s) = \mathbb{E}[r_t | s_t = s], \tag{1}$$

where $V(s)$ is the state value function. r_t is the reward value in state s_t.

$$Q(s,a) = \mathbb{E}[r_t | s_t = s, a_t = a], \tag{2}$$

where $Q(s,a)$ is the state/action value function. a_t is the action in state s_t. In the optimal scheduling problem of a smart grid, the Q-function can help the agent to evaluate the possible stable states and optimization effects of the grid after adjusting the control parameters of a cell and device at a certain time. The optimal policy is the optimal action chosen by the agent in different states to maximize the cumulative reward. This policy can effectively solve multi-objective optimization problems in smart grids. In optimal grid control, the optimal policy involves cooperative scheduling of multiple devices to maximize the stability, efficiency, and economy of the grid. Value iteration is a method based on the state value function that solves the optimal policy by iteratively updating the value of the state. The core idea of value iteration is to incrementally estimate the value of each state and then derive the optimal policy through the value function. The update process of the value iteration can be formulated as follows:

$$V_{t+1}(s) = \max_a \left(r(s,a) + \gamma \sum_{s'} P(s'|s,a) V_t(s') \right), \tag{3}$$

where $V_{t+1}(s)$ is the value of state s at iteration t. γ is the weight used by the discount factor to determine the future reward. $r(s,a)$ is the immediate reward after state s takes policy a. $P(s'|s,a)V_t(s')$ is the state transition probability, which represents the probability of transitioning to state s_{t+1} after action a_t is performed in state s_t. The main difference between policy iteration and value iteration is that it updates the policy directly instead of relying on the update of the state/value function. Smart grids have complex topologies and huge parameters, so choosing a policy iteration based approach is more helpful to improve the computational efficiency. Policy iteration consists of two main steps: policy evaluation and policy improvement. The update formulation for a policy evaluation is as follows:

$$V^{\pi}(s) = r(s,\pi(s)) + \gamma \sum P(s'|s,\pi(s))V^{\pi}(s'), \tag{4}$$

where $V^{\pi}(s)$ is the expected reward obtained after taking an action following policy $\pi(s)$ in state s. $\gamma \sum P(s'|s,\pi(s))V^{\pi}(s')$ is the value-weighted sum of all possible states s'. $V^{\pi}(s')$ is the value function of the next state. $\pi(s)$ is a deterministic policy, and $\pi(s)$ is replaced by $\sum_a \pi(a|s)$ if the policy is a random policy. Policy improvements improve the current policy by selecting the optimal action for each state. Policy improvements require the construction of an action/value function $Q^{\pi}(s)$, based on which a better policy is selected.

$$Q^{\pi}(s,a) = r(s,a) + \gamma \sum P(s'|s,\pi(s))V^{\pi}(s'), \tag{5}$$

where $Q^{\pi}(s,a)$ is the expected future reward after performing action a in state s. $Q^{\pi}(s,a)$ can calculate the potential payoff of different actions. The relation between $V^{\pi}(s)$ and $Q^{\pi}(s,a)$ is

$$V^{\pi}(s) = Q^{\pi}(s,\pi(s)). \tag{6}$$

The new policy $\pi'(s)$ can select more optimal actions, such that $V^{\pi'}(s)$ is at least no worse than $V^{\pi}(s)$. The updated formula is as follows:

$$\pi'(s) = arg\max_a Q^{\pi}(s,a) = arg\max_a (r(s,a) + \gamma \sum P(s'|s,a)V^{\pi}(s')), \tag{7}$$

where $\pi(s')$ is the new policy that maximizes $Q^{\pi}(s,a)$. Policy iteration continuously optimizes a policy by alternately executing policy evaluation and policy improvement steps until the policy converges to the optimal policy. Value iterations are straightforward to compute, and an optimal policy is directly estimated by the state/value function. Moreover, the convergence rate is fast and suitable for the case of small state space. The policy iteration can efficiently improve each policy and obtain the exact optimal policy. However, it requires performing a full policy evaluation for each policy, which has high computational complexity, and the update speed can be very slow in large-scale systems. Depending on the operational scenario and optimization objective, a large number of papers have been produced on the effective application of reinforcement learning in smart grids.

The research of reinforcement learning is gradually transitioning from the theoretical simulation environment to more complex and practical application scenarios [86,87]. The dynamic optimization problem of complex power systems under the scenario of a high proportion of new energy access has become an important application direction of reinforcement learning. Some studies have extensively employed reinforcement learning techniques in smart distribution grids and microgrids to enhance the interpretability and robustness of models by combining physical laws with reinforcement learning algorithms. In addition, the proposed multi-level optimization framework further promotes the collaborative application of single-agent and multi-agent methods in distributed systems [88–90].

Moreover, several papers have used biomass as a renewable energy source in conjunction with diesel power plants to solve the optimal control problem for isolated microgrids. By modeling the isolated microgrid as a Markov decision process, reinforcement learning is used to minimize the total system cost, which effectively addresses the resource shortage problem [91]. In the real-time regulation of distributed power grids, reinforcement learning algorithms have been gradually deployed in dynamic voltage controllers and load management systems to cope with sudden faults and complex constrained environments. At the same time, the technical integration of smart edge computing and hybrid reinforcement learning makes reinforcement learning more efficient and practical for applications.

As an important distributed energy management unit, microgrids can improve the consumption rate of renewable energy and enhance the resilience of a grid. However, the increasing size and number of microgrids brings entirely new challenges to the optimal dispatch of power systems, especially in the case of multi-microgrid cooperation coupled to the main grid. Traditional optimization methods, such as linear programming and heuristic algorithms, suffer from high computational complexity and poor real-time performance when dealing with complex dynamic environments [92,93]. Most existing studies focus on single-level optimization, lacking systematic studies and multi-objective cooperative optimization algorithms [94]. Moreover, due to the dynamic and nonlinear nature of power grids, traditional optimization algorithms are significantly limited in dealing with complex environmental changes. Reinforcement learning techniques have become an effective means to solve complex multi-layer optimization problems based on their ability to automatically adapt to environmental changes and self-learn [95,96]. However, most of the current research focuses on the internal cooperation of microgrids or the overall optimization scheme of microgrid clusters, and less on the cooperative optimization between the global system and microgrid clusters. The computational efficiency and convergence rate of reinforcement learning algorithms face challenges as the size of the nodes of the distribution network increases. Traditional optimization methods cannot effectively handle the hierarchical decision-making problem in smart grids. Multi-agent reinforcement learning provides an effective technical means to solve this problem by designing a collaborative mechanism among agents.

In summary, the application of reinforcement learning to smart grids has gradually evolved from a theoretical exploration to practical deployment, and its technical direction has expanded from simple single-agent models to complex multi-agent cooperative optimization, as well as from static problem solving to dynamic optimization and hierarchical control.

4. Two-Layer Reinforcement Learning Architecture for Distributed Smart Grid Management

Aiming at the reactive power optimization problem of smart grids, a two-layer reinforcement learning optimization framework is proposed. Consider a smart distribution network consisting of strip buses with a root node connected to a substation bus. The set of lines in the distribution network is $\mathcal{N} = \{1, \ldots, i, \ldots, N\}$. The square of the magnitude of the voltage amplitude of each node is denoted by v_i, and the voltage phase angle of the node is denoted by θ_i. A s mart distribution network includes loads, distributed power sources, capacitor banks, and reactive power compensation devices. The core task of the system is to deliver electrical energy from the transmission network to the end users while ensuring voltage stability and minimizing line losses. The reactive power compensation unit optimizes the power factor and reduces the reactive power loss by rapidly injecting or absorbing the reactive power load demand. Through the coordinated control of an on-load tap changer (OLTC), capacitor bank and voltage regulator, the optimal operation of the system can achieve voltage stability and power loss minimization in the distribution network. Specifically, voltage regulation in distribution networks requires precise coordinated control among multiple devices to ensure that the voltage at each node is within a reasonable range, thus maintaining the stability of the grid.

Smart distribution networks are modeled as radially distributed voltage control systems, where reactive power compensators are embedded in the control loops of each controllable node. The topology of a smart distribution network can be abstracted as a directed graph, where nodes represent individual distribution devices and loads, and edges represent power flows between devices. The root node is the connection point between the substation and the distribution network, and each device and load is connected through the distribution line, forming a radial network structure. Based on this topology, power flow calculations can be performed, optimal scheduling can be carried out, and control decisions can be made. In this case, voltage control depends not only on the regulation ability of individual devices but also on the global optimum achieved by the joint regulation of multiple devices.

Each node of a smart distribution network satisfies the following balance constraints:

$$p_j = -p_{ij} + R_{ij}I_{ij} + \sum_{j,k\in\zeta} p_{jk}, \tag{8}$$

$$q_j = -q_{ij} + X_{ij}I_{ij} + \sum_{j,k\in\zeta} q_{jk}, \tag{9}$$

$$I_{ij} = \left(p_{ij}^2 + q_{ij}^2\right), \tag{10}$$

where p_j and q_j are the active and reactive powers flowing into the jth node, respectively. p_{ij} and q_{ij} are the active and reactive powers flowing into node j from parent node i, respectively. R_{ij} and X_{ij} are the line impedance and reactance flowing through node j, respectively. l_{ij} is the square of the line current amplitude. $R_{ij}I_{ij}$ and $X_{ij}I_{ij}$ are the line losses. The voltage amplitude of each node is satisfied as follows:

$$v_j = v_i - 2(R_{ij}p_{ij} + X_{ij}q_{ij}) + (R_{ij}^2 + X_{ij}^2)/l_{ij}. \tag{11}$$

Each node is able to satisfy the voltage upper and lower limits $0.95 \leq v_j \leq 1.05$ to ensure that the voltage operates within a safe range.

Different devices have different response times to voltage stabilization in the face of load fluctuations and the stochastic nature of renewable energy. While reactive power compensation units and dynamic capacitors can respond to voltage fluctuations at the millisecond level, the tuning of transformer taps and capacitance states is mainly focused on optimizing the cooperative operation of a device, which typically corresponds to long timescales and is subject to its physical limitations. Scheduling in this phase requires taking into account both the load prediction and the output characteristics of the distributed power supply, as well as ensuring that equipment can quickly adapt to load fluctuations and voltage changes without compromising its lifetime.

A hierarchical reinforcement learning (HRL) framework is proposed to decompose the voltage control problem into long-term planning and target assignment at the upper level, as well as short-term planning and execution at the lower level, providing a control framework for effectively solving the control problem across time scales. This design enables the stepwise refinement of the scheduling policy, which leads to the efficient optimization of a power grid at different time scales.

The two-layer reinforcement learning framework is shown in Figure 4. The upper-level objective focuses on the optimization of power flow across the grid and the economy of equipment scheduling, focusing on the long-term scheduling of equipment and the minimization of operating costs through economical scheduling in the voltage control process. In this process, historical data, load prediction, and output features are used as observed quantities of agents to provide raw data for generating policies. The upper agent provides the initial adjustment amount for the lower schedule, whose optimization results in turn influence the execution effect of the long-term plan.

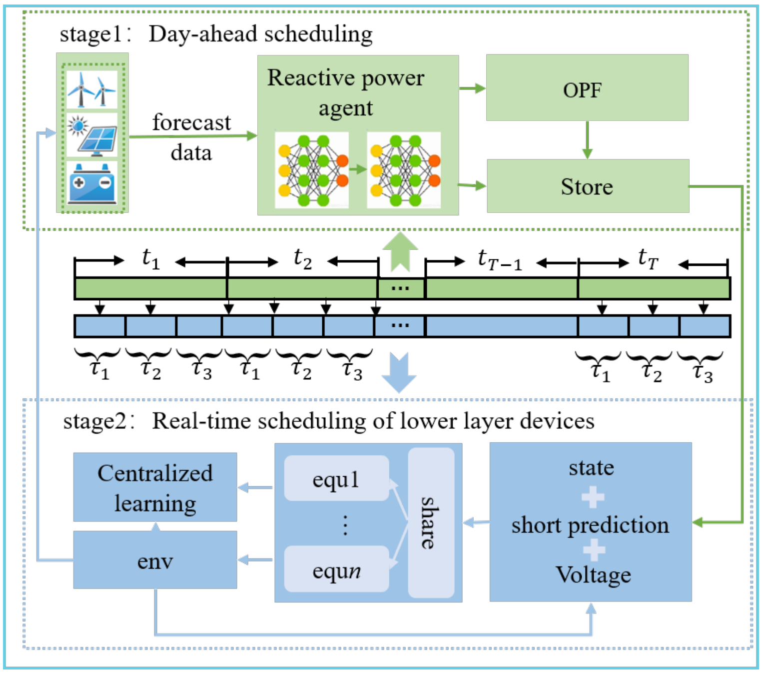

Figure 4. A two-layer reinforcement learning framework for a smart grid.

The upper-layer agents are responsible for global voltage optimization and the location and identification of fault regions. By monitoring the state of the whole network, the target regions and key nodes that need to be optimized are selected, and the corresponding voltage regulation task is assigned to lower-layer agents. Lower-layer agents are responsible for fine-tuning the scheduling of specific devices and responding quickly in the event of faults or emergencies.

The upper level has a time scale of 1 hour, and its main tasks are global voltage optimization, target assignment, and fault region identification. The agent optimization objective is to minimize voltage bias, line losses, and device operating costs. The specific objective function is as follows:

$$J_u = \min \sum_{i}^{n} (v_{vari} + p_{loss} + f_c), \tag{12}$$

where v_{vari} is the voltage deviation of each node. p_{loss} is the power loss in the process of power transmission. f_c is the operating cost of the equipment in the process of action execution.

The lower layer mainly handles short-term control tasks with a time scale of 15 min, focusing on fine-tuning and rapid responses at the device level. The main task of the lower layer is to perform operations such as dynamic capacitor switching status, voltage regulation, reactive power compensation, fast recovery and voltage stable regulation under fault conditions, according to the objectives of the upper layer. The objective function for the optimization of the lower layer is as follows:

$$J_l = \min \sum_{i}^{n} (v_{vari} + p_{loss} + c_{op}), \tag{13}$$

where v_{vari} is the voltage deviation. p_{loss} is the power loss after the device state switch. c_{op} is the cost of operating equipment. This objective function focuses on immediate responses of a device on a grid, ensuring that voltage returns to stability in a short period of time and reducing local line losses and equipment operation costs.

Through the hierarchical reinforcement learning framework, upper agents focus on global voltage optimization, fault area location, and recovery policy formulation, while lower agents perform scheduling tasks for specific devices to ensure voltage stability of the power grid in a short time. The interplay between the upper and lower layers allows the grid to be flexibly dispatched in the presence of load fluctuations, renewable energy uncertainties, and faults, thus maintaining efficient and stable operation. Since the time scale of the upper agent is 1 h, its main task of voltage regulation is to perform global optimization and equipment scheduling with the aim of achieving long-term voltage stability of the grid. However, in the face of an occurrence of a fast fault or voltage fluctuations, the response of the upper agent may be lagged. Therefore, it is necessary to rely on the fine control of the lower agent for a fast dynamic response. Lower-layer devices can perform device-level adjustments on short timescales to quickly respond to voltage deviations and faults in the grid, thus ensuring that the grid can recover stability in time in the event of faults. This hierarchical structure not only solves the control problem across time scales but also effectively improves the response capability and recovery speed of the grid. The proposed framework lays the theoretical foundation for the integration of distributed dispatch centers in smart grids. Depending on the different operation scenarios of a smart grid, selecting an appropriate reinforcement learning method to train an agent can effectively improve the operation.

5. Emerging Trends and Challenges in Smart Grid Research

By analyzing the application of existing reinforcement learning methods to a smart grid, it can be found that there are three core challenges when optimizing power systems:

1. Insufficient handling of safety constraints: Current reinforcement learning frameworks usually adopt the "a posteriori penalty" policy, which simply superimposes a constraint violation penalty term in the reward function. However, this approach leads to a conflict between safety constraints and the exploration space of the agent, which can lead to safety risks such as voltage overruns and line overloads while the agent pursues optimal economy.
2. Dimensional catastrophes due to multi-timescale coupling: Smart grid operations involve a strong coupling between millisecond-level transient control and hourly level scheduling decisions, resulting in a dramatic growth of the action space dimension. In the face of the diverse operational requirements of smart grids, it is urgent to build a more refined hierarchical distributed cooperative control mechanism that ensures efficient cooperation of each grid node on the basis of independent decision making and guarantees global stability and optimization.
3. Insufficient robustness and risk sensitivity: Under the influence of uncertain disturbances and complex constraints, smart grid control algorithms need to be more robust and risk-sensitive. Robustness ensures that the grid can remain stable in uncertain environments, while risk sensitivity requires that the control system can make adaptive decisions under uncertain conditions to minimize potential risks and losses.

Future breakthroughs will follow three main directions:

1. Deep integration of safety reinforcement learning and physical models: Lyapunov function constraints and safety layer embedding architecture are used to construct grid state maps to improve decision interpretability in grid control tasks. An organic

combination of reinforcement learning and physical models of dynamical systems will be implemented based on safety-constrained algorithms.

2. Collaborative optimization across time scales: Through a hierarchical reinforcement learning architecture, a coupled model across time scales is constructed to achieve efficient collaboration of control units in smart grids. Each agent makes independent decisions based on local information and collaborates with other agents to form a global optimal scheme to realize cross-regional collaborative scheduling, which not only protects data privacy but also reduces the redundancy of reserve capacity, so as to ensure the overall security and stability of the power grid system.
3. Risk assessment and adaptive decision making: Risk assessment can be carried out for different scenarios, as well as the flexible adjustment of strategies. By quantifying the probability distribution of extreme events, stochastic optimization and robust optimization are integrated into the adversarial training framework to balance the worst-case and expected performance, and a decentralized robust reinforcement learning algorithm is developed.

In line with the above directions, reinforcement learning and multi-agent optimization methods will become important tools for smart grid safety optimization scheduling. These approaches are able to adaptively adjust policies to cope with various uncertainties in the real-time operation of a power grid by interacting with the environment. In the future, reinforcement learning will play a more critical role in the intelligent transformation of power systems and provide solid technical support to achieve a high proportion of new energy consumption and ensure the efficient operation of smart grids.

6. Conclusions

In this work, the authors perform a thorough review of the development of smart grids and the application of reinforcement learning in smart grids. By analyzing the challenges and opportunities of optimal scheduling and voltage control problems in smart grids, this study proposes a two-layer reinforcement learning framework to lay the foundation for further research. Future research can further support the development of more efficient, stable, and scalable smart grid systems.

Author Contributions: Conceptualization, N.X.; methodology, N.X.; software, C.M.; validation, N.X.; formal analysis, Z.T. and J.B.; investigation, C.S.; resources, C.M.; data curation, N.X.; writing—original draft preparation, N.X.; writing—review and editing, N.X. and C.S.; visualization, Z.T.; supervision, C.M.; project administration, C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Science and Technology Project of the State Grid Corporation of China (5400-202456175A-1-1-ZN).

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Gill, S.; Kockar, I.; Ault, G.W. Dynamic optimal power flow for active distribution networks. *IEEE Trans. Power Syst.* **2014**, *29*, 121–131. [CrossRef]
2. Lin, C.; Wu, W.; Chen, X.; Zheng, W. Decentralized dynamic economic dispatch for integrated transmission and active distribution networks using multi-parametric programming. *IEEE Trans. Smart Grid* **2018** , *9*, 4983–4993. [CrossRef]
3. Sun, H.; Guo, Q.; Qi, J.; Ajjarapu, V.; Bravo, R.; Chow, J.; Li, Z.; Moghe, R.; Nasr-Azadani, E.; Tamrakar, U.; et al. Review of challenges and research opportunities for voltage control in smart grids. *IEEE Trans. Power Syst.* **2019**, *34*, 2790–2801. [CrossRef]

4. Chen, Y.; Pan, F.; Qiu, F.; Xavier, A.S.; Zheng, T.; Marwali, M.; Knueven, B.; Guan, Y.; Luh, P.B.; Wu, L.; et al. Security-constrained unit commitment for electricity market: Modeling, solution methods, and future challenges. *IEEE Trans. Power Syst.* **2023**, *38*, 4668–4681. [CrossRef]
5. Molzahn, D.K.; Dörfler, F.; Sandberg, H.; Low, S.H.; Chakrabarti, S.; Baldick, R.; Lavaei, J. A survey of distributed optimization and control algorithms for electric power systems. *IEEE Trans. Smart Grid* **2017**, *8*, 2941–2962. [CrossRef]
6. Dsouza, A.K.; Thammaiah, A.; Venkatesh, L.K.M. An intelligent management of power flow in the smart grid system using hybrid npo-atla approach. *Artif. Intell. Rev.* **2022**, *55*, 6461–6503. [CrossRef]
7. Zhou, M.; Zhai, J.; Li, G.; Ren, J. Distributed dispatch approach for bulk AC/DC hybrid systems with high wind power penetration. *IEEE Trans. Power Syst.* **2018**, *33*, 3325–3336. [CrossRef]
8. Antoniadou-Plytaria, K.E.; Kouveliotis-Lysikatos, I.N.; Georgilakis, P.S.; Hatziargyriou, N.D. Distributed and decentralized voltage control of smart distribution networks: Models, methods, and future research. *IEEE Trans. Smart Grid* **2017**, *8*, 2999–3008. [CrossRef]
9. Kekatos, V.; Zhang, L.; Giannakis, G.B.; Baldick, R. Voltage regulation algorithms for multiphase power distribution grids. *IEEE Trans. Power Syst.* **2016**, *31*, 3913–3923. [CrossRef]
10. Evangelopoulos, V.A.; Georgilakis, P.S.; Hatziargyriou, N.D. Optimal operation of smart distribution networks: A review of models, methods and future research. *Electr. Power Syst. Res.* **2016**, *140*, 95–106. [CrossRef]
11. Safdarian, F.; Kargarian, A.; Hasan, F. Multiclass learning-aided temporal decomposition and distributed optimization for power systems. *IEEE Trans. Power Syst.* **2021**, *36*, 4941–4952. [CrossRef]
12. Chen, Y.; Zhu, J.; Liu, Y.; Zhang, L.; Zhou, J. Distributed hierarchical deep reinforcement learning for large-scale grid emergency control. *IEEE Trans. Power Syst.* **2024**, *39*, 4446–4458.
13. Mu, C.; Wang, K.; Ni, Z.; Sun, C. Cooperative differential game-based optimal control and its application to power systems. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5169–5179.
14. Gao, Y.; Wang, W.; Yu, N. Consensus multi-agent reinforcement learning for volt-var control in power distribution networks. *IEEE Trans. Smart Grid* **2021**, *12*, 3594–3604. [CrossRef]
15. Naidu, B.R.; Bajpai, P.; Chakraborty, C.; Malakondaiah, M.; Kumar, B.K. Adaptive dynamic voltage support scheme for fault ride-through operation of a microgrid. *IEEE Trans. Sustain. Energy* **2023**, *14*, 974–986. [CrossRef]
16. Hu, D.; Ye, Z.; Gao, Y.; Ye, Z.; Peng, Y.; Yu, N. Multi-agent deep reinforcement learning for voltage control with coordinated active and reactive power optimization. *IEEE Trans. Smart Grid* **2022**, *13*, 4873–4886. [CrossRef]
17. Xu, S.; Xue, Y.; Chang, L. Review of power system support functions for inverter-based distributed energy resources-standards, control algorithms, and trends. *IEEE Open J. Power Electron.* **2021**, *2*, 88–105.
18. Wang, S.; Duan, J.; Shi, D.; Xu, C.; Li, H.; Diao, R.; Wang, Z. A data-driven multi-agent autonomous voltage control framework using deep reinforcement learning. *IEEE Trans. Power Syst.* **2020**, *35*, 4644–4654.
19. Yu, Y.; Ju, P.; Peng, Y.; Lou, B.; Huang, H. Analysis of dynamic voltage fluctuation mechanism in interconnected power grid with stochastic power disturbances. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 38–45.
20. Liu, Y.; Qu, Z.; Xin, H.; Gan, D. Distributed real-time optimal power flow control in smart grid. *IEEE Trans. Power Syst.* **2017**, *32*, 3403–3414.
21. Mohammadi, J.; Hug, G.; Kar, S. Agent-based distributed security constrained optimal power flow. *IEEE Trans. Smart Grid* **2018**, *9*, 1118–1130. [CrossRef]
22. Jain, H.; Mather, B.; Jain, A.K.; Baldwin, S.F. Grid-supportive loads? a new approach to increasing renewable energy in power systems. *IEEE Trans. Smart Grid* **2022**, *13*, 2959–2972.
23. Tazi, K.; Abbou, F.M.; Abdi, F. Multi-agent system for microgrids: Design, optimization and performance. *Artif. Intell. Rev.* **2020**, *53*, 1233–1292. [CrossRef]
24. Mu, C.; Liu, W.; Xu, W. Hierarchically adaptive frequency control for an EV-integrated smart grid with renewable energy. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4254–4263. [CrossRef]
25. Gambuzza, L.V.; Frasca, M. Distributed control of multiconsensus. *IEEE Trans. Autom. Control* **2021**, *66*, 2032–2044.
26. Gregoratti, D.; Matamoros, J. Distributed energy trading: The multiple-microgrid case. *IEEE Trans. Ind. Electron.* **2015**, *62*, 2551–2559.
27. Huang, Q.; Huang, R.; Hao, W.; Tan, J.; Fan, R.; Huang, Z. Adaptive power system emergency control using deep reinforcement learning. *IEEE Trans. Smart Grid* **2020**, *11*, 1171–1182.
28. Lowe, R.; Wu, Y.; Tamar, A.; Harb, J. Multi-agent Actor-Critic for mixed cooperative-competitive environments. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017.
29. Rashid, T.; Samvelyan, M.; Witt, C.S.D.; Farquhar, G.; Foerster, J.; Whiteson, S. Qmix: Monotonic value function factorisation for deep multi-agent reinforcement learning. *J. Mach. Learn. Res.* **2021**, *21*, 1–51.
30. Hao, J.; Yang, T.; Tang, H.; Bai, C.; Liu, J.; Meng, Z.; Liu, P.; Wang, Z. Exploration in deep reinforcement learning: From single-agent to multiagent domain. *IEEE Trans. Neural Netw. Learn. Syst.* **2024**, *35*, 8762–8782. [CrossRef]

31. Nguyen, V.; Wang, C.; Hsieh, Y. Electrification of highway transportation with solar and wind energy. *Sustainability* **2021**, *13*, 5456. [CrossRef]
32. Li, X.; Fang, Z.; Li, F.; Xie, S.; Cheng, S. Game-based optimal dispatching strategy for distribution network with multiple microgrids leasing shared energy storage. *Proc. CSEE* **2022**, *42*, 6611–6625.
33. Liu, H.; Wu, W. Two-stage deep reinforcement learning for inverter-based Volt-VAR control in active distribution networks. *IEEE Trans. Smart Grid* **2021**, *12*, 2037–2047.
34. DAsl, K.; Seifi, A.R.; Rastegar, M.; Dabbaghjamanesh, M.; Hatziargyriou, N.D. Distributed two-level energy scheduling of networked regional integrated energy systems. *IEEE Syst. J.* **2022**, *16*, 5433–5444.
35. Xu, G.; Lin, Z.; Wu, Q.; Tan, J.; Chan, W.K.V. Bi-level hierarchical model with deep reinforcement learning-based extended horizon scheduling for integrated electricity-heat systems. *Electr. Power Syst.* **2024**, *229*, 110195.
36. Siu, J.Y.; Kumar, N.; Panda, S.K. Command authentication using multiagent system for attacks on the economic dispatch problem. *IEEE Trans. Ind. Appl.* **2022**, *58*, 4381–4393.
37. Zheng, S.; Trott, A.; Srinivasa, S.; Parkes, D.C.; Socher, R. The AI economist: Optimal economic policy design via two-level deep reinforcement learning. *Sci. Adv.* **2022**, *8*, 13332.
38. Chamandoust, H. Optimal hybrid participation of customers in a smart micro-grid based on day-ahead electrical market. *Artif. Rev.* **2022**, *55*, 5891–5915. [CrossRef]
39. She, B.; Li, F.; Cui, H.; Zhang, J.; Bo, R. Fusion of microgrid control with model-free reinforcement learning: Review and vision. *IEEE Trans. Smart Grid* **2023**, *14*, 3232–3245.
40. Zhou, Q.; Shahidehpour, M.; Li, Z.; Xu, X. Two-layer control scheme for maintaining the frequency and the optimal economic operation of hybrid ac/dc microgrids. *IEEE Trans. Power Syst.* **2019**, *34*, 64–75.
41. Gong, Z.; Liu, C.; Shang, L.; Lai, Q.; Terriche, Y. Power decoupling strategy for voltage modulated direct power control of voltage source inverters connected to weak grids. *IEEE Trans. Sustain. Energy* **2023**, *14*, 152–167.
42. Brandao, D.I.; Ferreira, W.M.; Alonso, A.M.S.; Tedeschi, E.; Marafão, F.P. Optimal multiobjective control of low-voltage AC microgrids: Power flow regulation and compensation of reactive power and unbalance. *IEEE Trans. Smart Grid* **2020**, *11*, 1239–1252.
43. Das, A.; Wu, D.; Ni, Z. Approximate dynamic programming with policy-based exploration for microgrid dispatch under uncertainties. *Int. Electr. Power Energy Syst.* **2022**, *142*, 108359.
44. Lee, J.T.; Anderson, S.; Vergara, C.; Callaway, D.S. Non-intrusive load management under forecast uncertainty in energy constrained microgrids. *Electr. Power Syst. Res.* **2021**, *190*, 106632.
45. Liu, W.; Zhuang, P.; Liang, H.; Peng, J.; Huang, Z. Distributed economic dispatch in microgrids based on cooperative reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 2192–2203. [PubMed]
46. Mu, C.; Zhang, Y.; Gao, Z.; Sun, C. ADP-based robust tracking control for a class of nonlinear systems with unmatched uncertainties. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *50*, 4056–4067. [CrossRef]
47. Mu, C.; Sun, C.; Wang, D.; Song, A. Adaptive tracking control for a class of continuous-time uncertain nonlinear systems using the approximate solution of HJB equation. *Neurocomputing* **2017**, *260*, 432–442.
48. Shi, Z.; Wang, W.; Huang, Y.; Li, P.; Dong, L. Simultaneous optimization of renewable energy and energy storage capacity with the hierarchical control. *CSEE J. Power Energy Syst.* **2022**, *8*, 95–104.
49. Esfahani, M.M.; Hariri, A.; Mohammed, O.A. A multiagent-based game-theoretic and optimization approach for market operation of multimicrogrid systems. *IEEE Trans. Ind. Inform.* **2019**, *15*, 280–292.
50. Hong, S.-H.; Lee, H.-S. Robust energy management system with safe reinforcement learning using short-horizon forecasts. *IEEE Trans. Smart Grid* **2023**, *14*, 2485–2488.
51. Wang, K.; Mu, C.; Ni, Z.; Liu, D. Safe reinforcement learning and adaptive optimal control with applications to obstacle avoidance problem. *IEEE Trans. Autom. Sci. Eng.* **2024**, *21*, 4599–4612.
52. Wen, S.; Xiong, W.; Qiu, J. MPC-based frequency control strategy with a dynamic energy interaction scheme for the grid-connected microgrid system. *J. Frankl. Inst.* **2020**, *357*, 2736–2751.
53. Bi, W.; Shu, Y.; Dong, W.; Yang, Q. Real-time energy management of microgrid using reinforcement learning. In Proceedings of the 2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Xuzhou, China, 16–19 October 2020; pp. 38–41.
54. Xiong, L.; Tang, Y.; Mao, S.; Liu, H.; Meng, K.; Dong, Z.; Qian, F. A two-level energy management strategy for multi-microgrid systems with interval prediction and reinforcement learning. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2022**, *69*, 1788–1799.
55. Zhu, Y.; Nie, C.; Chen, B. Study on multi game cooperative scheduling of microgrid cluster system under hybrid time-scale. *Power Syst.* **2020**, *47*, 3249–3260.
56. Dreglea, A.; Foley, A.; Häger, U.; Sidorov, D.; Tomin, N. Hybrid renewable energy systems, load and generation forecasting, new grids structure, and smart technologies. In *Solving Urban Infrastructure Problems Using Smart City Technologies*; Vacca, J.R., Ed.; Elsevier: Amsterdam, The Netherlands, 2021; pp. 475–484.

57. Ding, L.; Lin, Z.; Shi, X.; Yan, G. Target-value-competition-based multi-agent deep reinforcement learning algorithm for distributed nonconvex economic dispatch. *IEEE Trans. Power Syst.* **2023**, *38*, 204–217.
58. Chen, W.; Li, T. Distributed economic dispatch for energy internet based on multiagent consensus control. *IEEE Trans. Autom. Control* **2021**, *66*, 137–152.
59. He, Q.; Ding, L.; Kong, Z.-M.; Hu, P.; Guan, Z.-H. Distributed scheme for line overload mitigation with linearized ac power flow. *IEEE Trans. Circuits Syst. II Express Briefs* **2022**, *69*, 2877–2881.
60. Liu, C.; Xu, X.; Hu, D. Multiobjective reinforcement learning: A comprehensive overview. *IEEE Trans. Syst. Man Cybern. Syst.* **2015**, *45*, 385–398.
61. Khamis, M.A.; Gomaa, W. Adaptive multi-objective reinforcement learning with hybrid exploration for traffic signal control based on cooperative multi-agent framework. *Eng. Appl. Artif.* **2014**, *29*, 134–151. [CrossRef]
62. Dong, P.; Xu, L.; Lin, Y.; Liu, M. Multi-objective coordinated control of reactive compensation devices among multiple substations. *IEEE Trans. Power Syst.* **2018**, *33*, 2395–2403. [CrossRef]
63. Nowak, S.; Chen, Y.C.; Wang, L. Distributed measurement-based optimal der dispatch with estimated sensitivity models. *IEEE Trans. Smart Grid* **2022**, *13*, 2197–2208.
64. Zhang, Q.; Dehghanpour, K.; Wang, Z.; Qiu, F.; Zhao, D. Multi-agent safe policy learning for power management of networked microgrids. *IEEE Trans. Smart Grid* **2021**, *12*, 1048–1062. [CrossRef]
65. Yan, Z.; Xu, Y. Real-time optimal power flow: A lagrangian based deep reinforcement learning approach. *IEEE Trans. Power Syst.* **2020**, *35*, 3270–3273. [CrossRef]
66. Liu, Y.; Li, Y.; Xin, H.; Gooi, H.B.; Pan, J. Distributed optimal tie-line power flow control for multiple interconnected ac microgrids. *IEEE Trans. Power Syst.* **2019**, *34*, 1869–1880. [CrossRef]
67. Mu, C.; Shi, Y.; Xu, N.; Wang, X.; Tang, Z.; Jia, H.; Geng, H. Multi-objective interval optimization dispatch of microgrid via deep reinforcement learning. *IEEE Trans. Smart Grid* **2024**, *15*, 2957–2970. [CrossRef]
68. Lin, N.; Orfanoudakis, S.; Cardenas, N.O.; Giraldo, J.S.; Vergara, P.P. Powerflownet: Power flow approximation using message passing graph neural networks. *Int. J. Electr. Power Energy Syst.* **2024**, *160*, 110112. [CrossRef]
69. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In Proceedings of the International Conference on Machine Learning (ICML), Stockholm, Sweden, 10–15 July 2018; pp. 1861–1870.
70. Yang, Y.; Yang, Z.; Yu, J.; Xie, K.; Jin, L. Fast economic dispatch in smart grids using deep learning: An active constraint screening approach. *IEEE Internet Things J.* **2020**, *7*, 11030–11040. [CrossRef]
71. Han, X.; Mu, C.; Yan, J.; Niu, Z. An autonomous control technology based on deep reinforcement learning for optimal active power dispatch. *Int. J. Electr. Power Energy Syst.* **2023**, *145*, 108686. [CrossRef]
72. Li, H.; Wang, L.; Lin, D.; Zhang, X. A nash game model of multi-agent participation in renewable energy consumption and the solving method via transfer reinforcement learning. *Proc. CSEE* **2019**, *39*, 3249–3260.
73. Li, M.; Wei, W.; Chen, Y.; Ge, M.-F.; Catalão, J.P.S. Learning the optimal strategy of power system operation with varying renewable generations. *IEEE Trans. Sustain. Energy* **2021**, *12*, 2293–2305. [CrossRef]
74. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [CrossRef]
75. Liu, H.; Wu, W. Online multi-agent reinforcement learning for decentralized inverter-based volt-var control. *IEEE Trans. Smart Grid* **2021**, *12*, 2980–2990.
76. Sun, X.; Qiu, J. Two-Stage Volt/Var Control in Active Distribution Networks With Multi-Agent Deep Reinforcement Learning Method. *IEEE Trans. Smart Grid* **2021**, *12*, 2903–2912.
77. Zhang, X.; Liu, Y.; Duan, J.; Qiu, G.; Liu, T.; Liu, J. DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources. *IEEE Trans. Power Syst.* **2021**, *36*, 5465–5475. [CrossRef]
78. Li, X.; Luo, F.; Li, C. Multi-agent deep reinforcement learning-based autonomous decision-making framework for community virtual power plants. *Appl. Energy* **2024**, *360*, 122813.
79. Cao, D.; Zhao, J.; Hu, W.; Ding, F.; Huang, Q.; Chen, Z.; Blaabjerg, F. Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs. *IEEE Trans. Smart Grid* **2021**, *12*, 4137–4150.
80. Hou, X.; Guo, Z.; Wang, X.; Qian, T.; Zhang, J.; Qi, S.; Xiao, J. Parallel learner: A practical deep reinforcement learning framework for multi-scenario games. *Knowl.-Based Syst.* **2022**, *236*, 107753. [CrossRef]
81. Dong, L.; Lin, H.; Qiao, J.; Zhang, T.; Zhang, S.; Pu, T. A coordinated active and reactive power optimization approach for multi-microgrids connected to distribution networks with multi-actor-attention-critic deep reinforcement learning. *Appl. Energy* **2024**, *373*, 123870.
82. Abdessameud, A.; Polushin, I.G.; Tayebi, A. Distributed coordination of dynamical multi-agent systems under directed graphs and constrained information exchange. *IEEE Trans. Autom. Control* **2017**, *62*, 1668–1683. [CrossRef]

83. Li, J.; Zhang, R.; Wang, H.; Liu, Z.; Lai, H.; Zhang, Y. Deep reinforcement learning for voltage control and renewable accommodation using spatial-temporal graph information. *IEEE Trans. Sustain.* **2024**, *15*, 249–262. [CrossRef]
84. Banerjee, S.; Balaban, E.; Shirley, M.; Bradner, K.; Pavone, M. Contingency planning using Bi-level Markov Decision Processes for space missions. In Proceedings of the 2024 IEEE Aerospace Conference, Big Sky, MT, USA, 2–9 March 2024; pp. 1–9.
85. Wang, K.; Mu, C. Learning-based control with decentralized dynamic event-triggering for vehicle systems. *IEEE Trans. Ind.* **2023**, *19*, 2629–2639.
86. Chi, P.; Wang, Z.; Liao, H.; Li, T.; Wu, X.; Zhang, Q. Application of artificial intelligence in the new generation of underwater humanoid welding robots: A review. *Artif. Intell. Rev.* **2024**, *57*, 306.
87. de Mars, P.; O'Sullivan, A. Applying reinforcement learning and tree search to the unit commitment problem. *Appl. Energy* **2021**, *302*, 117519.
88. Mu, C.; Peng, J.; Sun, C. Hierarchical multiagent formation control scheme via Actor-Critic learning. *IEEE Trans. Neural Networks Learn. Syst.* **2023**, *34*, 8764–8777.
89. Zhou, G.; Tian, W.; Buyya, R.; Xue, R.; Song, L. Deep reinforcement learning-based methods for resource scheduling in cloud computing: A review and future directions. *Artif. Intell. Rev.* **2024**, *57*, 1–42.
90. Mu, C.; Wang, K.; Sun, C. Learning control supported by dynamic event communication applying to industrial systems. *IEEE Trans. Ind. Inform.* **2021**, *17*, 2325–2335.
91. Kozlov, A.N.; Tomin, N.V.; Sidorov, D.N.; Lora, E.E.S.; Kurbatsky, V.G. Optimal Operation Control of PV-Biomass Gasifier-Diesel-Hybrid Systems Using Reinforcement Learning Techniques. *Energies* **2020**, *13*, 2632. [CrossRef]
92. Han, Y.; Zhang, K.; Li, H.; Coelho, E.A.A.; Guerrero, J.M. MAS-based distributed coordinated control and optimization in microgrid and microgrid clusters: A comprehensive overview. *IEEE Trans. Power Electron.* **2018**, *33*, 6488–6508.
93. Xu, Y.; Dong, Z.; Li, Z.; Liu, Y.; Ding, Z. Distributed optimization for integrated frequency regulation and economic dispatch in microgrids. *IEEE Trans. Smart Grid* **2021**, *12*, 4595–4606.
94. Xu, Y.; Dong, Z.; Li, Z.; Liu, Y.; Ding, Z. Day-ahead optimal dispatching of hybrid power system based ondeep reinforcement learning. *Cogn. Comput. Syst.* **2022**, *4*, 351–361.
95. Mu, C.; Wang, K.; Qiu, T. Dynamic event-triggering neural learning control for partially unknown nonlinear systems. *IEEE Trans. On Cybern.* **2022**, *52*, 2200–2213.
96. Guo, G.; Zhang, M.; Gong, Y.; Xu, Q. Safe multi-agent deep reinforcement learning for real-time decentralized control of inverter based renewable energy resources considering communication delay. *Appl. Energy* **2023**, *349*, 121648. [CrossRef]