

Article

A Deep Reinforcement Learning Approach for Microgrid Energy Transmission Dispatching

Shuai Chen ^{1,2}, Jian Liu ¹, Zhenwei Cui ¹, Zhiyu Chen ³, Hua Wang ¹ and Wendong Xiao ^{2,*} 

¹ Kunlun Digital Technology Co., Ltd., Beijing 100007, China; chenshuai03@cnpc.com.cn (S.C.); liujian8080@petrochina.com.cn (J.L.); cuizhenwei@petrochina.com.cn (Z.C.); wanghua@petrochina.com.cn (H.W.)

² School of Automation and Electrical Engineering, University of Science and Technology Beijing, Beijing 100083, China

³ State Grid Information & Telecommunication Branch, Beijing 100761, China; zhiyu-chen@sgcc.com.cn

* Correspondence: wdxiao@ustb.edu.cn

Abstract: Optimal energy transmission dispatching of microgrid systems involves complicated transmission energy allocation and battery charging/discharging management and remains a difficult and challenging research problem subject to complex operation conditions and action constraints due to the randomness and volatility of new energy. Traditional microgrid transmission dispatching mainly considers the matching of the demand side and the supply side from a macro perspective, without considering the impact of line loss. Therefore, a Hierarchical Deep Q-network (HDQN) approach for microgrid energy dispatching is proposed to address this issue. The approach takes the power flow of each line and the battery charging/discharging behavior as decision variables to minimize the system operation cost. The proposed approach employs a two-layer agent optimization architecture for simultaneously processing the discrete and continuous variables, with one agent making upper layer decisions on the charging and discharging behavior of the batteries, and the other agent making lower layer decisions on the transmission energy allocation for the line. The experimental results indicate that the proposed approach achieves better performance than existing approaches.



Citation: Chen, S.; Liu, J.; Cui, Z.; Chen, Z.; Wang, H.; Xiao, W. A Deep Reinforcement Learning Approach for Microgrid Energy Transmission Dispatching. *Appl. Sci.* **2024**, *14*, 3682. <https://doi.org/10.3390/app14093682>

Academic Editor: Alessandro Lo Schiavo

Received: 3 April 2024
Revised: 22 April 2024
Accepted: 22 April 2024
Published: 26 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: microgrid; optimal energy transmission dispatching; deep Q-network; deep reinforcement learning

1. Introduction

With the rapid development of the economy, industrial and residential power consumption is increasing day by day, and energy demand presents a diversified and personalized development trend. At present, long-distance transmission struggles meet the requirements of multi-scale distribution networks for power supply reliability and diversity. In order to address these power needs, research on the efficient operation of microgrids has received increasing attention and become a strong support for the green development of a low-carbon society. A microgrid is a regional power generation, power distribution and power consumption system composed of distributed power sources, energy storage equipment, power electronic energy conversion devices, power loads, monitoring and protection devices, and control components. Its purpose is to use distributed power sources flexibly and efficiently, solve the problem of full consumption of diversified energy sources and realize the stable operation of the power system in the region. However, due to the intermittency, randomness and uncontrollable nature of new energy power generation, new energy alone cannot provide a sustainable and stable power supply. Therefore, it is necessary to use adjustable power to meet the diversified demands of power loads.

As an important part of the new power system, the microgrid is being promoted on a large scale as the main distributed new energy consumption method, and it plays a key role

in maintaining the smooth operation of the distribution power grid. The deployment of microgrid systems can effectively address the power needs of remote areas, islands, or areas that cannot be connected to a larger grid. With the further research into and application of microgrids, how to efficiently manage power distribution and energy transmission in microgrid systems to maximize economic benefits has become a significant research direction. However, different from the traditional large power grid, the operation and dispatching of microgrids faces many challenges: (1) The uncertainty of intermittent renewable energy power generation makes it difficult for a microgrid to make accurate day-ahead dispatching plans based on prediction, resulting in the imbalance between supply and demand; (2) Bidirectional power flow caused by connecting distributed renewable energy power generation into the grid can easily cause busbar voltage fluctuation, resulting in the instability of microgrid operation [1]. Therefore, due to the complexity of energy types and application environments, the overall economic operation optimization of microgrids faces certain challenges [2,3].

Compared with the optimal dispatching of the current large power grid, the optimal dispatching of the microgrid is more complicated. Distributed power sources in the microgrid have different forms and different operation characteristics. New energy sources such as photovoltaic and wind power are greatly affected by the weather and environment, and their power capacity is small. Many scholars have put forward their own solutions to the problems faced by optimal dispatching of the microgrid. The economic and social benefits of microgrid operation are usually considered in the optimization of microgrid dispatching to minimize the operation cost and ensure the health and power balance of the system. The literature [4] aims at minimizing energy production and transmission cost, power loss, and load-dumping operations, to improve system performance to the maximum extent in general. The literature [5] manages the variables of the power system through optimal power control and supplies the load at the minimum or a reasonable cost without violating the system constraints. In order to improve energy efficiency, the literature [6] proposes a LAN energy management method based on optimal energy flow to realize accurate control of distributed energy storage.

Although the above studies have solved the power distribution problem of microgrid lines to a large extent, these studies have not fully considered the grid structure of the low-voltage microgrid and its power distribution characteristics. Due to the low-voltage power supply characteristics of the microgrid, the line losses of its transmission lines cannot be ignored. The traditional research usually ignores the influence of line loss, which has some deviation in Hierarchical application. Therefore, in order to solve this problem, a comprehensive optimization dispatching model of microgrid energy considering line loss is proposed in this paper. According to the characteristics of microgrids, a multi-time Markov decision model was constructed, and a Hierarchical deep Q-network (HDQN) method was designed. The optimal decision estimation of discrete variables and continuous variables was realized through a neural network, so as to obtain the approximate optimal dispatching strategy, which involves complicated transmission energy allocation and battery charging/discharging management.

In essence, the microgrid energy transmission dispatching problem studied in this paper is a constrained non-linear programming problem. However, the current research mainly focuses on the overall dispatching of energy and lacks in-depth analysis and research on the energy distribution of each power consumption unit; that is to say, it is difficult to allocate the power of each line.

Considering the relationship between power and voltage/current, the microgrid optimal transmission dispatching problem studied in this paper is a special nonlinear programming problem, that is, a quadratic programming problem with quadratic constraints. Therefore, in this paper, power flow control is included in the behavioral reward as the goal, to form a microgrid power controller based on deep reinforcement learning, in which the adjustable resources of load, diesel generator and energy storage device are time-varying. The paper adopts a two-stage control mode. Firstly, the output power of

the system should meet the load demand of the microgrid. Secondly, the adjustability and continuity of the power output or input are ensured when charging/discharging the energy storage. Finally, in the case of considering line loss, the power on all lines and the actions of battery charging/discharging are controlled to ensure the dynamic balance among renewable energy, energy storage, diesel generators and loads, so as to ensure the optimal steady state operation of the microgrid.

Considering that the energy supply of the microgrid is provided by non-dispatchable energy sources (e.g., photovoltaic and wind power) and dispatchable energy sources (e.g., diesel generation and battery), its energy dispatching optimization is to optimize power generation output and distribution line flow by regulating the supply of schedulable energy and reducing the operation cost of the microgrid. Different from the high-voltage transmission system, the distinct low-voltage power supply characteristics of a microgrid result in significantly higher power losses as the line lengthens, due to the increased current levels required for transmitting the same amount of power. Therefore, the optimal dispatching of energy in the microgrid should consider the power loss on the distribution line. However, most of the existing research does not consider the power loss of the transmission line, making the operation cost predicted by various models higher. Therefore, taking into account the power loss of distribution line, this paper reduces the energy loss and operation cost of the microgrid by adjusting the transmission power of each line.

This paper is organized into six sections. Section 1 is a general introduction. Section 2 introduces the work related to the optimal dispatching of microgrids. Section 3 presents the energy transmission dispatching problem in microgrid systems. Section 4 proposes the deep reinforcement learning approach model and framework. Section 5 introduces the experimental results and discusses them through case studies. The conclusion is drawn in Section 6.

2. Related Work

The energy transmission dispatching problem involves determining the optimal schedule for the transmission of energy over a network of transmission lines, taking into account factors such as the availability of energy sources, the demand for energy, and the constraints and limitations of the transmission system. This problem is particularly complex due to the large number of variables involved and the need to account for uncertain factors such as weather patterns and unpredictable changes in energy demand. To solve the energy transmission dispatching problem, advanced mathematical models and optimization techniques are used to analyze and optimize the energy delivery process. These models take into account factors such as energy demand, transmission line capacity, transmission losses and the availability of energy sources and use sophisticated algorithms to determine the optimal schedule for energy transmission. Overall, energy transmission dispatching is a critical component of the energy industry, as it enables the reliable and efficient delivery of energy to customers while optimizing the use of resources and minimizing cost.

In order to optimize the energy dispatching strategies of microgrids in intermittent and random environments, many scholars have proposed model-based optimization dispatching methods. The literature [7] proposes a dual MPC multi-energy microcomputer optimization dispatching strategy based on the TUBE model prediction control (TMPC). The model mentioned by the multi-energy microgrid has a good robustness. A rogue mixed integer two-order cone planning (R-MISOCP) model is proposed to use the elastic optimization dispatching of microgrid to achieve minimal cost operation of the microgrid [8]. The literature [9] proposes an optimal energy management strategy, considering the MMG network that minimizes the operation cost of running constraints and carbon emissions and realizes the optimal dispatching of energy. The literature [10] presents an energy management method for multi-energy microgrids based on DQN, which overcomes the challenges of complex energy trading and multi-energy coupling, demonstrating superior convergence, stability and operational efficiency in a three-MEMG test system. The literature [11] proposes a two-stage random planning model such as new energy forces and electricity

prices and carries out an optimization scheme for predict control through the model. These methods need to establish accurate microgrid models for their solutions, but, in fact, most of the micro-power grid systems cannot complete the accurate construction of the model. Under the premise of a large number of assumptions, the optimized dispatching of the microcomputer grid is realized. Therefore, when the environment becomes complex and controllable variables increase significantly, the model-driven methods have problems such as difficulty building, disaster and low-computing efficiency, and the scope of application is limited. At the same time, considering that the microgrid is a complex real-time dynamic system, the control method based on a model-driven control can no longer achieve the accurate optimization control of the microgrid.

In order to reduce the dependence on the model, the data-driven method is proposed in large quantities to solve the optimized dispatching problem of microgrids and improve the economic goals as much as possible under various constraints. Chen et al. proposed a modified deep Q-network to solve the power-charging and -discharging problem of user-side battery energy storage, reducing the cost and energy consumption of the charging and discharging of the industrial park [12]. Paudyal et al. transformed the mixed integer nonlinear programming problem into a nonlinear programming problem based on actual needs and established a three-phase power distribution optimization power flow model, thereby reducing the computational burden and facilitating the Hierarchical implementation and application [13]. The literature [14] adopted Long-Short-Term memory (LSTM), featuring the past renewable power generation and load sequences, and continuously made optimal operation decisions online without relying on prediction models. The literature [15] proposed a multi-intelligent physical enhancement learning method with an attention mechanism. You can learn the optimal strategy of microgrids without complex system modeling and significantly reduce the cost of the microgrid. An online dispatch method was used based on the real-time energy management of microgrids, based on imitation learning (IL) [16]. A real-time optimization method based on value function was proposed to minimize the total energy cost [17]. In addition, many people use simulation annealing (SA) and adaptive dynamic programming (ADP) to process the optimized dispatching problem of microgrids [18–21]. Significantly, various studies have not conducted actual analysis on the energy distribution of various energy-consuming units, and it is difficult to grasp the distribution of the energy flow on the transmission line.

At present, on the basis of comprehensive consideration of the randomness of distributed power output and the volatility of supply and demand prices, the mainstream microgrid operation cost optimization uses the output power of diesel power generation (DG), photovoltaic power generation (PV), wind-powered electricity generation, energy storage battery and electrical load as the control variables [22–24]. Compared with high-voltage transmission systems, microgrids are low-voltage transmission networks, and the resistance of transmission lines has an important impact on the economic dispatch of energy. The transmission resistance causes more obvious power loss in the low-voltage network. In order to further improve the economic operation capability of the microgrid, energy distribution that takes into account the resistance characteristics of the transmission line is a direction that needs further attention. Therefore, it is necessary to consider the power loss of the transmission line for optimizing and dispatching the low-voltage microgrid. At the same time, we can dynamically optimize the energy dispatching, thereby reducing operation cost [25].

3. Problem Formulation

As shown in Figure 1, this paper considers a microgrid system, which consists of PVs, DGs, batteries and loads. PV and DG are important carriers of energy required to support the operation of microgrids and can provide energy for user side loads and energy storage batteries. Due to the close correlation between the power output of photovoltaic power generation and sunlight intensity, its large volatility restricts the accurate dispatching of energy. The power output of DG and energy storage batteries is relatively stable, and it is

an energy source that can be easily dispatched, which can ensure the stable operation and timely response of the microgrid. The energy storage battery can be used as both a load and a power source in the system. When the power output of PV and DG cannot meet the energy demand of the load, the energy storage battery maintains the balance between the supply and demand of electric energy through a discharging operation. When the power output of PV and DG exceeds the energy demand of the load, the battery stores energy through a charging operation to increase the consumption ratio of non-dispatchable energy.

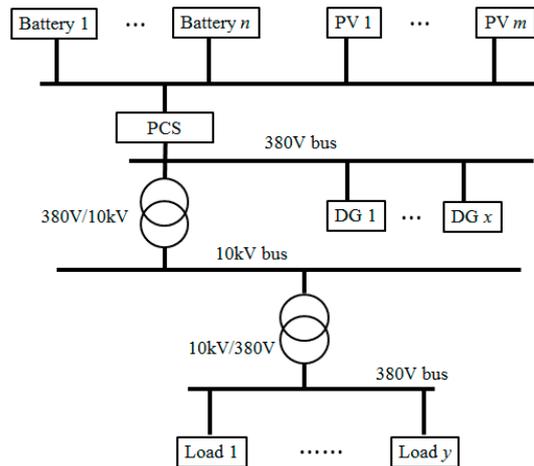


Figure 1. The framework of a microgrid system.

The paper mainly studies the power balance problem among the power supply side, the load side and the energy storage side, as well as the power mismatch problem among the controllable generation unit, controllable energy storage facility and adjustable load. At the same time, if the power generation side does not have enough energy to meet the demand of the loads, the charging and discharging of energy storage facilities and adjustable loads will be considered. This process will continue until the mismatch problem is resolved.

3.1. Photovoltaic Power Generation System Model

The power output of PV is directly related to sunlight intensity and ambient temperature. Its mathematical model can be represented by

$$P_{pv} = h_{pv} Q_{pv} \frac{G_T}{G_{STC}} [1 + \theta(T_c - T_{STC})] \tag{1}$$

where p_{pv} and h_{pv} are the output power and derating factor of PV, respectively [26]. Generally, the derating factor is set to 0.9. Q_{pv} represents the capacity of PV. G_T and G_{STC} are the light intensity at the working point and the light intensity under standard conditions, respectively. θ and T_c are the temperature coefficient and battery temperature, respectively. T_{STC} is the reference temperature of the energy storage battery. In actual operation, the reference temperature of the energy storage battery is set to 25 °C.

3.2. Diesel Power Generation System Model

The diesel power generation is a dispatchable energy source, and the relationship between its output power and operation cost can be given by

$$C_{dg}(P_{dg}(t)) = \alpha(P_{dg}(t)\Delta t)^2 + \beta P_{dg}(t)\Delta t + \gamma \tag{2}$$

where $C_{dg}(P_{dg}(t))$ and $P_{dg}(t)$ are the generation cost and output power of DG; α , β and γ are set to 0.00085, 0.11 and 6, respectively, in this paper [18].

3.3. Energy Storage Battery Model

Energy storage batteries play a very important role in energy balance in microgrids and can regulate the supply and demand of electric energy [27]. In this paper, the battery charging and discharging model can be modeled by

$$E_{b,t+1} = E_{b,t} - P_{b,t} \times \eta(P_{b,t}) \tag{3}$$

where $E_{b,t}$ and $E_{b,t+1}$ are the battery energy at time t and $t + 1$, respectively [28]. $P_{b,t}$ is the power output at time t of the energy storage battery. $\eta(P_{b,t})$ is the total efficiency of the energy storage battery either charging or discharging. It can be represented by

$$\eta(P_{b,t}) = 0.898 - 0.173|P_{b,t}|/P_r \quad s.t. \quad P_r > 0 \tag{4}$$

where P_r is the rated output power of the energy storage battery. The value of $P_{b,t}$ is positive, negative and zero for the charging operation, discharging operation and non-operation of the energy storage battery, respectively.

3.4. Operation Conditions Constraints

Power balance of microgrid is the primary condition for the stable operation, and the power demand and supply in each time period need to satisfy the power balance equation. The equation is defined as

$$\sum_{i=1}^y P_{load\ i,t} = \sum_{i=1}^m P_{pv\ i,t} + \sum_{i=1}^x P_{dg\ i,t} + \sum_{i=1}^n P_{b\ i,t} \tag{5}$$

$$P_{dg}^{min} \leq P_{dg,t} \leq P_{dg}^{max} \tag{6}$$

$$0 \leq P_{pv,t} \leq P_{pv}^{max} \tag{7}$$

where P_{dg}^{min} , P_{dg}^{max} , P_{pv}^{max} are the lower and upper limits of generator output power and the upper limits of photovoltaic output power, respectively.

The output power of DG and PV is constrained by the upper and lower boundaries in Formula (6) and Formula (7). The working time of the energy storage battery is limited by Formula (8). Its power rating and maximum charging/discharging rate are limited by Formula (9).

$$E_{b,t}^{min} \leq E_{b,t} \leq E_{b,t}^{max} \tag{8}$$

$$P_{b,t}^{min} \leq P_{b,t} \leq P_{b,t}^{max} \tag{9}$$

When the energy storage battery provides energy for q loads, the total power allocated by the transmission line must meet with the constraints

$$0 \leq \sum_{i=1}^q u_{i,t} \leq P_{b,t}^{max} \tag{10}$$

Correspondingly, when the distributed power sources in the microgrid provide energy for f energy storage batteries, the power distributed by the transmission line needs to meet the constraints

$$0 \leq \sum_{i=1}^f u_{i,t} \leq P_{b,t}^{max} \tag{11}$$

$$P_{l,t} = I_{h,t}^2 R_h \tag{12}$$

$$I_{h,t} = \frac{u_{h,t}}{V_m} \tag{13}$$

The voltage V_m of the microgrid is generally 220 V. At the same time, the line loss of the h -th transmission line at time t can be calculated based on Formulas (12) and (13).

3.5. Objective Function

The goal of the optimization problem is to search the optimal power flow of the transmission line at each time period under the given state variables and calculate the minimum cost based on Formulas (5)–(13). Its expression is realized by

$$\min(C_t) = \sum_t^{+\infty} C_{dg,t}(x, u) + C_{m,t}(x, u) + C_{l,t}(x, u) \quad (14)$$

The operation cost of microgrid C_t mainly consists of the fuel cost $C_{dg,t}$, the operation and maintenance cost $C_{m,t}$ and the line loss cost $C_{l,t}$. These costs can be calculated by

$$C_{dg,t}(x, u) = U_p \times P_{dg,t} \quad (15)$$

$$C_{m,t}(x, u) = (\eta_1 \times P_{pv,t} + \eta_2 \times P_{dg,t}) \quad (16)$$

$$C_{l,t}(x, u) = \theta_1 \times P_{l,t} + \theta_2 \times P_{b,t} \quad (17)$$

The price of the diesel is usually set as $U_p = \text{USD } 1.2$, $\eta_1 = 0.0096$, $\eta_2 = 0.088$, $\theta_1 = 0.08$, $\theta_2 = 0.01$ [29]. From Formulas (12), (14) and (17), we can find that both the constraints and the objective function contain quadratic terms. Therefore, the problem can be understood as a quadratically constrained quadratic programming (QCQP) problem.

When the energy supply of the distributed energy source can meet the energy consumption demand of the microgrid, the energy storage battery can perform a charging operation to store energy. Correspondingly, when the energy supply is insufficient, the diesel power generation and energy storage system can generate enough energy to meet the operation needs of the microgrid.

4. The Proposed HDQN Approach

4.1. DQN

DQN is a typical reinforcement learning method based on value function, combining deep learning and Q-learning, and has good convergence performance. The structure of DQN is composed of the current value network, target value network, error function and experience replay unit. It estimates action-value functions using deep neural networks (DNN). In order to solve the problem of instability and non-convergence of the DNN approximation of the action-value function, the experience replay mechanism and target network can be adopted.

DQN evaluates the action through the target network. The size of the action-value function corresponding to the different actions is estimated to find the maximum action of the motion value function and the decision of the smart body. From the principle of the DQN approach, it can be seen that DQN can only handle limited action values. As the action space increases, DQN's computing volume increases sharply, so DQN is usually used to process the task of discrete action space.

4.2. Hierarchical Deep Q-Network Approach

In the process of a microgrid optimization decision, the agent can meet the actual demand of the load through a series of battery-charging and -discharging power decisions and the power distribution of each line. However, the charging and discharging decision of the battery is a discrete variable decision, and the power distribution of each line is a continuous variable decision. Therefore, when considering the line loss of the low-voltage distribution power grid, the energy dispatching problem of the microgrid is a mixed-decision problem with complex constraints, including discrete action-continuous variables. To solve these problems, an optimization approach based on DQN is proposed in this paper. The neural network is used to estimate the action values, so as to realize the decision of continuous variables. At the same time, the discrete action decision is realized

by combining with the DQN network, and, finally, the energy optimization dispatching of the microgrid is realized.

The action space of comprehensive optimization dispatching of micrometers includes two types of actions. One is the charging/discharging of energy storage equipment, which is a discrete action. The other is the specific transmission power of each line, which is a continuous energy distribution. The energy storage equipment at the time of the charging and discharging decisions will affect the future energy status of the microgrid system and has long-term regulatory significance. Therefore, this paper designs a double-layer HDQN framework in which the upper-layer agent is used for the charging and discharging decisions of energy storage devices, and the lower-layer agent receives the decision-making actions of the upper layer. Based on the upper-layer decision, an Actor-Critic network is used to obtain the optimal action that minimizes the system operation cost. All actions of the upper-layer agent and the lower-layer agent constitute the entire action space at that moment, acting together on the environment to obtain the next state.

The HDQN approach framework is shown in the Figure 2, including environment, experience pool, rewards and different neural networks. Agents at both the upper and lower layers can sense the current state of the environment. The evaluation network of the upper-layer agent can solve the long-term battery charging and discharging decisions, and the action and evaluation network of the lower-layer agent can solve the optimal line power distribution and output of each controllable unit at the current moment.

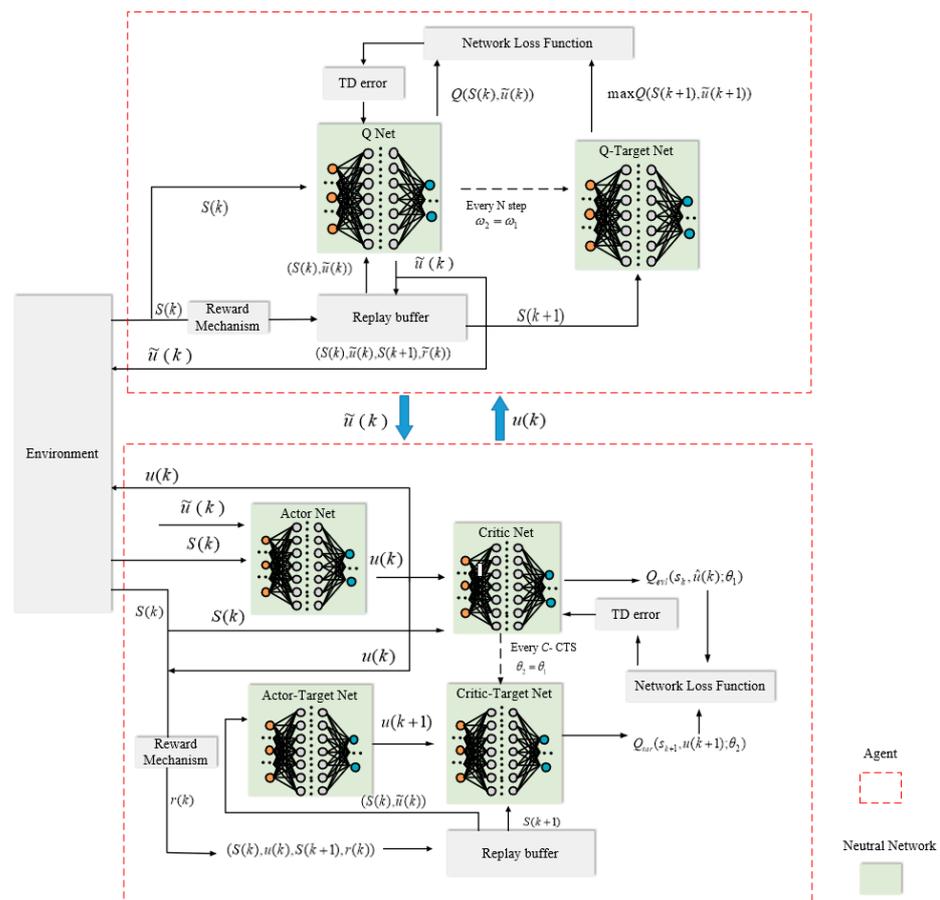


Figure 2. The framework diagram of the HDQN approach.

4.3. Upper Layer Model Design

4.3.1. Rewards

The return r obtained by the agent after selecting actions according to the load demand, power output and battery state of charge at time t has a negative value. Therefore, a higher

return means a lower operation cost. The objective function can be obtained through cumulative rewards, and this design allows the objective Function (14) to be included in the rewards:

$$\begin{aligned} \tilde{r}_t &= -\kappa C_t \\ \kappa &= \begin{cases} 1, & P_{pv,t} - P_{load,t} \geq 0 \\ -1, & P_{pv,t} - P_{load,t} < 0 \end{cases} \end{aligned} \tag{18}$$

where C_t is the system operation cost, and κ is an indicator function. When the photovoltaic output is greater than the load demand, its value is 1, which gives the agent a positive reward. When the photovoltaic output is less than the load demand, its value is -1 , which is equivalent to giving the agent a penalty, thereby affecting the agent’s decision.

4.3.2. State Space

The microgrid system includes multiple components, such as distributed new energy, diesel generators, energy storage facilities and loads. The state space S is a collection of all states. The state $s(t)$ includes the state of the energy storage battery at time t , the power demand of the load, the power output of the generator and the output power of the photovoltaic system:

$$s(t) = \{P_{pvm,t}, P_{dg,t}, E_{bn,t}, P_{loady,t}\} \tag{19}$$

where $E_{bn,t}$ is the energy of the n -th battery at time t . $P_{loady,t}$, $P_{pvm,t}$ and $P_{dg,t}$ are the power demand of the i -th load, output power of the m -th photovoltaic and output power of DG.

4.3.3. Action Space

Combining practical scenarios and considering simplifying training difficulty, the action space of the upper layer agent mainly includes three situations: charging, discharging, and no operation of the battery. When the value is 1, the battery is in the charging state, when the value is -1 , the battery is in the discharging state, and when the value is 0, the battery has no operation.

$$\tilde{u}(t) = \{1, 0, -1\} \tag{20}$$

4.4. Lower Layer Model Design

4.4.1. Rewards

The lower-layer agent minimizes system operation costs after upper-layer actions are given. The reward function is set based on the system objective function, which is to achieve cost optimization through cumulative rewards:

$$r_t = \begin{cases} -C_t & \text{sum}(a_{i,t}) \geq P_{load\ i,t} \\ \lambda C_t & \text{sum}(a_{i,t}) < P_{load\ i,t} \end{cases} \tag{21}$$

where $\lambda \in \mathbb{R}^+$ is a positive natural number. By introducing λ and adjusting the reward function appropriately, the penalty factor is higher when the generated power cannot meet the load demand. Therefore, under the condition of power balance, the system gives priority to meeting the load demand while meeting various constraints.

4.4.2. State Space

The state space of the lower-layer agent is consistent with that of the upper-layer agent.

4.4.3. Action Space

The action space A includes the output of controllable power on all lines. In this paper, all actions are the transmission power on each line, which is taken as a continuous value but is limited by the inherent attributes of the energy storage device, load and so on, such

as the maximum charging and discharging power of the battery. The action space can be expressed as

$$u(t) = \{u_{1,t}, \dots, u_{h,t}\} \quad (22)$$

where $u_{h,t}$ is the power flow of the h -th transmission line at time t . The optimized control law $u^*(t)$ is found by the operational decision $s(t)$.

4.5. Implementation of HDQN

To realize the optimal charging and discharging management and energy transmission dispatching, this paper uses the HDQN approach to control the actions of various devices in the microgrid. The upper-layer agent performs calculations based on the observed environmental state and its strategy at time t , thereby selecting a corresponding action and transmitting the action to the lower-layer agent at the same time. Based on this, the lower-layer agent calculates according to its strategy to obtain a set of determined actions. All actions of the upper and lower layers of the agent act on the running environment and obtain the next environment state. The algorithm steps are shown in Algorithm 1.

Algorithm 1. HDQN algorithm for microgrid

- 1: Randomly initialize the Q-network, the Actor network and the Critic network with weights $\theta^q, \theta^u, \theta^c$.
 - 2: Initialize the target network Q-target network, the Actor-target network and the Critic-target network with weights $\tilde{\theta}^q, \tilde{\theta}^u, \tilde{\theta}^c$.
 - 3: Initialize replay buffer $\{R_1, R_2\}$
 - 4: Initialize T, t
 - 5: **for** episode = 1, M **do**
 - 6: Receive the initial observation state S_1
 - 7: **for** $t = 1, T$ **do**
 - 8: With probability ε_1 , selects \tilde{U} random action \tilde{u}
 - 9: Otherwise selects $\tilde{u} = \max_u(s_t, u; \theta^q)$
 - 10: Execute action \tilde{u} , observe upper layer reward \tilde{r}_t and new state S_{t+1}
 - 11: Store transition $(s_t, \tilde{u}_t, \tilde{r}_t, s_{t+1})$ in R_1
 - 12: Sample a random minibatch of N_1 transitions $(s_i, \tilde{u}_i, \tilde{r}_i, s_{i+1})$ from R_1
 - 13: Set $y_i = \tilde{r}_i + \gamma \max_u Q(S_{i+1}, u; \theta^q)$
 - 14: Perform a gradient descent step on $L = \frac{1}{N_1} \sum_i (y_i - Q(s_i, \tilde{u}_i; \theta^q))^2$
 - 15: **end for**
 - 16: **for** $t = 1, T$ **do**
 - 17: With probability ε_2 selects U random action u
 - 18: Execute action (\tilde{u}, u) , observe lower-layer reward r_t and update state S_{t+1}
 - 19: Store transition s_t, u_t, r_t, s_{t+1} in R_2
 - 20: Sample a random minibatch of N_2 transitions (s_i, u_i, r_i, s_{i+1}) from R_2
 - 21: Set $y_i = r_i + \gamma \max_u Q(S_{i+1}, u; \theta^u, \theta^c)$
 - 22: Update critic by minimizing the loss $L = \frac{1}{N_2} \sum_i (y_i - Q(s_i, u_i; \theta^u, \theta^c))^2$
 - 23: Update the actor policy using the sampled policy gradient: $\nabla_{\theta^u} J \approx \frac{1}{N_2} \sum_i \nabla_u Q(s, u | \theta^c) \Big|_{s=s_i, u=u(s_i)} \nabla_{\theta^u} u(s | \theta^u) \Big|_{s_i}$
 - 24: Every N steps $\theta = \theta^q, \theta = \theta^u, \theta = \theta^c$
 - 25: **end for**
 - 26: **end for**
 - 27: return (\tilde{u}, u)
-

5. Experimental Analysis

This experiment considers a typical microgrid system, including a photovoltaic system, a diesel generator, two batteries and two loads [21]. The time interval is $\Delta t = 1$ h. Reference [30] for the typical daily load and irradiation data. The batteries capacities are set to 100 kWh in the experiment. The rated power output and maximum charging/discharging rate of the batteries are 16 kW. The initial energy of the batteries is 100 kWh. The rated power of PV and DG are 210 kW and 120 kW, respectively.

In this experiment, the Q-network of the upper-layer agent has three fully connected hidden layers. Through a large number of simulation experiments, the number of neurons is set to 300, 300 and 150 respectively. The activation function is the M-ReLU, whose hidden layer has 200 and 150 M-ReLU neurons [12]. The upper and lower layers use the same network parameters. The power transmission dispatching in this paper is a periodic task with a 24 h cycle. In this experiment, some initial parameters are summarized in Table 1.

Table 1. Parameters of the HDQN for the Microgrid Energy Dispatching Problem.

Parameter	Value
Actor learning rate	0.0005
Critic learning rate	0.0001
Batch size	256
Steps in one episode (T)	24
Discount factor	0.95
Iterates	1000
Calculation accuracy	0.0001
Replay memory	50,000

Figures 3 and 4 show the cost curves of the SA, ADP and proposed approach under the same conditions. Figure 3 shows the total cost generated by the three dispatching methods. From the experimental results, it can be seen that the total cost does not increase substantially between 09:00 and 16:00. The reason is that photovoltaic is the main power supply method during this period, and the total cost mainly comes from transmission line losses. At other times, the photovoltaic output drops to zero. At the same time, the diesel generator and the batteries serve as the main power supply for the load. Therefore, the cost of power supply is relatively high. Figure 4 shows the cost per time step of the SA, ADP and proposed approach. These experimental results show that the proposed method spends less on each time step than the SA and ADP, and the dispatching effect is better. The results show that the proposed approach costs less than the SA and ADP in each time step and has a better dispatching effect. The cost of the SA, ADP and proposed approach to provide services for all loads in one day is USD 1309.49, USD 1091.92 and USD 1050.58, respectively. Compared with the other two approaches, the proposed approach saves USD 258.91 and USD 41.34, respectively. Obviously, the proposed approach provides the optimal dispatching scheme, which can support the operation of the microgrid with minimal operation cost.

Figures 5 and 6 show the energy outputs of PV and DG at each time period. Through the analysis of the results, the energy outputs of PV and DG have a strong correlation. DG does not operate between 11:00 and 15:00, while the output power of PV is in a high operation state. At the same time, the energy output of PV can meet the energy demand of the load and provide energy for charging the energy storage battery. At other times, PV and DG work together to meet the energy demand of the load.

Figure 7 shows the power loss within 24 h, and the results show that power loss is worth paying attention to in the economical operation of microgrids. The total power losses of the SA and the ADP are 199.91 kW and 130.87 kW, respectively. Using the proposed approach, the total power losses are 113.77 kW, saving 86.14 kW and 17.1 kW in one day, respectively.

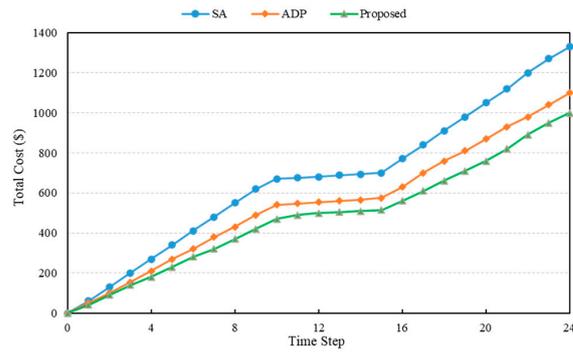


Figure 3. The total cost within 24 h for three approaches.

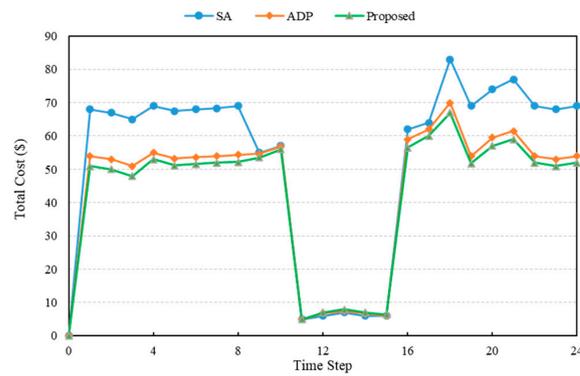


Figure 4. The phase cost within 24 h for three approaches.

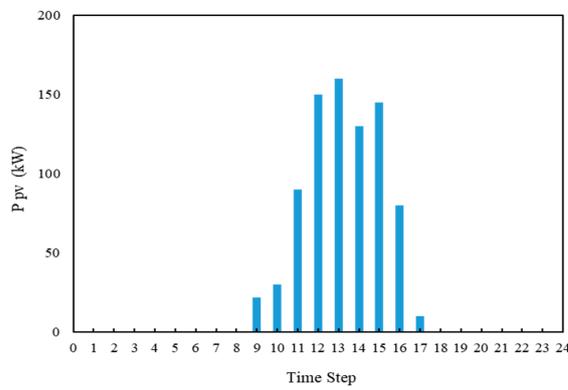


Figure 5. The energy output of PV within 24 h.

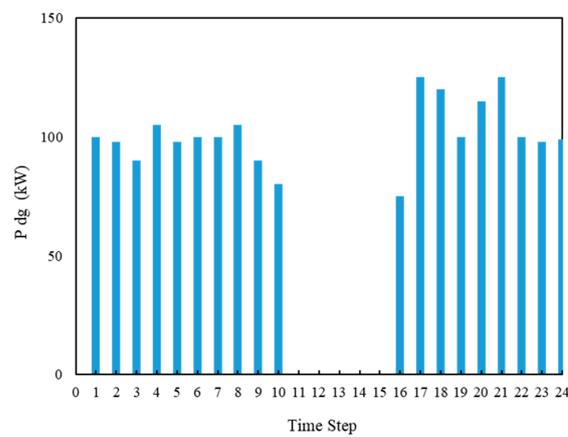


Figure 6. The energy output of DG within 24 h.

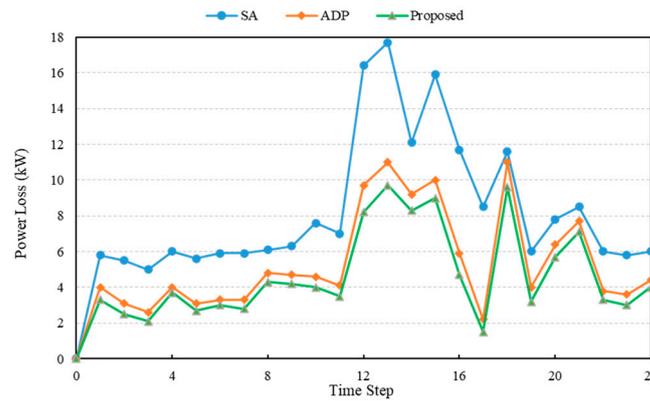


Figure 7. The power loss of the microgrid within 24 h.

The discount factor in the approach indicates the impact of the current behavior on the future, and the larger the discount factor, the more the model focuses on the impact generated by the current behavior. Therefore, the discount factor has an important impact on the convergence of the approach. In Figure 8, it is clear that different discount factors have certain influences on the system, but the whole system is within the error range. Different discount factors can make the approach converge to the optimal value and have the same convergence performance. It can be seen that the approach is effective for solving the problem.

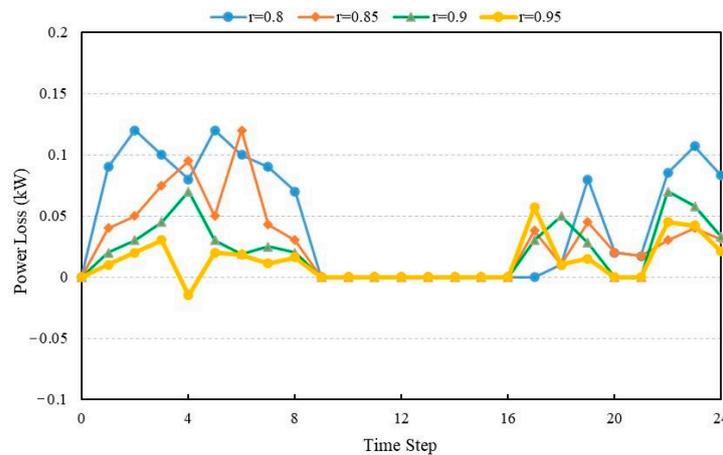


Figure 8. The cost error of each time step to different discount factors.

As can be seen above, when the total power supply of the microgrid is greater than the total demand, the proposed approach can achieve optimization of the operation and maintenance cost. But, if there is a lack of diesel generators in the microgrid (or if the output power of the diesel generators is zero), the total power supply may not fully meet the total demand, and the total power quality will be decreased. In this case, it is necessary to dispatch the power supply for different loads to ensure that the overall operation of the microgrid is satisfied to the greatest extent in the case of power shortages.

Figures 9 and 10 illustrate the total cost and step cost of using the SA, ADP and proposed approach without a diesel generator. It is still clear that in the absence of a diesel generator, the total cost and cost per time step obtained by using the proposed approach are lower than the SA and the ADP. The total cost of the ADP is USD 80.4 per day, while the total cost of the SA is USD 122.5 per day. The proposed approach saves USD 12 and USD 54.5 per day, with savings rates of about 15% and 44%, respectively. When a microgrid system does not have a diesel generator, the most important costs during the system operation are the line loss and the loss of energy storage batteries.

Without a diesel generator, the total cost is reduced, but the quality of the power supply is greatly affected, and the total power supply cannot fully meet the load demand.

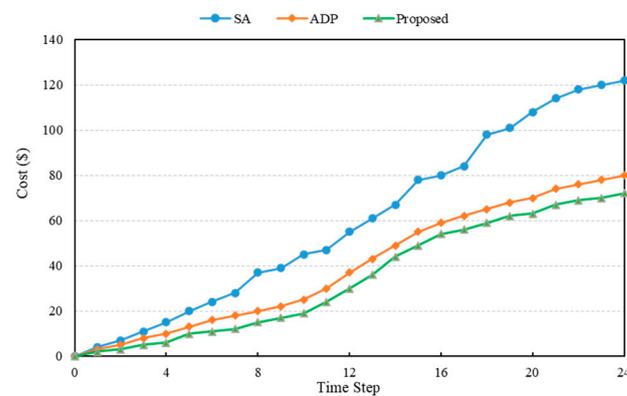


Figure 9. The total cost without DG within 24 h.

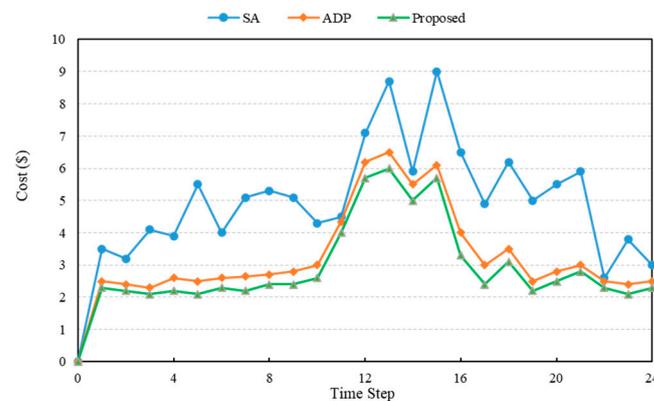


Figure 10. The step cost without DG within 24 h.

6. Conclusions

In this paper, the optimal energy transmission dispatching approach of the microgrid is introduced. An optimal approach for energy transmission dispatching based on an HDQN is proposed to achieve energy storage battery management and transmission energy allocation of the microgrid. To reduce complexity, the system is designed in two layers, each represented by an agent. The upper layer first makes charging and discharging decisions for the battery, and the lower layer uses the power on each transmission line as a decision variable, based on the decision results of the upper layer, to obtain a decision result that minimizes the overall operation cost of the system. Simulation results show that the proposed method combines the advantages of the HDQN to solve the problem of power allocation optimization in the microgrid. In continuous variable optimization, the proposed method achieves better performance than the SA and ADP.

The energy storage system is considered as a whole in the paper. In fact, there are also some scientific problems about the charging and discharging dispatching and battery life of different batteries within the energy storage system. Combining the line power distribution proposed in the paper, battery charging and discharging dispatching, and the management of batteries in energy storage systems, is a worthwhile issue for comprehensive energy management and dispatching in microgrids.

Future research will explore how to determine the optimal capacity and location of distributed power sources under stable electricity loads, taking into account the degradation of batteries and the diesel generator's impact on the decision-making process of the approach. Additionally, as the scale of the microgrid increases, its computational complex-

ity will exhibit an exponential rise. In the next step, we will prioritize the consideration of balancing the scale of the microgrid with the efficiency of the approach.

Author Contributions: Conceptualization, S.C.; methodology, S.C. and Z.C. (Zhenwei Cui); visualization, Z.C. (Zhenwei Cui); investigation and software, Z.C. (Zhiyu Chen) and J.L.; supervision, H.W. and W.X.; writing—review and editing, S.C., J.L. and W.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: The authors are grateful to Yueming Guo from Kunlun Digital Technology Co., Ltd. and Kunqi He from the School of Information Engineering, China University of Geosciences (Beijing) for their assistance in managing and organizing data, as well as providing resources for our research. Their contributions are greatly appreciated.

Conflicts of Interest: Shuai Chen, Jian Liu, Zhenwei Cui and Hua Wang were employed by the company Kunlun Digital Technology Co., Ltd. Zhiyu Chen was employed by the company State Grid Information & Telecommunication Branch. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

- Ji, Y.; Wang, J.; Xu, J.; Li, D. Data-driven online energy scheduling of a microgrid based on deep reinforcement learning. *Energies* **2021**, *14*, 2120. [[CrossRef](#)]
- Zhou, B.; Zou, J.; Chung, C.Y.; Wang, H.; Liu, N.; Voropai, N.; Xu, D. Multi-microgrid energy management systems: Architecture, communication, and dispatching strategies. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 463–476. [[CrossRef](#)]
- Liu, W.; Shen, J.; Zhang, S.; Li, N.; Zhu, Z.; Liang, L.; Wen, Z. Distributed secondary control strategy based on Q-learning and pinning control for droop-controlled microgrids. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 1314–1325. [[CrossRef](#)]
- Arwa, O.; Folly, A. Reinforcement learning techniques for optimal power control in grid-connected microgrids: A comprehensive review. *IEEE Access* **2020**, *8*, 208992–209007. [[CrossRef](#)]
- García-Triviño, P.; de Oliveira-Assís, L.; Soares-Ramos, E.P.P.; Sarrias-Mena, R.; García-Vázquez, C.A.; Fernández-Ramírez, L.M. Supervisory Control System for a Grid-Connected MVDC Microgrid Based on Z-Source Converters With PV, Battery Storage, Green Hydrogen System and Charging Station of Electric Vehicles. *IEEE Trans. Ind. Appl.* **2023**, *59*, 2650–2660. [[CrossRef](#)]
- Pu, X.; Lin, H.; Jin, J.; Lin, Q.; Wu, K.; Liu, F. Nergy management method of energy local network with distributed energy storage based optimal energy flow. *J. Guangxi Univ. (Nat. Sci. Ed.)* **2020**, *45*, 284–297.
- Lin, W.; Chen, F.; Deng, H.; Shao, Z. Tube Model Predictive Control based Optimal Dispatching of a Multi-energy Microgrid. In Proceedings of the 2022 7th Asia Conference on Power and Electrical Engineering, Hangzhou, China, 15–17 April 2022; pp. 880–884.
- Zografou-Barredo, N.-M.; Patsios, C.; Sarantakos, I.; Davison, P.; Walker, S.L.; Taylor, P.C. MicroGrid resilience-oriented dispatching: A robust MISOCP model. *IEEE Trans. Smart Grid* **2020**, *12*, 1867–1879. [[CrossRef](#)]
- Zhong, X.; Zhong, W.; Liu, Y.; Yang, C.; Xie, S. Optimal energy management for multi-energy multi-microgrid networks considering carbon emission limitations. *Energy* **2022**, *246*, 123428. [[CrossRef](#)]
- Xiao, H.; Pu, X.; Pei, W.; Ma, L.; Ma, T. A Novel Energy Management Method for Networked Multi-Energy Microgrids Based on Improved DQN. *IEEE Trans. Smart Grid* **2023**, *14*, 4912–4926. [[CrossRef](#)]
- Li, Z.; Zang, C.; Zeng, P.; Yu, H. Combined two-stage stochastic programming and receding horizon control strategy for microgrid energy management considering uncertainty. *Energies* **2016**, *9*, 499. [[CrossRef](#)]
- Chen, S.; Jiang, C.; Li, J.; Xiang, J.; Xiao, W. Improved Deep Q Network for User-Side Battery Energy Storage Charging and Discharging Strategy in Industrial Parks. *Entropy* **2021**, *23*, 1311. [[CrossRef](#)] [[PubMed](#)]
- Paudyal, S.; Canizares, C.; Bhattacharya, K. Optimal operation of distribution feeders in smart grids. *IEEE Trans. Ind. Electron.* **2011**, *58*, 4495–4503. [[CrossRef](#)]
- Shuai, H.; He, H. Online dispatching of a residential microgrid via Monte-Carlo tree search and a learned model. *IEEE Trans. Smart Grid* **2020**, *12*, 1073–1087. [[CrossRef](#)]
- Gao, G.; Wen, Y.; Tao, D. Distributed energy trading and dispatching among microgrids via multiagent reinforcement learning. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 10638–10652. [[CrossRef](#)] [[PubMed](#)]

16. Gao, S.; Xiang, C.; Yu, M.; Tan, K.T.; Lee, T.H. Online optimal power dispatching of a microgrid via imitation learning. *IEEE Trans. Smart Grid* **2021**, *13*, 861–876. [[CrossRef](#)]
17. Dumas, J.; Dakir, S.; Liu, C.; Cornélusse, B. Coordination of operational planning and real-time optimization in microgrids. *Electr. Power Syst. Res.* **2021**, *190*, 106634. [[CrossRef](#)]
18. Zhang, J.; Li, Z.; Wang, B. Within-day rolling optimal dispatching problem for active distribution networks by multi-objective evolutionary algorithm based on decomposition integrating with thought of simulated annealing. *Energy* **2021**, *223*, 120027. [[CrossRef](#)]
19. Li, S.; Zhou, X.; Guo, Q. Research on microgrid optimization based on simulated annealing particle swarm optimization. In Proceedings of the E3S Web of Conferences, Shanghai, China, 16–18 August 2019; Volume 118, p. 01038.
20. Wei, Q.; Liu, D.; Lewis, F.L.; Liu, Y.; Zhang, J. Mixed iterative adaptive dynamic programming for optimal battery energy control in smart residential microgrids. *IEEE Trans. Ind. Electron.* **2017**, *64*, 4110–4120. [[CrossRef](#)]
21. Li, J.; Chen, S.; Jiang, C.; Liu, F. Adaptive dynamic programming approach for micro-grid optimal energy transmission dispatching. In Proceedings of the 2020 the 39th Chinese Control Conference, Shenyang, China, 27–29 July 2020; pp. 6190–6195.
22. Zhu, H.; Wang, H.; Zhang, D.; Dai, W.; Wu, T. Optimal Dispatching of Micro-Grid Multi-Energy System Considering Two-Dimensions Price-Based Demand Response. In Proceedings of the 2021 6th Asia Conference on Power and Electrical Engineering, Chongqing, China, 8–11 April 2021; pp. 572–577.
23. Liu, W.; Zhan, J.; Chung, C.Y.; Li, Y. Day-ahead optimal operation for multi-energy residential systems with renewables. *IEEE Trans. Sustain. Energy* **2019**, *10*, 1927–1938. [[CrossRef](#)]
24. Chen, L.; Wu, J.; Wu, F.; Tang, H.; Li, C.; Xiong, F. Energy flow optimization method for multi-energy system oriented to combined cooling, heating and power. *Energy* **2020**, *211*, 118536.
25. Basu, K.; Chowdhury, S.; Chowdhury, S. Micro-grid: Energy management by loss minimization technique. *Int. J. Energy Environ.* **2011**, *2*, 267–276.
26. Jiang, Y.; Yang, Y.; Tan, S.-C.; Hui, S.Y.R. Power loss minimization of parallel-connected distributed energy resources in DC microgrids using a distributed gradient algorithm-based hierarchical control. *IEEE Trans. Smart Grid* **2022**, *13*, 4538–4550. [[CrossRef](#)]
27. Liu, Z.; Liu, S.; Li, Q.; Zhang, Y. Optimal Day-ahead Dispatching of Islanded Microgrid Considering Risk-based Reserve Decision. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 1149–1160. [[CrossRef](#)]
28. Braz Pontes, L.; Percy Molina Rodriguez, Y.; Luyo Kuong, J.; Espinoza, H.R. Optimal allocation of energy storage system in distribution systems with intermittent renewable energy. *IEEE Lat. Am. Trans.* **2021**, *19*, 288–296. [[CrossRef](#)]
29. Sun, J.; Hu, C.; Liu, L.; Zhao, B.; Liu, J.; Shi, J. Two-stage correction strategy-based real-time dispatch for economic operation of microgrids. *Chin. J. Electr. Eng.* **2022**, *8*, 42–51. [[CrossRef](#)]
30. Zeng, P.; Li, H.; He, H.; Li, S. Dynamic energy management of a microgrid using approximate dynamic programming and deep recurrent neural network learning. *IEEE Trans. Smart Grid* **2019**, *10*, 4435–4445. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.