



Article

Joint Optimization of Energy Storage Sharing and Demand Response in Microgrid Considering Multiple Uncertainties

Di Liu ¹, Junwei Cao ^{2,*} and Mingshuang Liu ³¹ Department of Automation, Tsinghua University, Beijing 100084, China; kfliudi@163.com² Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China³ Shenzhen Tencent Computer System Co., Ltd., Shenzhen 518057, China; mshuangliu@tencent.com

* Correspondence: jcao@tsinghua.edu.cn

Abstract: Energy storage (ES) is playing an increasingly important role in reducing the spatial and temporal power imbalance of supply and demand caused by the uncertainty and periodicity of renewable energy in the microgrid. The utilization efficiency of distributed ES belonging to different entities can be improved through sharing, and considerable flexibility resources can be provided to the microgrid through the coordination of ES sharing and demand response, but its reliability is affected by multiple uncertainties from different sources. In this study, a two-stage ES sharing mechanism is proposed, in which the idle ES capacity is aggregated on the previous day to provide reliable resources for real-time optimization. Then, a two-layer semi-coupled optimization strategy based on a deep deterministic policy gradient is proposed to solve the asynchronous decision problems of day-ahead sharing and intra-day optimization. To deal with the impact of multiple uncertainties, Monte Carlo sampling is applied to ensure that the shared ES capacity is sufficient in any circumstances. Simulation verifies that the local consumption rate of renewable energy is effectively increased by 12.9%, and both microgrid operator and prosumers can improve their revenue through the joint optimization of ES sharing and demand response.



Citation: Liu, D.; Cao, J.; Liu, M. Joint Optimization of Energy Storage Sharing and Demand Response in Microgrid Considering Multiple Uncertainties. *Energies* **2022**, *15*, 3067. <https://doi.org/10.3390/en15093067>

Academic Editor: Dimitrios Katsaprakakis

Received: 22 March 2022

Accepted: 14 April 2022

Published: 22 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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: energy storage; demand response; deep reinforcement learning; multiple uncertainties; Monte Carlo sampling

1. Introduction

The microgrid (MG) is a subset of a power system with self-control capabilities, usually composed of distributed generators, loads, energy storage facilities, etc. [1]. Different from the distribution network, the MG can operate in islanded mode or grid-connected mode, and from the perspective of the upper-level system, the MG is an independent entity in the power system [2]. To deal with the problem of environmental pollution, renewable energy has developed rapidly in recent years. In an MG with high proportion of renewable energy, distributed renewable power generation devices, e.g., rooftop photovoltaics (PV), are often installed on the demand side, which turns consumers into prosumers with dual attributes of demand and supply. The uncertainty of renewable energy output brings certain challenges to the efficient operation of the MG.

Energy storage (ES) is considered to be an effective means to deal with the fluctuation of renewable energy power generation, and its installation is rapidly increasing around the world [3]. Since ES can be owned by microgrid operators (MGOs) or prosumers, in order to improve the utilization efficiency of ES, mechanisms for ES capacity sharing have been proposed, which are mainly divided into two modes, i.e., centralized ES sharing and distributed ES sharing.

In the centralized mode, ES is invested and operated by MGO or independent ES operators, and prosumers purchase the required ES capacity. Stackelberg game theory

is used to analyze the relationship among participants [4], and the market framework is designed to maximize the revenue of system [5]. An offline optimization approach for single MG equipped with ES is proposed in [6], and the energy cost of the conventional energy drawn from the main grid is minimized. A two-layer energy management system for MGs is proposed in [7], and ES is used to minimize the total operational cost, as well as deal with the uncertainty of renewable energy. Other similar studies can be seen in [8–11].

In the distributed mode, ES is owned by each prosumer, and the capacity can be shared by prosumers through incentives or transactions [12–14]. Affected by changes in power supply and demand, the shared ES capacity required by MGO in each time slot varies. Excessive shared ES capacity causes waste of capacity resources, while insufficient shared ES capacity affects the adjustment of MGO. However, due to the influence of uncertainty, both the shared ES capacity required by MGO and the ES capacity that prosumers can share fluctuate, and how to obtain an appropriate shared ES capacity in each time slot has not been well resolved in existing research.

Moreover, demand response (DR) is also recognized as an effective means to use the adjustable resources on the demand side to improve the flexibility of the MG, and it mainly includes two types: one is price-based (e.g., [15–18]), and the other is incentive-based (e.g., [19–23]).

In the area of price-based DR, time-of-use (TOU) is widely applied due to its stability, and the social costs can be reduced by utilizing the temporal complementarity of end-users [15,16]. Real-time price with higher flexibility is also the concern of many researchers, and the desirable usage behaviors are elicited through appropriate mechanisms and online optimization approaches [17,18].

Incentive-based DR can provide flexible schedulable resources for the system operator, which is conducive to the collaborative optimization of DR and other flexible resources [19]. Incentive-based DR is usually implemented during peak load periods, and consumers are directly subsidized according to their response [20–23]. Since the reduced load demand in DR includes transferable loads such as electric vehicles and delayable loads such as temperature control loads, DR will cause load rebound in subsequent time periods, which affects the cumulative revenues of the whole day. Some studies have paid attention to this problem and considered the load reflection phenomenon in the optimization, e.g., [24–27], but the uncertainty of load rebound caused by prosumer behavior has been ignored in these studies.

Although massive mechanisms and optimization strategies on DR and ES sharing have been proposed in the existing studies, the joint optimization considering their mutual influence and multiple uncertainties needs to be further investigated. On one hand, under the constraint of power balance, the change in shared ES power will not only affect its own marginal cost, but also change the reduced power and marginal cost in DR, and vice versa. On the other hand, the superposition of uncertainties existing in intraday optimization of MG and day-ahead ES sharing, respectively, brings a great challenge to the maintenance of power balance, as well as the maximization of MGO's revenues.

To solve optimization problems in complex environments, deep reinforcement learning (DRL) with strong learning ability can be applied, which has been proven to be efficient in some studies [28]. Considering that the behavior of the prosumer in DR and ES capacity sharing is difficult to accurately model, the model-free DRL algorithm should be applied, and in order to improve the utilization efficiency of historical datasets, the offline DRL algorithm, i.e., deep deterministic policy gradient (DDPG), is selected in this paper. Based on the replay buffer, the DDPG algorithm can realize the reuse of historical data, and has achieved good results in the optimization control of MG, such as battery charging control, motor control, voltage control, etc. [29–31].

In the scenario of this paper, MGO obtains the appropriate shared ES capacity through incentives, and the incentives need to be formulated on the previous day. However, the required ES capacity is affected by intraday optimization, which cannot be known in advance, i.e., MGO has to formulate incentives with incomplete information. Meanwhile,

both the required ES capacity and the ES capacity that can be shared by prosumers are affected by uncertainty, so the DDPG algorithm cannot be directly applied to solve the optimization problem in this paper.

In order to improve the utilization efficiency of distributed ES, a two-stage ES sharing mechanism based on incentives is proposed, in which MGOs can obtain the required ES capacity to reduce operating costs, while prosumers can revenue from the sharing of idle ES capacity. Then, a two-layer semi-coupled optimization strategy based on DDPG is proposed to solve the decision-making problem with incomplete information, and Monte Carlo sampling is applied to deal with the influence of uncertainty. The main importance and contributions are summarized as follows.

(1) A two-stage optimization framework is proposed to realize the cooperation of DR and ES sharing. Compared with the existing studies that only focus on DR, such as [20–23], or only focus on ES sharing, such as [12–14], joint optimization can more fully release the adjustable potential of resources on the demand side, so as to improve the revenues of MGO and the local consumption of renewable energy.

(2) Since the required ES capacity in the day-ahead ES sharing is determined by real-time optimization and cannot be known in advance, a two-layer semi-coupled optimization strategy based on DDPG is proposed to realize asynchronous optimization of coupled decision-making problems that are distributed in different time slots.

(3) Multiple uncertainties caused by prediction errors, prosumer behavior, etc., are considered as fully as possible. Differently from existing studies, such as [23–27], which ignore the uncertainty of load rebound in DR, Bayesian transition probability is introduced to describe the uncertainty caused by prosumer behavior.

(4) To deal with the impact of multiple uncertainties on ES capacity sharing, Monte Carlo sampling is applied in the network training of the proposed algorithm. Compared with the existing research that ignores the impact of uncertainty on ES sharing, such as [32–35], the optimization strategy proposed in this paper can ensure that sufficient shared ES capacity for real-time optimization can always be obtained at the lowest cost in any scenario.

The rest of this paper is organized as follows: Section 2 illustrates the framework of the system and introduces the sharing mechanism of distributed ES. The modeling of MGO and prosumers is presented in Section 3. Section 4 proposes the two-layer semi-coupled optimization strategy based on DDPG, and numerical simulation is given in Section 5. Finally, conclusions are drawn in Section 6.

2. System Architecture and Sharing Mechanism of Distributed Energy Storage

This section first introduces the system architecture for ES sharing and DR in MG, then explains the time sequence of actions in the cooperative scheduling.

2.1. System Architecture

As shown in Figure 1, this paper focuses on residential community MG, and the considered system consists of an MGO and different types of prosumers, e.g., residential, apartment or commercial buildings, etc., assuming that prosumers are equipped with rooftop PV and ES of different capacities. Meanwhile, the prosumers' load consists of adjustable loads (e.g., electric vehicles, water heaters, and air conditioners, etc.) and non-adjustable loads (e.g., lighting and refrigerators, etc.). All demand-side devices are controlled by smart terminals, i.e., prosumers can make optimization decisions based on changes in external signals. The daily power consumption of residential buildings is usually tens of kwh [36], so the daily power consumption of MG containing multiple buildings is usually hundreds of kwh. The renewable energy power source in residential communities is mainly rooftop PV, and in high-penetration residential communities, the maximum output of PV can reach several hundred kilowatts or even higher. With proper PV array arrangement, rooftop PV can be aggregated into a stable PV power source [37]. To store excess PV output, ES devices are equipped in homes within buildings. Since the

capacity of household ES devices is around 10 kWh [38], assuming that each household is equipped with ES devices, the total ES capacity in the MG can reach thousands of kWh.

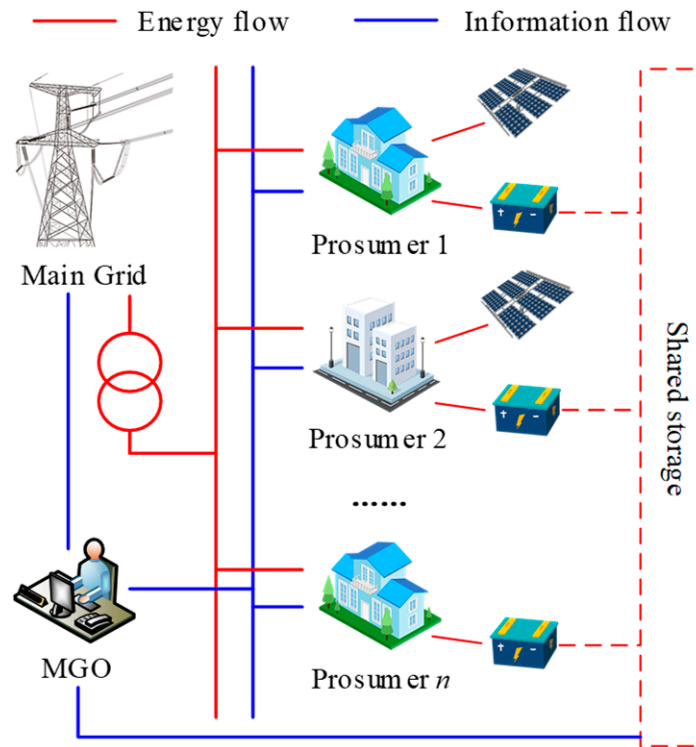


Figure 1. Energy storage sharing architecture.

MGO is an independent entity that obtains revenues by providing energy supply services to prosumers. Normally, MGO purchases electricity from the main grid with the real-time price (RTP) and sells it to prosumers with TOU tariff. When the RTP is higher than TOU, the sales of electricity to prosumers cause deficits to MGO, which can be alleviated by reducing load demand through DR. Moreover, MGO can aggregate prosumers’ idle ES capacity through incentives and improve its own revenues through proper charging/discharging operations.

2.2. Sharing Mechanism and Action Sequence

The variables that MGO needs to optimize include the incentive price for ES sharing, the charging/discharging power of the ES, and the incentive price for DR. The time sequence of the actions is shown in Figure 2.

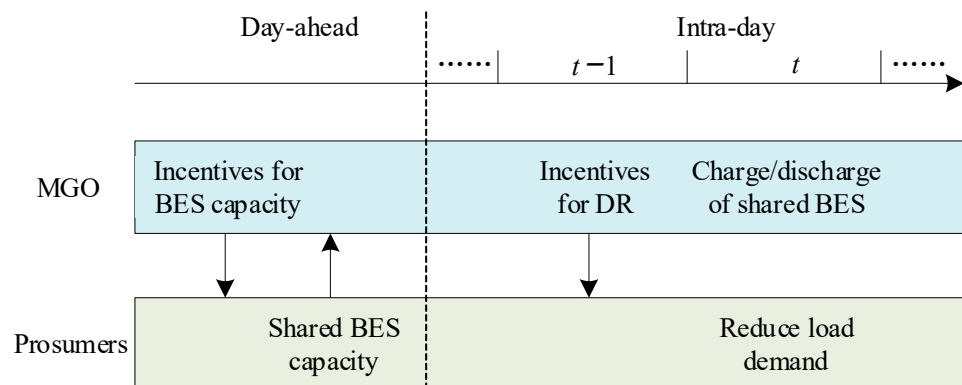


Figure 2. Time sequence of actions.

Assume that a day is divided into T time slots, and each time slot is indexed by $t \in \{0, 1, \dots, T\}$. As shown in Figure 2, on the previous day, MGO announces the incentive prices of ES sharing to prosumers, and each prosumer feeds back the shared ES capacity accordingly. It should be noted that the incentive price of each time slot varies with the required ES capacity. In time slot $t - 1$, MGO announces the incentive price of DR for the next time slot according to the prosumers' net load demand and the RTP in the main grid. Subsequently, in time slot t , MGO controls the charging/discharging power of the shared ES, and interacts with the main grid to purchase or sell power to maintain the power balance of the MG.

3. Modeling of Prosumer and MGO

3.1. Prosumer Modeling

We consider a set of prosumers $i = \{1, \dots, n\}$, whose load is provided by the MGO. Prosumers can obtain revenues by sharing their idle ES capacity. Charging/discharging reduces the cycle life of ES, and the cost per unit capacity increases with the depth of charge/discharge [39], i.e., prosumers have to bear the increasing marginal cost for sharing ES capacity. The cost of sharing ES can be expressed as

$$C_{i,t}^S = f_{i,t}(O_{i,t})O_{i,t}, \quad (1)$$

where $C_{i,t}^S$ is the total cost of the i -th prosumer in time slot t in ES sharing, $O_{i,t}$ is the capacity of the shared ES, and $f_{i,t}(\cdot)$ is a monotonically increasing function, representing the unit cost in the ES sharing.

The goal of prosumers participating in ES sharing is to maximize their own revenues. Therefore, the capacity of ES shared by prosumers under incentive λ_t^{CS} is:

$$\begin{aligned} O_{i,t}^* = \operatorname{argmax} & (f_{i,t}(O_{i,t})O_{i,t} - \lambda_t^{CS}O_{i,t}), \\ \text{s.t. } & O_{i,t}^* \leq O_{i,t}^I, \end{aligned} \quad (2)$$

where λ_t^{CS} is the incentive price for ES sharing in time slot t , $O_{i,t}^*$ is the shared ES capacity of the i -th prosumer in time slot t , and $O_{i,t}^I$ is the installed ES capacity of the i -th prosumer.

In the time slot where RTP is higher than TOU, MGO has to bear the deficit to provide electricity to prosumers, and it can reduce the prosumers' load demand through DR to alleviate the deficit. Load reduction brings economic revenues to prosumers, but unfortunately forces them to suffer the loss of comfort. According to economic theory, the relationship between the load reduction of prosumers and the DR incentive price can be expressed by demand function:

$$\Delta P_{i,t}^L = g_{i,t}(\lambda_t^L), \quad (3)$$

where $\Delta P_{i,t}^L$ is the amount of load reduction, λ_t^L is the DR incentive price in time slot t , and $g_{i,t}(\cdot)$ is a monotonically increasing function. It should be pointed out that due to the different characteristics of prosumers, $g_{i,t}(\cdot)$ may have different forms. Notably, the model-free DRL algorithms used in this paper can be adapted to different demand functions through learning.

Since the prosumers' adjustable load is composed of reducible load and transferable load, the reduction in the load causes changes in the load demand in the subsequent time slots [40], so the load of prosumers in each time slot is

$$P_{i,t}^{co,L} = P_{i,t}^L - P_{i,t}^{PV} + \sum_{j=0}^{t-1} \eta_{i,j,t} \Delta P_{i,j,t}^L, \quad (4)$$

where $P_{i,t}^{PV}$ is the output power of PV equipped by the i -th prosumer in time slot t , $P_{i,t}^{co,L}$ is the corrected load considering load rebound, and $\eta_{i,j,t}$ is the influence coefficient of load reduction in time slot j on the load of time slot t , which reflects the relationship between

load reduction and load rebound. Load rebound is the phenomenon that the load demand will be higher than the normal load during a period of time after DR, and it is mainly caused by transferable loads such as temperature control load, electric vehicle, etc. The power of load rebound is related to the depth of DR [24,25]. Considering that the uncertainty of prosumer behavior will affect the power of load rebound, Bayesian transition probability is introduced to describe the relationship of load rebound and DR:

$$\eta_{i,j,t} = \frac{\sum_{j=0}^{t-1} \sum_{m=0}^M p\left(P_{i,t}^L = P_{i,t-1}^L + \Delta P_{m,i,j}^L \mid \Delta P_{m,i,j}^L\right) \Delta P_{m,i,j}^L}{\sum_{j=0}^{t-1} \sum_{m=0}^M \Delta P_{m,i,j}^L}, \tag{5}$$

Assuming that there are M loads participating in load reduction in time slot j , $\Delta P_{m,i,j}^L$ is the power of the m -th reduced load, and the probability of m -th reduced load transferring to time slot t is $p\left(P_{i,t}^L = P_{i,t-1}^L + \Delta P_{m,i,j}^L \mid \Delta P_{m,i,j}^L\right)$.

3.2. MGO Modeling

In each time slot, MGO can control the charging/discharging power of the shared ES, and the dynamic transition of ES is as follows [41]:

$$SOC_t^E = SOC_{t-1}^E + \eta_t^c P_t^c \Delta t - \frac{1}{\eta_t^d} P_t^d \Delta t, \tag{6}$$

$$0 \leq SOC_t^E \leq \sum_{i=1}^N O_{i,t}^*, P_t^c \leq P_t^{c,max}, P_t^d \leq P_t^{d,max}, \tag{7}$$

where η_t^c and η_t^d are the charging/discharging efficiency of ES, respectively, P_t^c and P_t^d are the charging/discharging power in time slot t , respectively, and SOC_t^E is the energy stored in the shared ES and should not exceed the total shared ES capacity $\sum_{n=1}^N S_{i,t}^*$ in each time slot.

Due to the uncertainty of the prosumer behavior, the expected shared ES capacity that MGO can achieve by announcing the ES sharing incentive price λ_t^{CS} in each time slot is:

$$E\left(O_{i,t}^{sum}\right) = E\left(\sum_{i=1}^N \left(O_{i,t}^* + \varepsilon_{i,t}^{CS}\right)\right), \tag{8}$$

where E is the expectation of the function and $\varepsilon_{i,t}^{CS}$ is the error between the actual value and the theoretical value of the i -th prosumer's shared ES capacity.

Similarly, the prosumer's actual response under DR incentive price λ_t^L also has an error, and the expected response of prosumers in time slot t can be expressed as follows:

$$E\left(\Delta P_t^{L,sum}\right) = E\left(\sum_{i=1}^N \left(\Delta P_{i,t}^L + \varepsilon_{i,t}^{DR}\right)\right), \tag{9}$$

where $\varepsilon_{i,t}^{DR}$ is the error between the actual value and the theoretical value of the i -th prosumer's response in time slot t . Then, the total expected actual load demand of prosumers in each time slot is:

$$E\left(P_t^{L,sum}\right) = E\left(\sum_{i=1}^N P_{i,t}^{co,L}\right) - E\left(\Delta P_t^{L,sum}\right), \tag{10}$$

In each time slot, the power balance constraint has to be satisfied:

$$P_t^{L,sum} + P_t^S + P_t^c = P_t^P + P_t^d, \tag{11}$$

where P_t^P is the electricity purchased by MGO from the external grid during peak hours, and P_t^S is the electricity sold by MGO to the external grid during valley hours.

Due to the limited capacity of the transformer connecting the MG with the main grid, the power that MGO interacts with the main grid needs to satisfy the following constraints:

$$0 \leq P_t^P \leq P^{tr}, \quad 0 \leq P_t^S \leq P^{tr}, \quad (12)$$

where P^{tr} is the maximum transmission power of the transformer. The costs and revenues of MGO due to the power interaction between the MG and the main grid are as follows:

$$C_t^P = \lambda_t^P P_t^P, \quad U_t^S = \lambda_t^S P_t^S, \quad (13)$$

where C_t^P and U_t^S are the costs of purchasing power from the main grid and the revenues from selling power to the main grid, respectively. λ_t^P is the RTP of the external electricity wholesale market.

MGO sells electricity to prosumers in accordance with TOU tariffs. To encourage prosumers to install distributed PV, MGO needs to purchase all the excess PV power of prosumers:

$$C_t^I = \begin{cases} \lambda_t^{\text{TOU}} P_{i,t}^{\text{co,L}}, & P_{i,t}^{\text{co,L}} \geq 0 \\ \psi \lambda_t^{\text{TOU}} P_{i,t}^{\text{co,L}}, & P_{i,t}^{\text{co,L}} < 0 \end{cases}, \quad (14)$$

where C_t^I is the cost/revenue caused by the energy interaction with the prosumers, λ_t^{TOU} is the TOU tariff, ψ is the price coefficient of MGO purchasing prosumers' excess PV power, and in order to promote local consumption of PV power, $\psi \in [0, 1]$.

The total cost of the MGO in time slot t is as follows:

$$C_t^{\text{sum}} = C_t^I + C_t^P + \sum_{i=1}^N C_{i,t}^S + \sum_{i=1}^N \Delta P_{i,t}^L \lambda_t^L - U_t^S, \quad (15)$$

The optimization goal of MGO is to minimize the total cost of the whole period:

$$\max \sum_{t=0}^T C_t^{\text{sum}} \quad \text{s.t.} \quad (6) \sim (12). \quad (16)$$

4. Problem Formulation and Solution

In this section, we use the Markov decision process (MDP) to describe the decision-making actions of MGO, then introduce the proposed two-layer semi-coupled optimization strategy, as well as the method of applying Markov sampling in the reverse training of the neural network to deal with the influence of uncertainty.

4.1. Construction of Decision Problems Based on MDP

A standard Markov process consists of a set of 5 tuples, i.e., $M = (S, A, T, R, \gamma)$, where S the state space, A is the action space, T is the transition probability of the state after the action is executed, R is the reward for action, and γ is the discount factor.

The actions performed by MGO include incentive price for ES sharing λ_t^{CS} , the charging/discharging power of the ES P_t^c and P_t^d , and the incentive price for DR λ_t^L . Since the sharing of ES capacity is completed on the previous day, DR and power control of the shared ES are completed intraday, the action space can be divided into two subspaces:

$$a_t^{\text{DA}} = (\lambda_t^{\text{CS}}), \quad a_t^{\text{RT}} = (\lambda_t^L, P_t^c, P_t^d). \quad (17)$$

The purpose of action a_t^{DA} is to obtain sufficient capacity according to the required ES capacity $\hat{O}_t^{\text{target}}$ in each time slot, so the state is set to be $S_t^{\text{DA}} = (\hat{O}_t^{\text{target}})$. With the goal of maximizing revenues, the value of $\hat{O}_t^{\text{target}}$ is determined by many factors, e.g., RTP,

PV output, etc., and affected by multiple uncertainties, so it cannot be obtained directly. Through the reverse training of the two-layer semi-coupled network proposed in this paper and Monte Carlo sampling, the value of $\hat{O}_t^{\text{target}}$ can be obtained, which is discussed in detail in the next part. In order to satisfy the power balance constraint, the interactive power P_t^S and P_t^P with the main grid are set to be passive variables, and their values are calculated according to the power balance constraint.

In the intraday optimization, MGO needs to comprehensively consider RTP, load demand, PV output power, shared ES capacity, etc., to maximize the cumulative revenues throughout the day. The state space for action a_t^{RT} is:

$$S_t^{\text{RT}} = (P_t^L, P_t^{\text{PV}}, \lambda_t^P, \lambda_t^{\text{TOU}}, \psi). \quad (18)$$

Although the goal of MGO's optimization is to maximize the revenues throughout the day, the sub-goals of the actions a_t^{DA} and a_t^{RT} are different. On the previous day, MGO's goal is to minimize the cost for ES sharing while ensuring that the shared ES capacity is not less than the target ES capacity $\hat{O}_t^{\text{target}}$. Therefore, the reward is set as follows:

$$r_t^{\text{DA}} = -\sum_{i=1}^N C_{i,t}^S - \varphi \sqrt{\max(\hat{O}_t^{\text{target}} - O_t, 0)}, \quad (19)$$

where O_t is the shared ES capacity traded on the previous day.

The reward consists of two parts. The first is the cost for ES sharing, and the latter is the penalty for the shortage of shared ES capacity. When the shared ES capacity is insufficient, the operating efficiency of the MG is reduced, and even the power balance constraint may be violated. Therefore, the penalty can ensure that there is always sufficient shared ES capacity for real-time adjustment. Meanwhile, in order to ensure that the value of the penalty does not grow too fast, the difference between $\hat{O}_t^{\text{target}}$ and O is squared, and a correction coefficient φ is added to make its value match the previous item.

In the real-time adjustment, the goal of the actions is to maximize the revenue with the premise of satisfying the constraints, so the reward of the actions is set as

$$r_t^{\text{RT}} = -C_t^{\text{sum}} - D_t^S - D_t^{\text{tr}}, \quad (20)$$

where D_t^S and D_t^{tr} are penalties for violation of constraints (7) and (12), respectively:

$$D_t^{\text{SOC}} = \begin{cases} \max(D_t^S - O_t, 0), & D_t^S \geq 0 \\ -D_t^{\text{SOC}}, & D_t^S < 0 \end{cases}, \quad (21)$$

$$D_t^{\text{tr}} = \left| \max(P_t^S, P_t^P) - P^{\text{tr}} \right|. \quad (22)$$

Then, the cumulative rewards are:

$$R_t^{\text{DA}} = \sum_{k=t}^T \gamma_{\text{DA}}^{k-t} r_t^{\text{DA}}, \quad R_t^{\text{RT}} = \sum_{k=t}^T \gamma_{\text{RT}}^{k-t} r_t^{\text{RT}}, \quad (23)$$

where γ_{DA} and γ_{RT} are discount factors.

4.2. Two-Layer Semi-Coupled Optimization Strategy Based on DDPG

MGO needs to perform actions on the previous day and within the day, respectively. However, the target ES capacity $\hat{O}_t^{\text{target}}$ is unknown during the day-ahead action, such that the reward caused by the action cannot be evaluated. Inspired by the learning and memory capabilities of neural networks, a semi-coupled two-layer network based on DDPG is proposed to solve this problem, in which the actor-critic networks are established for the day-ahead action and the intra-day actions, respectively, and Monte Carlo sampling is

introduced in the training process to deal with the influence of multiple uncertainties, as shown in Figure 3.

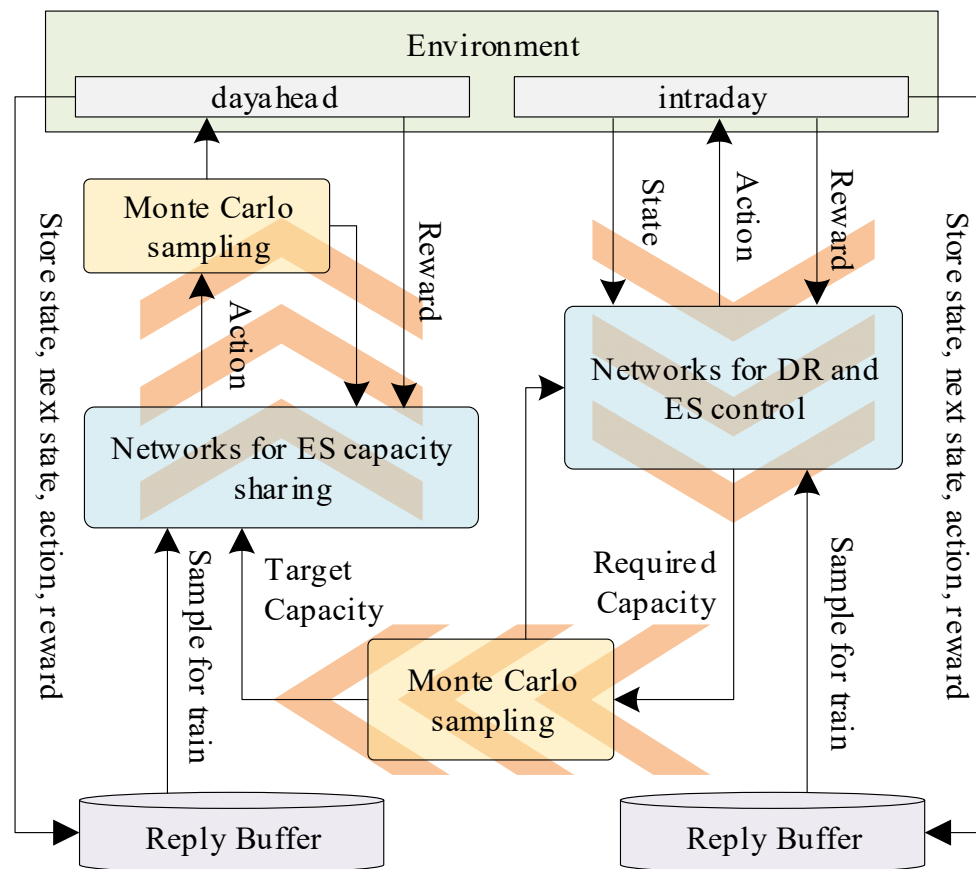


Figure 3. Two-layer semi-coupled network based on DDPG.

Since DDPG is an offline learning algorithm, although the actions are executed in a time sequence in practice, in the training of networks, the intraday networks for DR and ES control can be trained first. The shared ES capacity is unknown in advance, so we first make the following assumptions, which are guaranteed to be established by the day-ahead action.

Assumption 1. *The required ES capacity in each time slot can always be satisfied.*

With Assumption 1, the networks for ES capacity sharing can be trained. Since the actor–critic networks for day-ahead action and intraday actions are all based on DDPG, they have similar training processes, and the training processes are shown in the subsequent analysis. For the sake of simplicity, some subscripts are omitted.

To measure the performance of action a_t , we set the value function based on the Bellman equation as:

$$Q^\mu(S_t, a_t) = E[r_t + \gamma Q^\mu(S_{t+1}, \mu(S_{t+1}))], \tag{24}$$

where $\mu(S_{t+1})$ is the policy of action.

Let θ^Q and θ^μ be the parameters of critic network and actor network, respectively, and $\theta^{Q'}$ and $\theta^{\mu'}$ be the parameters of target critic network and target actor network, respectively. To train the critic network, define the loss function as

$$L(\theta^Q) = \frac{1}{H} \sum_i (y_i - Q(S_i, a_i | \theta^Q))^2, \tag{25}$$

where H is the number of samples from the reply buffer, and y_i is calculated using target network:

$$y_i = r_i + \gamma Q^{\mu'}(S_{i+1}, \mu'(S_{i+1} | \theta^{\mu'})) | \theta^{Q'} \quad (26)$$

Then, the policy gradient to train the action network is:

$$\nabla_{\theta^{\mu}} J \approx \frac{1}{H} \sum_i \nabla_a Q(S, a | \theta^Q) \Big|_{S=S_i, a=\mu(S_i)} \nabla_{\theta^{\mu}} \mu(S | \theta^{\mu}) \Big|_{S_i}, \quad (27)$$

where $\nabla_{\theta^{\mu}}$ is the gradient of θ^Q and ∇_a is the gradient of a .

Moreover, since DDPG is a deterministic strategy, random noise needs to be added when exploring the environment:

$$\tilde{\mu}(S) = \mu(S | \theta^{\mu}) + \mathcal{N}_t \quad (28)$$

where \mathcal{N}_t is random noise.

The parameters of the target networks are updated by copying:

$$\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}, \quad (29)$$

$$\theta^{\mu'} \leftarrow \tau \theta^{\mu} + (1 - \tau) \theta^{\mu'}, \quad (30)$$

where τ is the update rate coefficient, θ^Q and θ^{μ} are the parameter vectors of critic network and action network, respectively, and $\theta^{Q'}$ and $\theta^{\mu'}$ are the parameter vectors of target critic network and target action network, respectively.

The goal of day-ahead optimization is to provide sufficient ES capacity for intraday optimization at the minimum cost. The optimization problem can be divided into two sub-problems, i.e., the calculation of the aggregated ES target capacity $\hat{O}_t^{\text{target}}$ of each time slot, and the formulation of the ES shared incentive price λ_t^{CS} .

The uncertainty of PV output, prosumers' load and response all affect the value of $\hat{O}_t^{\text{target}}$, so Monte Carlo sampling is applied to obtain the expected value of $\hat{O}_t^{\text{target}}$. Sample K_E times in each training and uses the maximum value as the target ES capacity as follows:

$$\hat{O}_t^{\text{target}} = \max_j O_{t,j}, \quad j = 1, 2, \dots, K_E. \quad (31)$$

where $O_{t,j}$ is the required ES capacity in the j -th sample.

The ES capacity shared by prosumers under specific incentives also has uncertainty. In order to make the shared ES capacity always meet the needs of intraday adjustment, Monte Carlo sampling is also applied to determine the incentive price for ES sharing. In network training, for a certain action, i.e., the ES incentive price λ_t^{CS} , all shared ES capacity of prosumers are sampled with K_B times, and the minimum value is used to calculate the reward to ensure that Assumption 1 holds under the worst case. Then, the value of O_t in network training is as follows:

$$O_t(a_t^{\text{DA}}) = \min_j \sum_{i=1}^n O_{i,j,t}(a_t^{\text{DA}}), \quad j = 1, 2, \dots, K_B, \quad (32)$$

where $O_{i,j,t}(a_t^{\text{DA}})$ is the shared ES capacity in the j -th sample of the i -th prosumer with action a_t^{DA} . Then, the reward for training can be calculated using (19), and the network can be trained accordingly.

The detailed reverse training process of the two-layer semi-coupled network is presented in Algorithm 1.

Algorithm 1. The detailed reverse training process of the two-layer semi-coupled network.

Randomly initialize critic network Q_{RT} and actor μ_{RT} for DR and ES control with weight θ_{RT}^Q and θ_{RT}^μ , initialize critic network Q_{DA} and actor μ_{DA} for DR and ES control with weight θ_{DA}^Q and θ_{DA}^μ
 Initialize target network Q'_{RT} and μ'_{RT} for DR and ES control with weight $\theta'_{RT}^Q \leftarrow \theta_{RT}^Q$ and $\theta'_{RT}^\mu \leftarrow \theta_{RT}^\mu$, initialize target network Q'_{DA} and μ'_{DA} for DR and ES control with weight $\theta'_{DA}^Q \leftarrow \theta_{DA}^Q$ and $\theta'_{DA}^\mu \leftarrow \theta_{DA}^\mu$
 Initialize reply buffer R_{RT}, R_{DA}
 for episode = 1, M do
 Receive initial observation state S_1^{RT}
 for $t = 1, T$ do
 Select action a_t^{RT} according to (28), (29)
 Execute action a_t^{RT} and observe reward and the next state
 Store transition $(S_t^{RT}, S_{t+1}^{RT}, a_t^{RT}, r_t^{RT})$ in reply buffer R_{RT}
 Sample a random minibatch of H transitions from R_{RT}
 Update critic by minimizing (25)
 Update actor policy using (27)
 Update the target networks according to (30), (31)
 End for
 End for
 for episode = 1, M do
 Receive initial observation state according to (32)
 for $t = 1, T$ do
 Select action a_t^{DA} according to (28) and (29)
 Execute action a_t^{DA} and observe reward according to (33)
 Store transition $(S_t^{DA}, S_{t+1}^{DA}, a_t^{DA}, r_t^{DA})$ in reply buffer R_{DA}
 Sample a random minibatch of H transitions from R_{DA}
 Update critic by minimizing (25)
 Update actor policy using (27)
 Update the target networks according to (30) and (31)
 End for
 End for

5. Simulation Experiment

5.1. Settings of Simulation Environment

In order to verify the performance of the proposed algorithm, 180 days of data are used to perform the simulation with 140 days as the training set, 20 days as the validation set, and 20 days as the test set. Although the dataset contains data from different seasons, seasonal differences are not specially considered because the prosumer's load data and PV output data contain seasonal characteristics, which can be learned by the algorithm. The RTP is taken from the Pennsylvania–New Jersey–Maryland (PJM) electricity market. The power demand and PV output power are based on the real data in the PJM electricity market [42], but it is scaled down in proportion. In this paper, MG participates in the electricity market as an independent entity. In order to encourage local consumption of renewable energy, the price of surplus PV sold to the main grid is lower than the RTP [43], and the price coefficient ψ is set to be 0.5. The TOU tariff in Table 1 is set according to existing research results in [44]:

Table 1. TOU Tariff.

Time Slots	0:00–6:00 (Flat)	6:00–22:00 (Peak)	22:00–23:00 (Flat)
Price (USD/kWh)	0.02	0.04	0.02

The prosumers' response functions for ES capacity sharing incentive f_i and DR incentive g_i are both assumed to be quadratic functions, i.e., $O_i = \alpha_{CS}(\lambda_i^{CS})^2 + \beta_{CS}\lambda_i^{CS} + \zeta_{CS} + \varepsilon_{CS}$

and $\lambda_t^{\text{DR}} = \alpha_{\text{DR}}(\Delta P_t)^2 + \beta_{\text{DR}}\Delta P_t + \zeta_{\text{DR}} + \varepsilon_{\text{DR}}$, where the values of all parameters are shown in Table 2. The uncertainties in RTP, PV output, prosumers' load demand, prosumers' response to DR incentive, and response to ES capacity sharing incentive are all assumed to follow a normal distribution with mean 0 and standard deviation 0.03.

Table 2. Parameter settings.

Parameter	α_{CS}	β_{CS}	ζ_{CS}	α_{DR}	β_{DR}	ζ_{DR}	e_{CS}	σ_{CS}
Value	3×10^{-9}	1×10^{-7}	1×10^{-4}	5×10^{-7}	1×10^{-6}	1×10^{-3}	0	0.03
Parameter	σ_{PV}	P_c^{max}	P_d^{max}	e_{DR}	σ_{DR}	e_{L}	σ_{L}	e_{PV}
Value	0.03	200	200	0	0.03	0	0.03	0

In Table 2, e and σ are the expectation and standard deviation of each uncertain variable, respectively, and all uncertainties are calculated as follows:

$$X_U = X_O \times \left(1 + N(e, \sigma^2)\right) \quad (33)$$

where X_O denotes the original value without uncertainty, X_U denotes the value with uncertainty, and $N(e, \sigma^2)$ is a normal distribution with e as the expectation and σ as the standard deviation.

Assuming that the load rebound is affected by the past six time slots, and due to the influence of uncertainty, the load rebound coefficient follows the normal distribution with a standard deviation of 0.01, and the expected value of load rebound coefficient for each time slot is shown in Table 3:

Table 3. Expected value of load rebound coefficient.

Parameter	$\eta_{t-6,t}$	$\eta_{t-5,t}$	$\eta_{t-4,t}$	$\eta_{t-3,t}$	$\eta_{t-2,t}$	$\eta_{t-1,t}$
Value	0.01	0.02	0.03	0.05	0.1	0.2

In order to verify the advantages of the proposed method, four comparative cases are set up in the simulation experiment.

Case 1: The sharing and adjustment of distributed ES capacity is used to improve the revenues of MGO, without the consideration of DR; see, e.g., [12–14].

Case 2: DR is used to improve the revenues of MGO, but the idle capacity of prosumers' ES is not utilized; see, e.g., [20–23].

Case 3: The shared ES capacity of prosumers is fixed, i.e., the ES capacity aggregated by MGO in each time slot is the same; see, e.g., [10,11].

Case 4: The impact of multiple uncertainties is ignored in the sharing of ES capacity; see, e.g., [30–34].

All networks in the DRL adopt the fully connected network with three hidden layers, of which there are 256 neurons. The learning rate of the actor network and critic network for both sub-agents is set to be 0.0001 and 0.001, respectively. The algorithm is implemented using PyTorch 1.8.1 in Python 3.7.7. The case studies have been performed on a laptop with Intel(R) Core(TM) i7-9750H processor and one single NVIDIA GeForce GTX 1660 Ti GPU.

5.2. Performance Analysis of Intra-Day Joint Optimization

One day is selected for display in the 20 days of the test set. The RTP, TOU, load demand and PV output of the day are as follows:

Since PV power, prosumers' load, and RTP are all predicted values when performing actions, there is an error between them and actual values. The solid line in Figure 4 is the actual value of each parameter, and the shade is the fluctuation range of each parameter in 5000 Monte Carlo samples. It should be pointed out that the load fluctuations on the selected days are not large, but in the dataset, there are many days with large load fluctuations.

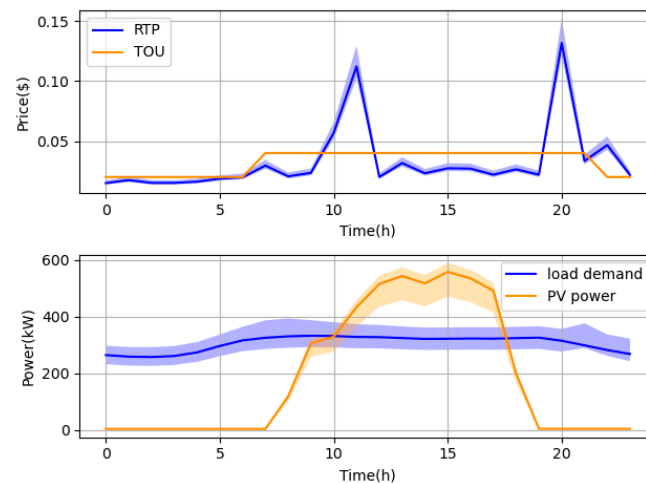


Figure 4. Price and power on selected day.

The power demand of the MG is first analyzed, and the results are as shown in Figure 5. The blue line is the PV power in the MG, the purple line is the original load demand in the MG, and the green, red and brown lines are the adjusted load demand in this paper, case 1 and case 2, respectively, including the prosumers' load demand and the charging/discharging power of shared ES. The original load demand fluctuates slightly throughout the day, but the PV output power varies greatly due to the influence of light intensity. Therefore, the PV output power in the MG from 10 to 17 o'clock is higher than the load demand, while there exists power shortages in the MG in other time slots.

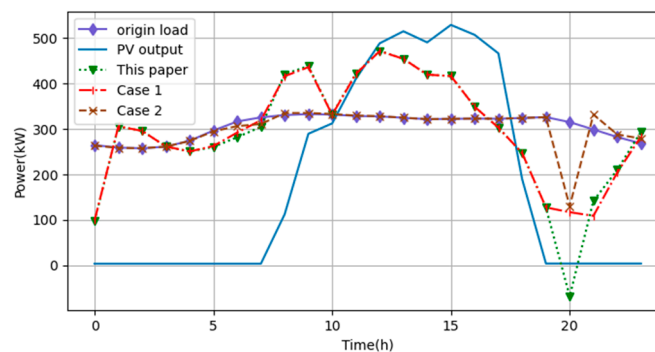


Figure 5. Changes in power demand in the microgrid before and after optimization.

In the optimization of this paper and case 1, the sharing of ES capacity is considered, and MGO can use the shared ES capacity to store excess PV output to increase the local consumption of PV. The total output of PV throughout the day is 4307.4 kWh. Before optimization, 73.6% can be consumed locally. After optimization using shared ES in this paper and case 1, the proportion of local consumption of PV output is increased to 86.5%. In case 2 where ES sharing is not considered, the local consumption rate of PV output is the same as that without optimization, indicating that the increase in the local consumption rate of PV output is mainly contributed by ES sharing.

Based on Figures 4 and 5, it can be seen that DR is mainly implemented in the time slots with higher RTP to reduce the deficit of MGO in the specific time slot, so as to improve the cumulative revenues of MGO throughout the day. Therefore, the revenues in each time slot are analyzed as shown in Figure 6:

Since the price of TOU is higher than that of RTP in most time slots, MGO can obtain revenues from power supply, while in other time slots with higher RTP, the revenue of MGO is negative, i.e., it has to bear the deficit to satisfy the energy demand of prosumers. Both the adjustment of ES in case 1 and DR in case 2 are effective means to reduce the deficit

and increase the total revenues of MGO. The total revenues of MGO throughout the day are USD 29.2 without optimization, while the total revenue increased by 74.0% to USD 50.8 in Case 1 and increased by 69.9% to USD 49.6 in Case 2. In the algorithm proposed in this paper, both DR and shared ES are considered, and the revenues of MGO are increased by 113.4% to USD 62.3. The results verify that the revenues of MGO can be further improved through the cooperation of DR and ES sharing compared to each method of them alone. Since the two methods have different effects on the revenues of each time slot of MGO throughout the day, the revenue improvement of each method compared with that without optimization is analyzed as follows:

The difference in revenue of MGO mainly appears in time slot 20 due to the extremely high RTP. Although the RTP in time slot 11 is also very high, the output of PV can satisfy the local load demand, so the change in RTP does not have a great impact on the revenues of MGO. Affected by the load rebound effect, the load demand in the subsequent time slots increases, thereby changing the revenues of MGO. After DR in time slot 20, load demand is increased in time slot 21, and since RTP is still higher than TOU in time slot 21, the revenues of MGO are reduced in Case 2. In comparison, ES sharing mainly affects the revenues of MGO in previous time slots. In the algorithm of this paper and Case 1, the revenues of MGO are reduced in time slots with high PV output, because MGO needs to pay for the shared ES capacity to store the exceeded PV power. Then, the stored power is used to satisfy the load demand of prosumers in time slots with high RTP, thereby improving the total revenues throughout the day.

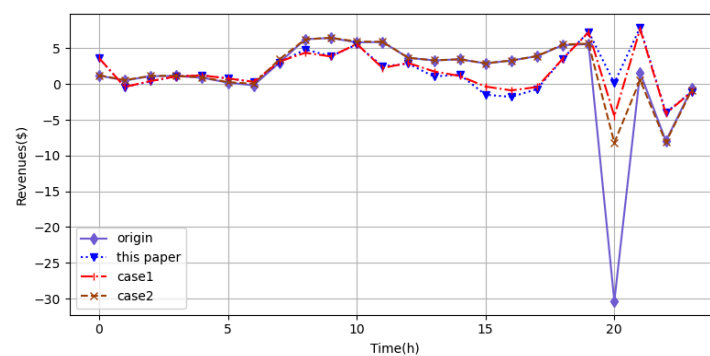


Figure 6. Revenues of MGO in each time slot.

In order to verify the stability of the optimization effect of the algorithm, the optimization effect of 10 consecutive days is counted in the test set, and the total revenue of these 10 days is shown in Table 4. It can be seen that the total revenue of MGO in these 10 days has reached USD 59.37, which is 29.30% and 9.18% higher than that of case 1 and case 2, respectively, indicating that the algorithm proposed in this paper has stable performance in continuous operation.

Table 4. Total revenues of MGO for 10 consecutive days.

Parameter	This Paper	Case1	Case2
Value	USD 59.37	USD 45.87	USD 54.38

5.3. Performance Analysis of Day-Ahead ES Capacity Sharing

However, MGO needs to pay for the shared ES capacity, so the change in required ES capacity and the corresponding power in each time slot are shown in Figure 7.

The unmarked solid line in Figure 7 is the value of each variable without the influence of uncertainty, the shadow is the fluctuation range of each variable in 5000 Monte Carlo samples, and the red solid line with the diamond mark is the upper limit of the required ES capacity, i.e., the target ES capacity in the day-ahead action.

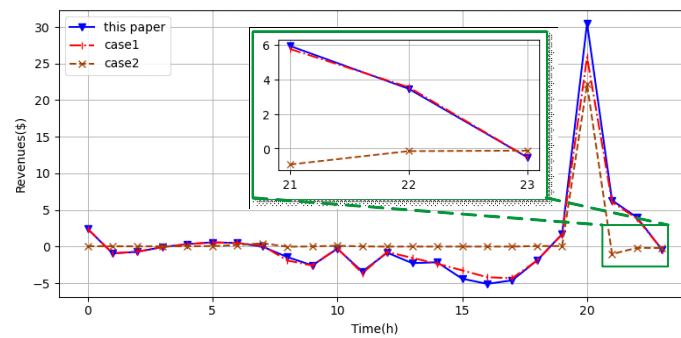


Figure 7. Improved revenues of MGO in each time slot.

As shown in Figure 8, in time slots with high PV output, the excess PV output is stored in the shared ES, and then the stored electricity is used to satisfy the load demand of prosumers in time slots with high RTP. As can be seen in Figure 5, the excess PV output is not completely stored, because the unit cost for ES sharing is increasing, and excess shared ES capacity leads to a decline in revenues. In addition to controlling the shared ES to absorb excess PV power, the algorithm in this paper can also track changes in RTP, store electricity when RTP is low, e.g., time slot 8 and 9, and supply power to prosumers in subsequent time slots, so as to further enhance the revenues of MGO.

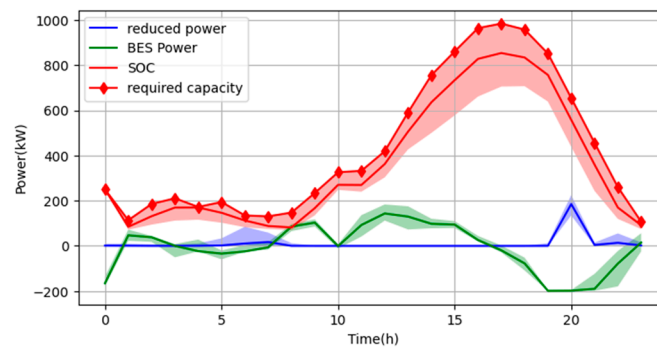


Figure 8. Power of shared ES and reduced load in DR.

The shared ES capacity in each time slot is also affected by uncertainty, and the results of ES sharing are as follows:

The solid line in Figure 9 represents the expected value of the shared ES capacity obtained by MGO, and the shaded part is the distribution interval of the ES capacity obtained by MGO in 5000 Monte Carlo samples. In Case 3, MGO predicts the maximum ES capacity required for the next day, and then aggregates shared ES based on this value. MGO does not need to make a decision for each time slot, which reduces the difficulty of ES sharing, but the shared ES capacity is not fully utilized, thereby reducing the revenues of MGO. In Case 4, the uncertainty in ES sharing has not been fully considered. Although its cost for shared ES capacity is the lowest, there may be insufficient ES capacity for intraday optimization. The Monte Carlo sampling process added in the algorithm proposed in this paper ensures that the shared ES capacity can always meet the needs of intraday optimization while minimizing the cost for shared ES capacity.

Prosumers can obtain revenues by sharing ES capacity, and the revenues in each time slot throughout the day are shown in Figure 10. Prosumers can obtain considerable revenues through ES capacity sharing, and the revenues are positively related to the shared ES capacity.

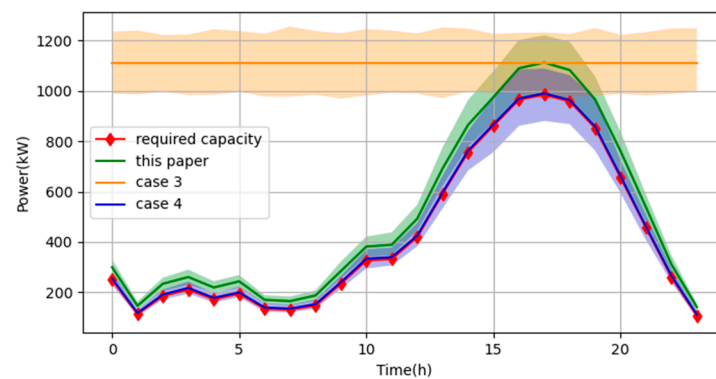


Figure 9. Results of ES capacity sharing.

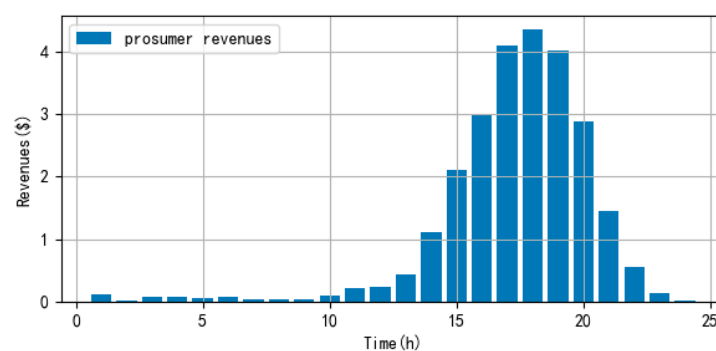


Figure 10. Revenues for prosumers by sharing ES capacity.

5.4. Performance Analysis of the Proposed Algorithm

The performance of Monte Carlo sampling determines whether the ES capacity required for intraday optimization can be met in the worst case. There are two independent Monte Carlo samples in the proposed algorithm. Because there are more uncertain factors when determining the required ES capacity, this sampling process is selected to analyze the impact of different sampling times on the determination of the required ES capacity, as shown in Figure 11:

In order to deal with the impact of multiple uncertainties and make the shared ES capacity satisfy the requirements of intraday optimization in the worst case, Monte Carlo sampling is applied to find the required ES capacity in the worst case. Too many Monte Carlo sampling times consume a lot of computing resources, and too few Monte Carlo sampling times may be difficult to accurately reflect the impact of uncertainty. Therefore, 10, 100, 1000 and 5000 Monte Carlo samples are set to verify the impact of sampling times on the required ES capacity. Each type of sampling is run 10 times, and the results are shown in Figure 10.

It can be seen from Figure 11a, due to too few sampling times, the boundary of the required ES capacity obtained by each group of Monte Carlo sampling is quite different. Moreover, compared with the groups of other sampling times, the upper boundary of the required ES capacity determined by the group with 10 sampling times is lower, which does not reflect the worst case. With the increase in the sampling times, the upper boundary of the required ES capacity determined by each group of samples is gradually stable. In Figure 11c,d, the upper boundary determined by each group of samples is relatively close, and the error is less than 35 kWh with the maximum required ES capacity is around 1000 kWh, indicating that the upper boundary of ES capacity required in the worst case can be stably found with sufficient sampling times.

The convergence and stability of the algorithm are important factors for evaluating its performance. This paper shows the loss convergence and reward changes of the network for DR and ES control, as shown in Figure 12.

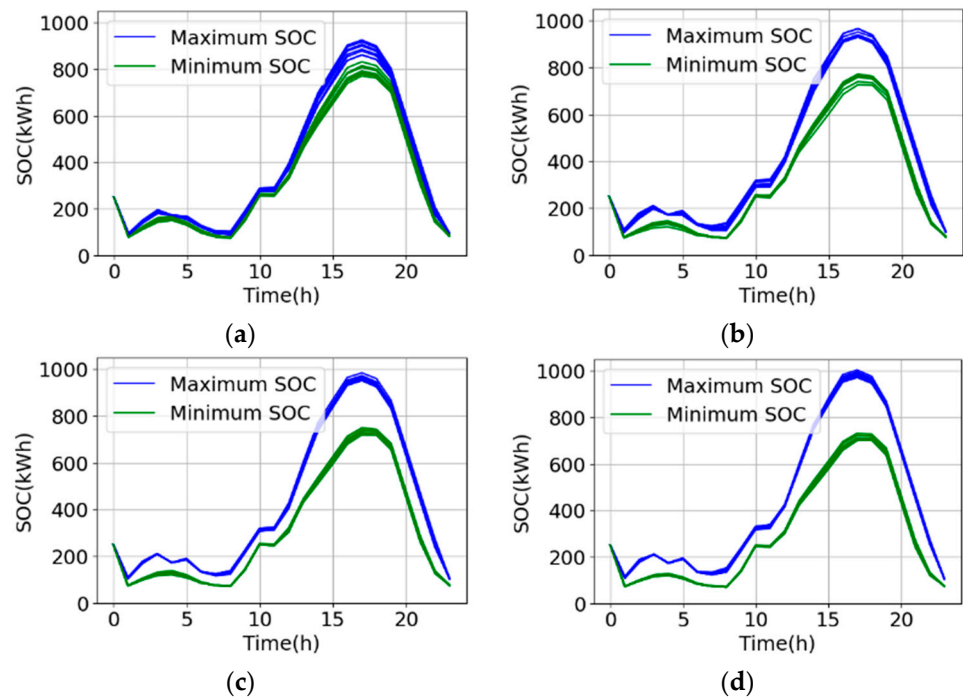


Figure 11. Effect of different Monte Carlo sampling times. (a) 10 samples. (b) 100 samples. (c) 1000 samples. (d) 5000 samples.

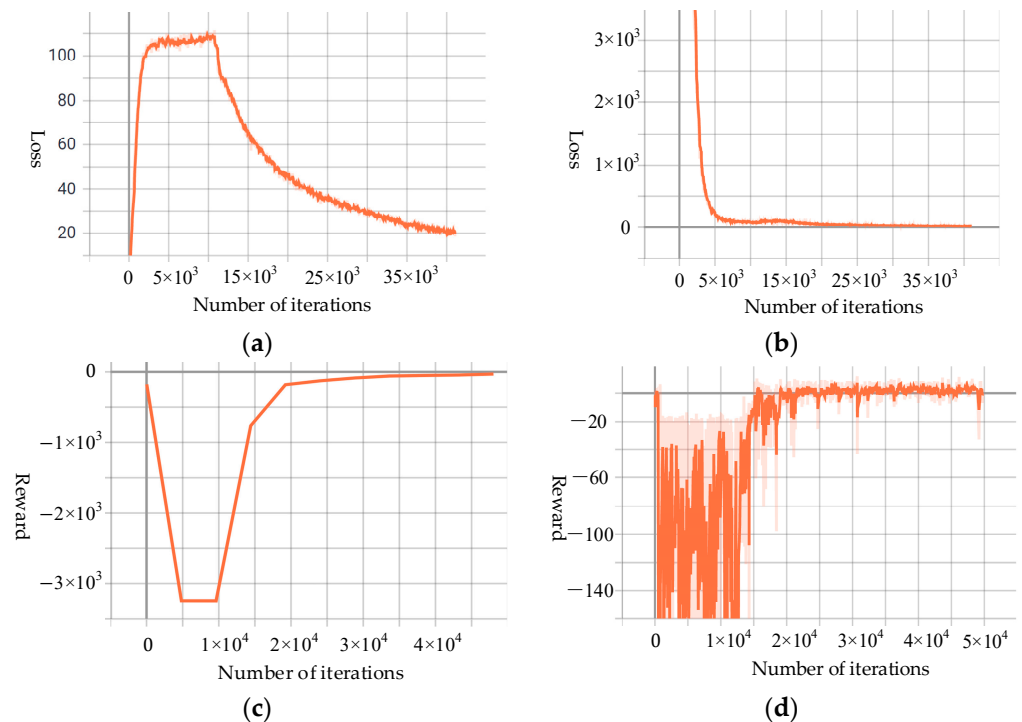


Figure 12. Effect of different Monte Carlo sampling times. (a) Actor loss. (b) Critic loss. (c) Mean reward. (d) Reward.

After about ten minutes of training, the network loss and reward of the algorithm tend to be stable, indicating that the algorithm has high training efficiency. It can be seen that the loss changes of actor network and critic network are relatively stable, and the reward can also converge smoothly after fluctuating at the beginning of training. This paper has tried several trainings, and most algorithms show similar convergence characteristics,

indicating that the algorithm has high stability. Increasing the learning rate can improve the convergence speed, but an excessive learning rate may lead to non-convergence. Therefore, in practical applications, a high learning rate should be used within an appropriate range. In addition, the depth and width of the network also affect the performance of the algorithm. A too-large network is not conducive to training, and a too-small network cannot meet the requirements of optimization. In practice, it is necessary to build an appropriate network according to the complexity of the optimization task.

6. Conclusions

This paper proposes a sharing mechanism for distributed ES and an optimization strategy to improve the social revenues and the local consumption of renewable energy through the collaboration of ES sharing and DR. Since the incentive for ES sharing, the incentive for DR, and the control of charge/discharge power of shared ES influence each other and are executed in different time slots, a two-layer semi-coupled optimization strategy is proposed to solve the asynchronous decision making of the coupling problem. Considering the influence of multiple uncertainties, Monte Carlo sampling is applied to ensure that MGO has sufficient shared ES capacity in any circumstances. Simulation shows that the local consumption of renewable energy can be effectively increased, and both MGO and prosumers can obtain revenues through ES sharing, thereby improving the social revenues of the MG. Simulation also shows that MGO can always obtain enough capacity at the lowest possible cost with the help of the proposed algorithm, which verifies the efficiency of the proposed algorithm.

In the future, the idle capacity of distributed ES can be aggregated into a virtual power plant (VPP) through sharing, and then provides flexible resources for the electricity market, or participate in peer-to-peer transactions to further improve the efficiency of distributed ES. Therefore, aggregation mechanisms and adjustment strategies need to be studied to ensure that VPP can provide reliable and sustainable adjustment capabilities at low cost.

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

Funding: This research was funded by SPECIAL RESEARCH PLAN FOR RHINOCEROS BIRD OF TENCENT BASIC PLATFORM TECHNOLOGY, grant number T102-TEG-2021110400001.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data of this paper came from the PJM power market, and the data acquisition website was: <https://www.pjm.com/> (accessed on 22 March 2022).

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

References

1. Morstyn, T.; Hredzak, B.; Agelidis, V.G. Distributed cooperative control of microgrid storage. *IEEE Trans. Power Syst.* **2015**, *30*, 2780–2789. [[CrossRef](#)]
2. Trivedi, R.; Khadem, S. Implementation of artificial intelligence techniques in microgrid control environment: Current progress and future scopes. *Energy AI* **2022**, *8*, 100147. [[CrossRef](#)]
3. Yan, X.; Gu, C.; Wyman-Pain, H.; Li, F. Capacity share optimization for multiservice energy storage management under portfolio theory. *IEEE Trans. Ind. Electron.* **2019**, *66*, 1598–1607. [[CrossRef](#)]
4. Mediawathe, C.P.; Shaw, M.; Halgamuge, S.; Smith, D.B.; Scott, P. An incentive-compatible energy trading framework for neighborhood area networks with shared energy storage. *IEEE Trans. Sustain. Energy* **2020**, *11*, 467–476. [[CrossRef](#)]
5. Szabó, D.Z.; Duck, P.; Johnson, P. Optimal trading of imbalance options for power systems using an energy storage device. *Eur. J. Oper. Res.* **2020**, *285*, 3–22. [[CrossRef](#)]
6. Rahbar, K.; Xu, J.; Zhang, R. Real-Time Energy Storage Management for Renewable Integration in Microgrid: An Off-Line Optimization Approach. *IEEE Trans. Smart Grid* **2015**, *6*, 124–134. [[CrossRef](#)]

7. Ju, C.; Wang, P.; Goel, L.; Xu, Y. A two-layer energy management system for microgrids with hybrid energy storage considering degradation costs. *IEEE Trans. Smart Grid* **2018**, *9*, 6047–6057. [[CrossRef](#)]
8. Wang, B.; Zhang, C.; Dong, Z.Y. Interval optimization based coordination of demand response and battery energy storage system considering SOC management in a microgrid. *IEEE Trans. Sustain. Energy* **2020**, *11*, 2922–2931. [[CrossRef](#)]
9. Du, Y.; Wu, J.; Li, S.; Long, C.; Onori, S. Hierarchical coordination of two-time scale microgrids with supply-demand imbalance. *IEEE Trans. Smart Grid* **2020**, *11*, 3726–3736. [[CrossRef](#)]
10. Walker, A.; Kwon, S. Analysis on impact of shared energy storage in residential community: Individual versus shared energy storage. *Appl. Energy* **2021**, *282*, 116172. [[CrossRef](#)]
11. Zhu, H.; Ouahada, K. A distributed real-time control algorithm for energy storage sharing. *Energy Build.* **2021**, *230*, 110478. [[CrossRef](#)]
12. Lai, S.; Qiu, J.; Tao, Y. Credit-based pricing and planning strategies for hydrogen and electricity energy storage sharing. *IEEE Trans. Sustain. Energy* **2022**, *13*, 67–80. [[CrossRef](#)]
13. Zhong, S.; Qiu, J.; Sun, L.; Liu, Y.; Zhang, C.; Wang, G. Coordinated planning of distributed WT, shared BESS and individual VESS using a two stage approach. *Int. J. Electr. Power Energy Syst.* **2020**, *114*, 105380. [[CrossRef](#)]
14. Hafiz, F.; de Queiroz, A.R.; Fajri, P.; Husain, I. Energy management and optimal storage sizing for a shared community: A multi-stage stochastic programming approach. *Appl. Energy* **2019**, *236*, 42–54. [[CrossRef](#)]
15. Yang, J.; Zhao, J.; Wen, F.; Dong, Z.Y. A framework of customizing electricity retail prices. *IEEE Trans. Power Syst.* **2018**, *33*, 2415–2428. [[CrossRef](#)]
16. Jacquot, P.; Beaudé, O.; Gaubert, S.; Oudjane, N. Analysis and implementation of an hourly billing mechanism for demand response management. *IEEE Trans. Smart Grid* **2019**, *10*, 4265–4278. [[CrossRef](#)]
17. Li, P.; Wang, H.; Zhang, B. A distributed online pricing strategy for demand response programs. *IEEE Trans. Smart Grid* **2017**, *10*, 350–360. [[CrossRef](#)]
18. Kim, S.; Giannakis, G.B. An online convex optimization approach to real-time energy pricing for demand response. *IEEE Trans. Smart Grid* **2017**, *8*, 2784–2793. [[CrossRef](#)]
19. Rana, M.; Rahi, K.; Ray, T.; Sarker, R. An efficient optimization approach for flexibility provisioning in community microgrids with an incentive-based demand response scheme. *Sustain. Cities Soc.* **2021**, *74*, 103218. [[CrossRef](#)]
20. Ellman, D.; Xiao, Y. Incentives to manipulate demand response baselines with uncertain event schedules. *IEEE Trans. Smart Grid* **2021**, *12*, 1358–1369. [[CrossRef](#)]
21. Muthirayan, D.; Kalathil, D.; Poolla, K.; Varaiya, P. Mechanism design for demand response programs. *IEEE Trans. Smart Grid* **2020**, *11*, 61–73. [[CrossRef](#)]
22. Chai, Y.; Xiang, Y.; Liu, J.; Gu, C.; Zhang, W.; Xu, W. Incentive-based demand response model for maximizing revenue of electricity retailers. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 1644–1650. [[CrossRef](#)]
23. Jindal, A.; Singh, M.; Kumar, N. Consumption-aware data analytical demand response scheme for peak load reduction in smart grid. *IEEE Trans. Ind. Electron.* **2018**, *65*, 8993–9004. [[CrossRef](#)]
24. Li, G.; Huang, Y.; Bie, Z. Reliability evaluation of smart distribution systems considering load rebound characteristics. *IEEE Trans. Sustain. Energy* **2018**, *9*, 1713–1721. [[CrossRef](#)]
25. Wei, C.; Xu, J.; Liao, S.; Sun, Y. Aggregation and scheduling models for electric vehicles in distribution networks considering power fluctuations and load rebound. *IEEE Trans. Sustain. Energy* **2020**, *11*, 2755–2764. [[CrossRef](#)]
26. Georges, E.; Cornélusse, B.; Ernst, D.; Lemort, V.; Mathieu, S. Residential heat pump as flexible load for direct control service with parametrized duration and rebound effect. *Appl. Energy* **2017**, *187*, 140–153. [[CrossRef](#)]
27. Wei, C.; Wu, Q.; Xu, J.; Sun, Y.; Jin, X.; Liao, S.; Yuan, Z.; Yu, L. Distributed scheduling of smart buildings to smooth power fluctuations considering load rebound. *Appl. Energy* **2020**, *276*, 115396. [[CrossRef](#)]
28. Hua, H.; Wei, Z.; Qin, Y.; Wang, T.; Li, L.; Cao, J. A review of distributed control and optimization in energy Internet: From traditional methods to artificial intelligence-based methods. *IET Cyber-Phys. Syst. Theory Appl.* **2021**, *6*, 63–79. [[CrossRef](#)]
29. Wei, Z.; Quan, Z.; Wu, J.; Li, Y.; Pou, J.; Zhong, H. Deep deterministic policy gradient-drl enabled multiphysics-constrained fast charging of lithium-ion battery. *IEEE Trans. Ind. Electron.* **2022**, *69*, 2588–2598. [[CrossRef](#)]
30. Nicola, M.; Nicola, C.-I.; Selișteanu, D. Improvement of PMSM sensorless control based on synergetic and sliding mode controllers using a reinforcement learning deep deterministic policy gradient agent. *Energies* **2022**, *15*, 2208. [[CrossRef](#)]
31. Zhang, J.; Li, Y.; Wu, Z.; Rong, C.; Wang, T.; Zhang, Z.; Zhou, S. Deep-reinforcement-learning-based two-timescale voltage control for distribution systems. *Energies* **2021**, *14*, 3540. [[CrossRef](#)]
32. Zhang, W.; Wei, W.; Chen, L.; Zheng, B.; Mei, S. Service pricing and load dispatch of residential shared energy storage unit. *Energy* **2020**, *202*, 117543. [[CrossRef](#)]
33. Hua, H.; Qin, Y.; Hao, C.; Cao, J. Stochastic optimal control for energy Internet: A bottom-up energy management approach. *IEEE Trans. Ind. Inform.* **2019**, *15*, 1788–1797. [[CrossRef](#)]
34. Alharbi, T.; Bhattacharya, K. Optimal scheduling of energy resources and management of loads in isolated/islanded microgrids. *Can. J. Electr. Comput. Eng.* **2017**, *40*, 284–294. [[CrossRef](#)]
35. Arghandeh, R.; Woyak, J.; Onen, A.; Jung, J.; Broadwater, R.P. Economic optimal operation of community energy storage systems in competitive energy markets. *Appl. Energy* **2014**, *135*, 71–80. [[CrossRef](#)]

36. Shcherbakova, A.A.; Shvedov, G.V.; Morsin, I.A. Power consumption of typical apartments of multi-storey residential buildings. In Proceedings of the 2020 International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), Moscow, Russia, 12–14 March 2020.
37. Satpathy, P.R.; Babu, T.S.; Shanmugam, S.K.; Popavath, L.N.; Alhelou, H.H. Impact of uneven shading by neighboring buildings and clouds on the conventional and hybrid configurations of roof-top PV arrays. *IEEE Access* **2021**, *9*, 139059–139073. [[CrossRef](#)]
38. Power Wall. Available online: <https://www.tesla.cn/powerwall> (accessed on 12 April 2022).
39. Shuai, H.; He, H. Online scheduling of a residential microgrid via Monte-Carlo tree search and a learned model. *IEEE Trans. Smart Grid* **2021**, *12*, 1073–1087. [[CrossRef](#)]
40. Muratori, M.; Rizzoni, G. Residential demand response: Dynamic energy management and time-varying electricity pricing. *IEEE Trans. Power Syst.* **2016**, *3*, 1108–1117. [[CrossRef](#)]
41. Hua, H.; Qin, Y.; Hao, C.; Cao, J. Optimal energy management strategies for energy Internet via deep reinforcement learning approach. *Appl. Energy* **2019**, *239*, 598–609. [[CrossRef](#)]
42. PJM. Real-Time Hourly LMPs. Available online: https://dataminer2.pjm.com/feed/rt_hrl_lmps (accessed on 21 March 2022).
43. Liu, N.; Yu, X.; Wang, C.; Li, C.; Ma, L.; Lei, J. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers. *IEEE Trans. Power Syst.* **2017**, *32*, 3569–3583. [[CrossRef](#)]
44. Zhao, X.; Gao, W.; Qian, F.; Ge, J. Electricity cost comparison of dynamic pricing model based on load forecasting in home energy management system. *Energy* **2021**, *229*, 120538. [[CrossRef](#)]