

Article

Decentralized Cooperative Power Dispatch Based on Multi-Agent Reinforcement Learning and Offline Digital Twin Technology for Building Integrated Photovoltaics and Energy Storage System Clusters

Qinwei Li ¹, Haowei Xing ^{2,3}, Han Zhu ^{1,*} and Zhengrong Li ¹ 

¹ College of Architecture and Urban Planning, Tongji University, Shanghai 201804, China; 18652735078@163.com (Q.L.); lizhengrong@tongji.edu.cn (Z.L.)

² The Architectural Design and Research Institute of Zhejiang University Co., Ltd., Hangzhou 310028, China; 0621059@zju.edu.cn

³ Center for Balance Architecture, Zhejiang University, Hangzhou 310058, China

* Correspondence: zhuhan@tongji.edu.cn

Abstract

Under carbon peaking and neutrality goals, building integrated with photovoltaics and energy storage system clusters (BIPECs) enable efficient on-site renewable energy use and can act as dispatch units for the public grid. However, BIPECs face significant uncertainties and are still under development. This study proposes a decentralized cooperative power dispatch model coupling a multi-agent proximal policy optimization (MAPPO) algorithm and offline digital twin (ODT) technology to optimize the photovoltaic (PV) power consumption of clusters despite limited data availability. An integrated BIPEC energy system model is established, and by leveraging the multi-agent system model of the BIPEC, the decentralized dispatch problem is converted into a fully cooperative multi-agent reinforcement learning (MARL) problem. A simulation-assisted ODT framework constructs a digital environment for MAPPO to augment data, conduct MAPPO training, and optimize the reward function, thereby obtaining power dispatch strategies. The results show that the proposed optimization model can obtain dispatch strategies that reflect a high degree of collaboration, reducing the cumulative power supply from the public grid by 0.55–2.56% per month compared to the non-cooperative self-generating and self-using strategy. This study presents the application of MARL in BIPECs by introducing a decentralized collaborative power dispatch methodology for building clusters, enhancing building energy efficiency and facilitating flexible collaborative power dispatch.

Keywords: decentralized power dispatch; photovoltaic system; energy storage system; multi-agent proximal policy optimization; digital twin



Academic Editors: Chi Yan Tso and Jianheng Chen

Received: 15 May 2026

Revised: 17 June 2026

Accepted: 18 June 2026

Published: 25 June 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Renewable energy resources (RERs) are green and efficient sources critical to achieving carbon neutrality [1,2]. RERs can be divided into distributed and centralized types, and building clusters play important roles in both forms. For distributed RERs, the building envelope can provide installation space for the generation system, and buildings consume power near the generation system to alleviate power transmission stress. Building integrated with photovoltaics (BIPV) is a widely adopted form [3], and an on-site energy storage system (ESS) can serve as an efficient buffer with a high return on investment to

balance power generation and consumption in the time domain [4]. Therefore, building integrated with photovoltaics and energy storage system (BIPE) can enhance the local consumption of distributed RERs. For centralized RERs, such as large-scale photovoltaic (PV) power generation systems, the output power exhibits high randomness, discontinuity, and instability. To enhance grid stability, building clusters, which account for approximately 40% of global final energy consumption [5,6], can participate in public power grid dispatch as demand response units for peak-load shifting. When several BIPEs form a cluster, referred to as a BIPEC, for cooperative dispatch, the load differences among buildings in the cluster can be fully utilized to further enhance dispatch flexibility. In summary, the BIPEC approach is both necessary and feasible for advancing carbon neutrality.

Owing to the large number and complex functions of the components in a BIPEC, optimal energy dispatch of the BIPEC must be studied. Centralized and decentralized control models exist for cooperative energy dispatch in integrated energy systems. In a centralized control model, a control center collects all unit information and determines their power transactions [7,8], whereas in a decentralized control model, each unit independently collects limited information that affects it and makes its own decisions. As the BIPEC approach is in a phase of rapid development, in such a complex and variable scenario, the decentralized dispatch control model performs better because of its strong flexibility and robustness [9,10].

Under a decentralized dispatch control model, the BIPEC power dispatch problem can be transformed into a fully cooperative multi-agent system (MAS) optimization problem [11]. The MAS-based approach treats each user in the cluster as an independent decision-making agent that shares information and experiences and learns from other agents. These agents make decisions independently to achieve individual or overall goals, such as improving the system operation economy and protecting user privacy [12,13].

Obtaining an optimal decentralized policy is a critical challenge in BIPEC power dispatch. Multi-agent deep reinforcement learning (MADRL) is a popular data-driven method that combines a deep reinforcement learning algorithm with MAS models [14]. MADRL can help a system consisting of multiple individuals optimize their decisions while comprehensively considering the collective interests of the whole system. MADRL has already been applied in decentralized dispatch in building energy systems [15]; therefore, introducing MADRL for decentralized cooperative power dispatch in BIPECs is feasible.

Among mature MADRL algorithms, multi-agent proximal policy optimization (MAPPO) as an on-policy method has gradually been used in various cooperative scenarios [16]. Wang et al. [17] proposed a deep reinforcement learning framework integrating a fully decentralized MAPPO algorithm with a Cubic Map neural network for drone-human collaborative data collection tasks in spatial crowdsourcing. Jia et al. [18] proposed a Lyapunov-based MAPPO algorithm with graph convolutional networks to maximize the average logarithmic data processing rate in the Internet of Vehicles system. Furthermore, MAPPO has been preliminarily applied to the management and dispatch of integrated energy systems. Zhang et al. [19] developed an energy-adaptive monitoring system for a solar sensor-powered smart animal farm, which integrates the MAPPO algorithm and an uncertainty maximization technique to resist multiple cyberattacks and enable high-quality data collection. Guo et al. [20] proposed a MAPPO algorithm incorporating the counterfactual baseline to optimize handover management and power allocation in a two-tier heterogeneous network, achieving higher system throughput and lower handover frequency. However, currently, limited studies have applied MADRL to distributed dispatch problems in integrated energy systems, and the appropriate MADRL method for the power dispatch problem specific to BIPEC requires further optimization.

Furthermore, when applied in engineering practice, all MADRL algorithms face the challenges of an accuracy gap [21,22], time or cost limitations [23,24], and insufficient security guarantees [25]. In particular, the training model of the MAPPO algorithm must be further adapted for the BIPEC, which currently lacks operational data and requires rapid deployment. To provide a reliable environment and sufficient datasets for acquiring, optimizing, and evaluating the dispatch strategy, digital twin (DT) technology can serve as a promising solution by mapping the physical space into a digital virtual space [26] and realizing data augmentation [27] for MAPPO training. Since this study lacks measured operational data of the BIPEC for real-time calibration, this study implements an offline DT (ODT) mode, which functions as a simulation-assisted training platform for data generation and policy optimization, rather than a fully synchronized real-time DT system.

In summary, for building cluster scenarios where each building is equipped with distributed photovoltaic generation and energy storage capabilities, limited research has optimized cluster power dispatch from the perspective of building energy consumption. Based on the above analysis, the main contributions of this paper are as follows.

- (1) A decentralized collaborative dispatch model for BIPECs based on MAPPO is established to achieve the optimization objective of minimizing the cumulative power supply from the public grid.
- (2) Simulation-assisted ODT technology is introduced to provide a digital training and experimental environment for MAPPO, enabling the acquisition and optimization of dispatch strategies. A five-dimensional DT framework is established to support future integration with real-time operational data during the operational stage.
- (3) A bidirectional optimization framework coupling dispatch strategy and ESS scheme based on ODT-MAPPO is proposed, enabling more comprehensive and effective BIPEC dispatch design.

The remainder of this paper is organized as follows. Section 2 establishes the power system model of the BIPEC and its corresponding MAS model, defining the power flow, communication flow, and constraints and proposing a decision-making process for power dispatch. Section 3 presents the design of an ODT-MAPPO model, in which the MAPPO algorithm and five-dimensional DT model for the BIPEC power dispatch problem and their coupling structure are analyzed. Section 4 presents a case study to demonstrate the feasibility and effectiveness of the developed approach. Finally, Section 5 summarizes the main conclusions.

2. BIPEC System Model

In this paper, the overall effect in a BIPEC is considered to be the goal of optimizing the internal power dispatch. In a BIPEC, each building is equipped with the PV power generation system and ESS, enabling it to act as a generator, consumer, and buffer for electricity. To realize the collaborative power dispatch of the BIPEC in a decentralized manner, the power exchange relationship and power flow within the system must be clarified. In response to low-carbon goals, economic games between buildings in the same BIPEC are not considered. This section establishes a BIPEC power system model, the corresponding MAS and the decentralized power dispatch decision-making process of a BIPEC based on the MAS.

2.1. BIPEC Power System Model

2.1.1. Composition of BIPEC System

Drawing on the energy models of integrated energy systems from the existing literature, including microgrids [28–30], integrated community energy systems [31–33], and

virtual power plants [34–36], the BIPE and BIPEC power system models used in this study are shown in Figure 1.

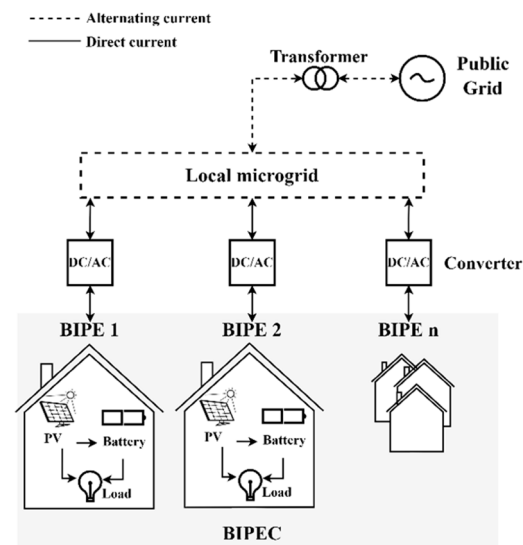


Figure 1. BIPEC system power schematic.

In this study, a rooftop PV system is adopted for BIPE as the distributed PV power generation system on each building side. This system does not require additional space and can reduce the impact on the appearance of the building. Moreover, the pitched roof of the existing building provides an ideal place for the tilt setting of the PV system, helping the PV modules obtain a longer sunshine duration and capture more solar radiation for cleaner electricity [37]. Consistent with BIPV systems, PV systems should be built as part of the roof envelope to reduce carbon emissions during the construction phase [38,39]. Additionally, it should be noted that the current PV model treats the photovoltaic system primarily as a rooftop power generation source and does not explicitly account for the architectural integration characteristics of BIPV in a strict sense.

Simultaneously, each BIPE is equipped with an independent ESS. Electrochemical ESSs, especially lithium-ion batteries (LIBs), possess high energy densities and fast response speeds and currently constitute a popular research topic. Rapid technological advances are being made in this area, and these systems serve as mainstream energy storage forms under the current new power system [40]. Therefore, the adoption of LIBs as the ESS for each BIPE in this study is reasonable and feasible.

Because the power output of the PV power generation system, the charging and discharging power of the battery, and the building electricity load are all in the form of direct current (DC), the power inside the building is DC. The power interaction between buildings in a BIPEC is realized through a local microgrid, and the current mature application of local microgrids is still based on alternating current (AC) power grids. Therefore, an AC microgrid is used in this study, which requires bidirectional DC/AC converters between each BIPE and the microgrid. Owing to the short distance between buildings and the low loss rate of AC transmission, the potential power storage and power buffer caused by the loss and time delay within the local microgrid are neglected. The power interaction between the BIPEC and public grid is also realized through the local microgrid. The power form is still AC; however, voltage conversion is required to realize cluster power dispatching and networking.

According to the control strategy, PV power generated in a building can be used for its own consumption, stored in its own battery, or delivered to the local microgrid for further distribution to other buildings or the public grid. The battery can be charged

by obtaining electricity from the PV power generation of a particular building or from the other buildings in the cluster. Considering the transmission loss, the battery and PV power generation systems of each building possess certain shared properties. However, the sharing process is not unrestricted, but rather also needs to balance the loss of power transmission and storage.

2.1.2. Power Balance Relationship and Constraints

The power balance of a BIPEC must consider three parts: the power balance inside the BIPE, power interaction between the buildings in the BIPEC, and interaction between the BIPEC and public grid. For the power balance inside a BIPE, the balance between the rooftop PV power generation, building electricity load, battery charging and discharging, and external interaction power must be established. The power balance inside a building can be expressed as:

$$E_{ci}(t) + E_{fi}(t) + E_i(t) = S_i(t) \quad (1)$$

where $E_{ci}(t)$ is the average charging power of the battery of building i at time step t . If $E_{ci}(t) > 0$, the battery is charged; otherwise, it is discharged. $E_{fi}(t)$ is the average net electric power output from building i to the local microgrid at time t . When $E_{fi}(t) > 0$, building i outputs power; otherwise, building i requests power from the local microgrid. $E_i(t)$ and $S_i(t)$ are the average electrical load power and average PV power generation of building i at time step t , respectively.

The charging and discharging actions of the battery at each time step are subject to the following constraint.

$$E_{ci,min} \leq E_{ci}(t) \leq E_{ci,max} \quad (2)$$

Two types of the loss are considered for the battery: self-discharge efficiency η_{SDC} and round-trip efficiency η_{RTE} . The relationship between η_{SDC} , η_{RTE} , state of charge (SOC), and charge or discharge power at each time step can be defined as:

$$SOC_i(t) = \frac{E_{ci}(t)\eta_{RTE}}{V_{max,i}} + SOC_i(t-1)\eta_{SDC} \quad (3)$$

where $V_{max,i}$ is the maximum capacity of the battery in building i .

The upper and lower limits of the SOC are determined by the depth of discharge (DoD).

$$SOC_i(t)_{min} \leq SOC_i(t) \leq SOC_i(t)_{max} \quad (4)$$

$$SOC_i(t)_{max} - SOC_i(t)_{min} = DoD(i) \quad (5)$$

The power interaction between buildings is based on a local microgrid. When buildings are connected to a local microgrid, AC/DC conversion is required, and a power balance exists in the building power network.

$$E'_{fi}(t) = \eta_{Converter}(i)E_{fi}(t) \quad (6)$$

where $E'_{fi}(t)$ is the average electrical power received by the local microgrid from building i at time step t .

The power interaction between the building cluster and public grid is also based on the local microgrid. Without considering the transmission loss and power delay in the microgrid, the power balance can be expressed as:

$$\sum_i^n E'_{fi}(t) + E'_g(t) = 0 \quad (7)$$

where n is the number of BIPEs in the cluster, and $E'_g(t)$ is the electric power obtained from the public grid in the microgrid.

Because the microgrid is a low-voltage AC network, whereas the public grid is a high-voltage AC network, transformers are required to adjust the voltage between the microgrid and public grid. The power balance between the microgrid and public grid can be expressed as follows.

$$E_g(t) = \eta_{Transform} E'_g(t) \quad (8)$$

where $E_g(t)$ is the power input from the public grid. When $E_g(t) > 0$, the public grid supplies power to the local microgrid and building cluster, and vice versa.

2.1.3. Objective Function

From the perspective of increasing PV power consumption, without adjusting the building energy load or considering the change in the loss ratio, the objective of cluster-level collaborative power dispatch is to minimize the cumulative electricity obtained from the public grid. The positive values of $E_g(t)$ in Equation (8) are summed to obtain the cumulative power transfer from the public grid over a period of time.

$$E_{in}(t) = \begin{cases} E_g(t), & \text{when } E_g(t) > 0 \\ 0, & \text{when } E_g(t) \leq 0 \end{cases} \quad (9)$$

$$\min Q = \sum_t E_{in}(t) \quad (10)$$

Under the above setting, there is no priority for the electricity consumption of all buildings in a BIPEC, and the influence of economic factors on the method of achieving optimal PV consumption is not considered. All buildings in a BIPEC are green, friendly and equal, which matches the development concept of green communities and green parks.

2.2. Fully Cooperative MAS Model for Dispatching

According to the power flow relationship within the cluster, the key to BIPEC power dispatch optimization is the decision of the battery charging and discharging power at each time step for each BIPE. Correspondingly, each BIPE must make effective decisions based on local observations during decentralized dispatch. An MAS for the BIPEC can be built to learn, acquire, and optimize the dispatch strategies for each BIPE.

The purpose of this study is to explore the optimal dispatch method for maximizing the overall PV power consumption of an entire cluster. All buildings are cooperative and have the same status; thus, a fully cooperative MAS model can be established for the BIPEC. If PV power generation and building energy consumption serve only as data inputs without control feedback, the dispatch action of each building is concentrated on the ESS. Therefore, the building power dispatch problem is decomposed into two steps: decentralized control of the building ESS and building power balance. Buildings serve as carriers that integrate the various energy systems. The distributed MAS structure shown in Figure 2 can be used to solve the BIPEC dispatch problem directly.

As shown in Figure 2, each BIPE has an autonomous control center corresponding to the building control agent (CA), which is responsible for receiving data and making power control decisions. The PV power generation agent (PVA) in a building is primarily used to predict the instant power generation of the building and send the data to the CA, whereas the load agent (LA) in the building is responsible for predicting the electricity load power of the building and sending the forecast data to the CA. The ESS agent (ESSA) monitors the status of the storage device and receives and executes the control commands issued by the CA. Only limited information is exchanged between buildings through each CA. Additionally,

a microgrid agent (MA) is introduced to monitor the power flow of the cluster, which does not have learning or decision-making functions. Each CA feeds back real-time power interaction information between the cluster and the grid. In essence, the CA of a BIPE is defined as an intelligent agent that makes action decisions and interacts with other CAs through the local microgrid and public power grid. All CAs in the cluster collectively constitute a fully cooperative MAS.

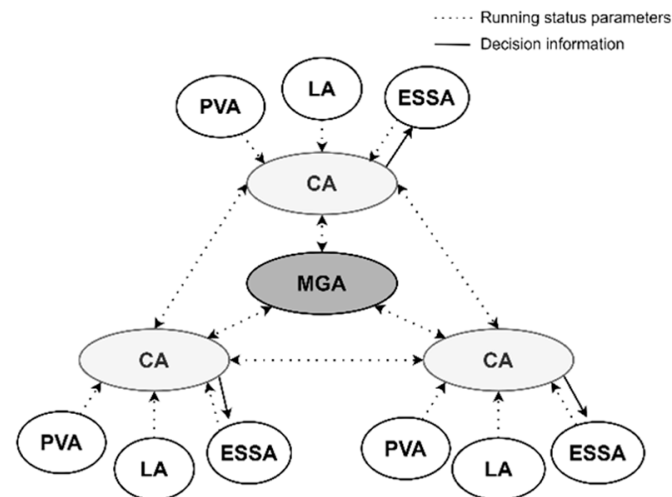


Figure 2. MAS structure for BIPEC power dispatch.

2.3. Energy Exchange Mechanism

The power generation and storage data in each building are transmitted in real time to the CA of each BIPE as part of the building properties. When participating in cluster power dispatch, each building primarily determines its outgoing power by controlling the storage equipment, which constitutes the interaction among the building, the microgrid, and the public power grid. In this study, for buildings belonging to the same owner, all buildings draw and supply power at equal priority, and all buildings can export power or obtain power from the outside; therefore, it is unnecessary to elaborate on the exact source or destination of power flows. The objective is to determine whether each building is exporting or receiving power from the outside at each moment, as well as the specific value of that power.

As a bridge connecting each building and the public power grid and realizing the power interactions, the local microgrid can serve as a “power pool” that accepts the “injection” and “export” of power from each building and the public grid. However, the microgrid has no storage capacity; therefore, the total net power received at each moment is zero, meaning the “water level” of the “pool” remains fixed at each moment. Furthermore, the platform does not consider delay, so “injection” and “export” are carried out simultaneously.

A local microgrid is used to balance the relationships both among the BIPEs within the cluster and between the BIPEC and the public grid. For the power dispatch of each building, only the net power output of the BIPE to the local microgrid needs to be determined to achieve the power dispatch of the cluster. For a certain BIPE, after determining the net power output of the building through power dispatch, the internal power balance of the building can be achieved using LIBs and Equation (1). The dispatch process is illustrated in Figure 3.

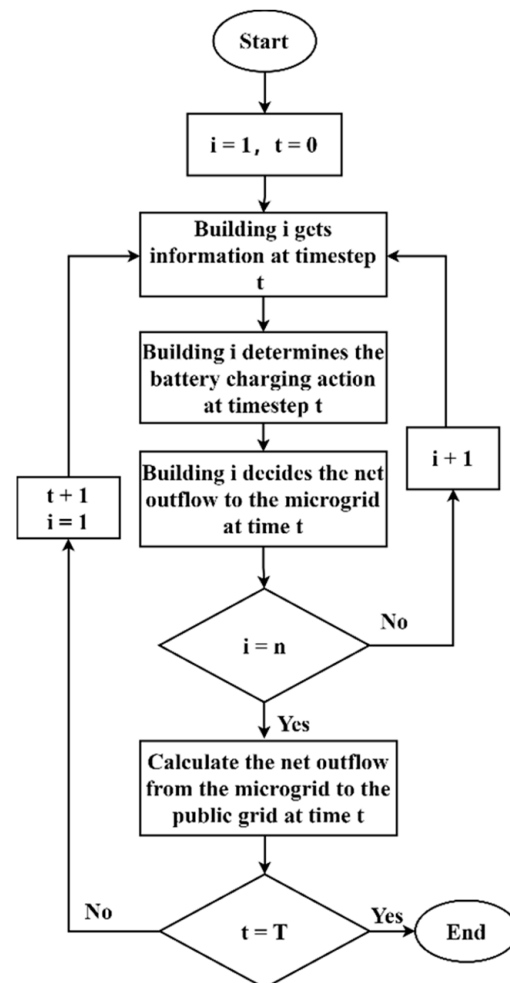


Figure 3. BIPE decision-making process for power dispatch.

3. ODT-MAPPO Power Dispatch Model for BIPEC

This section proposes a decentralized power dispatch strategy for BIPECs based on the MAPPO algorithm. Considering the sustainable development of BIPECs, DT technology is introduced to establish a coupled modeling framework with MAPPO. A five-dimensional DT framework is first established as the top-level architecture to support the full lifecycle of BIPEC dispatch, encompassing both offline training and online operation. Since measured and real-time data of the BIPEC are unavailable, this study implements the DT concept in an offline mode, which refers to as offline digital twin (ODT). Accordingly, an ODT-MAPPO model is developed for the training stage, in which the ODT functions primarily as a simulation-assisted training platform, while the operational stage involving real-time synchronization is presented as a future development direction.

3.1. Decentralized Power Dispatch Optimization Algorithm Based on MAPPO

To solve the power coordination dispatch problem of the abovementioned clusters, a cooperative MAPPO online optimization calculation [16] is adopted. The Markov decision process (MDP) is the foundational problem formulation for reinforcement learning, which can describe sequential decision-making problems. The dispatch problem can be formulated as an MDP, and the state at the next time step is determined by the action and state at the previous time step together with the environmental parameters at the current step, thus satisfying the standard MDP state transition property.

In the MAPPO algorithm, each BIPE is defined as an agent. Based on the BIPEC system model in Section 2.1, the factors influencing each BIPE's decision-making process are analyzed to define the state space of the algorithm. For a BIPE in a cluster, the factors affecting its power interaction with the external environment at each time step can be categorized into internal and partner factors. Internal factors include the PV power generation of the BIPE which determines its own power supply capacity, the building electricity demand, and the state of charge of its own energy storage system which determines its own power transfer capacity as a power buffer. Partner factors capture the status of other buildings in the cluster through publicly shared information. Additionally, whether the current day is a working day affects the building load profile and consequently influences peak shaving and load shifting performance. PV power generation and building electricity consumption are related to a specific moment in the day.

Therefore, for each agent, the state of agent i at time step t can be expressed as:

$$s_t^i = (S_i(t), E_i(t), SOC_i(t), holiday(t), t) \quad (11)$$

where $holiday(t)$ is a binary indicator (1 for holiday, 0 for workday), and t is the time value in the 24 h system. In this formulation, each agent encapsulates the electricity generation, consumption, and storage status of its corresponding building.

For each agent, the power dispatch is dynamically regulated by controlling the ESS to meet the demand. Correspondingly, the action of agent i is defined as:

$$a_t^i = \frac{E_{ci}(t)}{V_{max,i}} \quad (12)$$

which represents a continuous action corresponding to the percentage of each building's ESS charged or discharged at time step t .

Because dispatch is the result of the cooperation of all buildings in the cluster, all agents share the same objective and a common reward function. The objective is to minimize the cumulative power supply from the public grid during dispatch, which corresponds to maximizing the cumulative reward in MAPPO.

The MAPPO algorithm follows the Actor–Critic principle. During the training phase, all CAs share their own power states. The critic network can observe the global information and be updated in a unified manner, while each CA further updates its own actor network autonomously based on the critic network and its own state. Correspondingly, all CAs within a BIPEC communicate with each other and collaboratively learn the value network until stable and effective dispatch strategies are obtained. During the execution phase, each CA makes action decisions based solely on its own local observation s_t^i and its own actor network, without requiring any inter-building communication, thereby achieving fully decentralized dispatch.

MAPPO is a data-driven trial-and-error process, whereas BIPECs, in the process of rapid and continuous development, often require a reliable dispatch strategy when insufficient operational data are available. At present, several challenges remain in the process of obtaining a BIPEC power dispatch strategy through MAPPO training and deploying it in practice. (a) Limited reliability: Reinforcement learning (RL) training is highly dependent on the training environment. When a large difference exists between the training environment and the actual operation of power dispatch, the resulting dispatch strategy cannot cope with all situations encountered in the operation process. (b) Poor economic efficiency: The strategy acquisition process imposes requirements on both the training samples and the training environment, leading to high time and monetary costs. (c) Insufficient security: In the training process, communication and cooperation between

CAs and information sharing are required, which leads to information security problems. Moreover, artificial intelligence technologies, such as RL, can easily raise ethical or moral issues in decision-making, making it difficult to gain user trust. To address these challenges, the MAPPO training and power dispatch models must be optimized.

3.2. DT Five-Dimensional Framework for BIPEC

In this study, DT technology is used to build a simulation-assisted training environment in which MAPPO conducts trial-and-error learning for power dispatch. In this way, the iterative decision-making process that would otherwise occur in the real physical space is completed entirely in the digital space. Under this coupled form, MAPPO training does not rely entirely on real data collected in real time and has low requirements for the sample size of the actual power system operation data collection; trial and error in the actual power grid environment are not required, thus effectively reducing the training cost and improving the economy of strategy acquisition.

Referring to Qi et al. [41,42], combined with the system model of the BIPEC proposed in Section 2.1, a five-dimensional DT framework for the BIPEC is defined, as shown in Figure 4. Within this framework, the current study focuses on the ODT implementation, which is primarily used as a simulation-assisted data generation and policy-training platform to obtain training data when operational data are unavailable. The five-dimensional architecture is designed to support future integration with real-time operational data. In the actual operational stage, this framework can be used to implement online learning model iteration with updated running data.

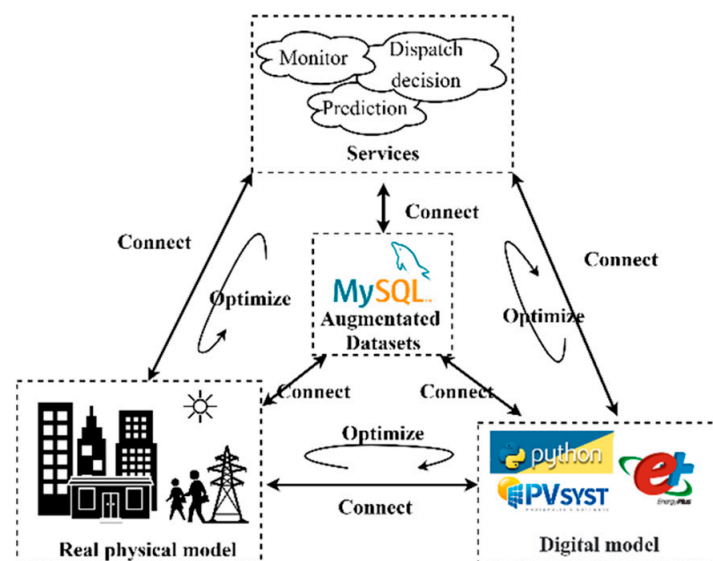


Figure 4. Schematic of BIPE digital twin 5D framework.

The dispatch of the building cluster in this study is based on the decentralized control mode of the MAS; therefore, the ODT model is constructed for each individual BIPE in the cluster, and the relevant data pertaining to building power generation, consumption, and storage are only transmitted among the component systems within the building. To optimize the power dispatch decision, the ODT model of each BIPE also needs to receive information about the interaction between the entire building group and power grid at each moment during the training stage. This information is obtained from the microgrid MA. Additionally, MySQL is used to store and retrieve the data, enabling offline operation of the platform system.

3.3. ODT-MAPPO Model for Optimal Dispatch Strategy

The ODT-MAPPO model, functioning as a simulation-assisted training platform in the current study, is designed to help the MAPPO algorithm obtain reliable and optimal dispatch policies for the BIPEC during the training process. Through simulation and data augmentation, the simulation-based ODT model is used to provide a digital environment for interaction, training, and verification, enabling the MAPPO algorithm to experience a wide range of possible operating situations during the training stage. Therefore, this framework strengthens the economy and reliability of the trained dispatch strategies.

According to the MAPPO framework for decentralized power dispatch defined in Section 3.1, the digital training environment of MAPPO needs to provide three components: the training state sample s_t^i , the instant state transfer, and the instant training reward feedback $R(t)$. An ODT-MAPPO training model for the power dispatch strategy training period is established, as shown in Figure 5. $S_i(t)$ and $E_i(t)$ in s_t^i are obtained based on the white-box model implemented in two mature simulation software packages: PVsyst 7.4.8 and EnergyPlus 23.2.0, respectively.

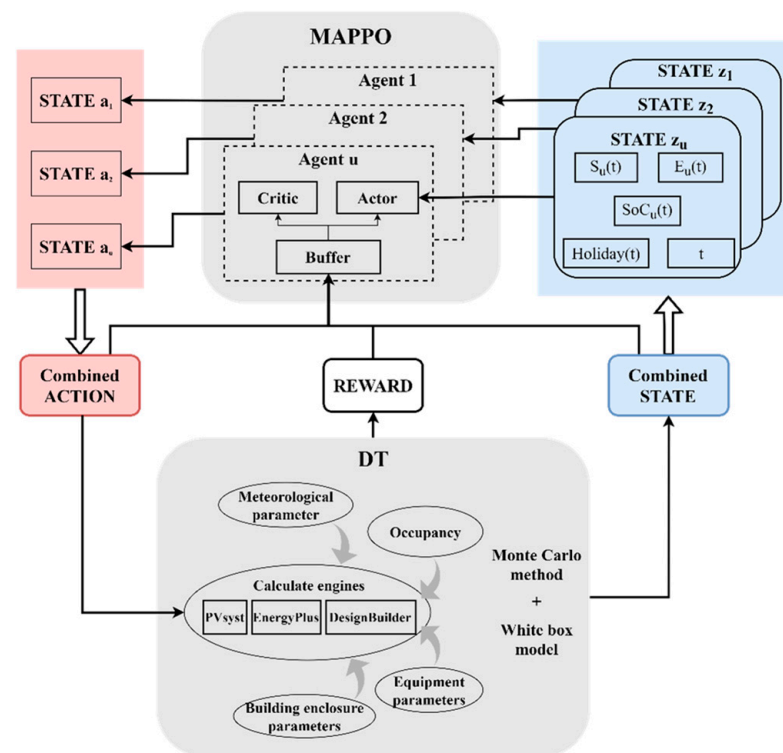


Figure 5. Structure of ODT-MAPPO coupling optimization model.

An example of data augmentation for building energy consumption data $E_i(t)$ is provided to illustrate the principle of training sample augmentation for a BIPEC. The occupancy rate distribution is one of the basic parameters of building energy consumption simulation calculations. It is the key factor that causes the changes in cold, heat, and humidity load of indoor air conditioning and further affects the state quantity $E_i(t)$. The distribution of the occupancy rate affects the use of equipment and lighting simultaneously.

The occupancy rate distribution usually follows certain rules, but randomness and uncertainty remain, which makes the building load different even under the same environmental conditions and on the same type of day. The existing indoor occupancy models include the fixed-schedule method, the random schedule method based on Monte Carlo simulation, and the time-autocorrelation random process model. However, the fixed-schedule method cannot provide a variety of occupancy schedules. The time-autocorrelation random

process model, which uses a Markov chain model as an example, is targeted at specific events. Since $E_i(t)$ corresponds to the electricity load of the entire building, the relevant events of personnel movement primarily comprise changes in occupancy status and arrival or departure events. However, most of the events occur randomly within 1 h; the load calculation time step in this study is 1 h, and the simulation time step is longer than the random event occurrence interval. Therefore, the time-autocorrelation random process model is also not applicable. Accordingly, a random schedule method based on Monte Carlo simulation is chosen to obtain a random occupancy schedule for buildings. Although this model cannot reflect the time autocorrelation or spatial correlation of personnel distribution in the building, it has no impact on the calculation of the total load of the building in this study and can ensure the diversity of the schedule. These random occupancy schedules are input into the white-box model of the building energy consumption to augment the training data.

Through the augmentation of the sample data, multiple groups of building energy consumption data for the same period of operation are obtained. When the augmented data are sufficient, multiple sets of energy consumption data at any time can represent all possible power consumption values and their distribution at that time. Therefore, 10 groups of state samples in the same month can be randomly selected, with the initial battery charging state of all BIPes set to $SOC_i(t)_{min}$. For any group of samples n , the average objective value in the month can be obtained based on the acquired strategy and state transfer law.

$$\bar{O} = \frac{\sum_{n=1}^{10} O_n}{10} \quad (13)$$

where O is the objective parameter, which can be Q .

Furthermore, the decision module and the monitoring module are designed for the operational stage of the DT framework; their implementation and validation require measured operational data and real-time data infrastructure.

The decision module is designed to enable each CA to collect real-time state information, including PV generation and building load powers through short-term forecasting models, battery SOC and public information of holidays and time indicators; this module feeds these inputs into the trained actor network to output the battery charging or discharging action at each time step. No inter-building communication would be required during this phase, as each CA operates solely on local observations. The monitoring module is intended to implement a deviation detection mechanism. Under future online operation, the DT would predict the expected battery SOC at the next time step and compare it with the actual measured value. If a significant or progressively increasing deviation is detected, this indicates battery parameter degradation, and the module issues an update warning, prompting recalibration of the battery model and retraining of the dispatch strategy.

Through these two modules, the DT framework is intended to establish multi-scale dynamic interaction with the MAPPO agents in future operational deployment. Compared with a conventional simulation environment, the proposed DT framework is designed to improve adaptability by establishing a bidirectional connection between the virtual model and the physical system, which is well recognized in the recent literature. Kabir et al. [43] surveyed DT for IoT-driven energy systems and concluded that DT enable real-time monitoring and adaptive management. Ugwu et al. [44] demonstrated that AI-driven DT for renewable energy grids can be continuously optimized through real-time data interaction and closed-loop feedback mechanisms. These capabilities motivate the five-dimensional DT architecture adopted in this study, of which the current ODT implementation realizes the offline training component.

4. Case Study and Data Experiments

A case study of a typical public building cluster in Shanghai is conducted to evaluate the proposed decentralized dispatch model through data experiments in three parts: data augmentation, strategy acquisition, and power dispatch. The objective is to verify the feasibility of the proposed decentralized dispatch strategy and assess its operational performance.

4.1. Basic Information

A cluster consisting of a hotel, a mall and an office building in Shanghai, China, is selected as the case study for decentralized dispatch. Because the accuracy of the prediction model is not the focus of this study, typical meteorological year data are adopted for the meteorological parameters in the ODT, and the occupancy schedules follow ASHRAE 90.1. PVsyst and EnergyPlus are used to simulate PV generation and building energy consumption, respectively. The dispatch duration is one month, using January (744 h) as an example, and the time step is set to 1 h. The maximum electrical storage capacities of the office building, hotel and mall are 520, 520, and 1200 kWh, respectively. For the rooftop PV systems of each building, based on available rooftop area and the typical meteorological year data for Shanghai, the PV modules are set at a fixed tilt angle of 30°, and the PV system capacities of the office building, hotel, and mall are 162, 108, and 1673 kWp, respectively. The upper and lower limits of the SOC are set to 100% and 0%, respectively.

Since the ESS adopts LIBs, this study sets $\eta_{SDC} = 99.99\%$, $\eta_{RTE}(i) = \begin{cases} 95\%, & \text{if } E_{ci}(t) \geq 0 \\ \frac{1}{95\%}, & \text{if } E_{ci}(t) < 0 \end{cases}$,

$$\eta_{Converter}(i) = \begin{cases} 97\%, & \text{if } E_{fi}(t) \geq 0 \\ \frac{1}{97\%}, & \text{if } E_{fi}(t) < 0 \end{cases}, \text{ and } \eta_{Transform} = \begin{cases} 98.5\%, & \text{if } E_g(t) \geq 0 \\ \frac{1}{98.5\%}, & \text{if } E_g(t) < 0 \end{cases}.$$

The actor and critic networks are constructed as fully connected neural networks. The learning rate is 5×10^{-4} , and the reward discount factor is 0.99. The simulations are performed in Python 3.10.0 on a computer equipped with an Intel Core i5 CPU and 16 GB RAM.

4.2. Results and Discussion

4.2.1. Data Augmentation Results and Analysis

To augment the building energy load data, the randomness of occupancy rate and corresponding lighting and equipment use are considered. According to Macdonald et al. [45], the random variation range of the occupancy schedule is usually $\pm 50\%$, following a normal distribution. The randomness is characterized by a jitter parameter ζ , and the occupancy rate at each time step can be expressed as:

$$r_{occ} = \zeta r_{occ, typical} \quad (14)$$

where $r_{occ, typical}$ is the occupancy rate corresponding to a typical schedule at each time step. Lighting and equipment use rates vary synchronously with occupancy rate through the same jitter parameter ζ .

When $\zeta > 0$, $r_{occ} > 1$ may occur, which corresponds to the real-time occupancy density being higher than the rated density. Occupancy can temporarily exhibit this phenomenon, but the lighting and equipment power cannot exceed their rated value; therefore, upper limits of the lighting and equipment use rate must be set. The corresponding use rates at each time step are:

$$r_{light} = \begin{cases} \zeta r_{light, typical}, & \text{when } \zeta r_{light, typical} \leq 1 \\ 1, & \text{when } \zeta r_{light, typical} > 1 \end{cases} \quad (15)$$

$$r_{ep} = \begin{cases} \zeta r_{ep,typical}, & \text{when } \zeta r_{ep,typical} \leq 1 \\ 1, & \text{when } \zeta r_{ep,typical} > 1 \end{cases} \quad (16)$$

Based on the typical schedules and jitter settings, numerous augmented schedules for occupancy, lighting, and equipment are generated. According to the principle of the Monte Carlo method, a large amount of generated data must be utilized to reflect the random distribution; therefore, Eppy 0.5.66 must be employed to batch-modify the white-box model established by EnergyPlus and generate building energy consumption results.

Using the above method, 10,000 groups of the occupancy rate, lighting use rate, and equipment use rate schedules for each building are input into EnergyPlus, yielding 10,000 groups of hourly building power data for an entire year. Figure 6 shows the hourly electricity consumption data for the three buildings after data augmentation on a certain working day.

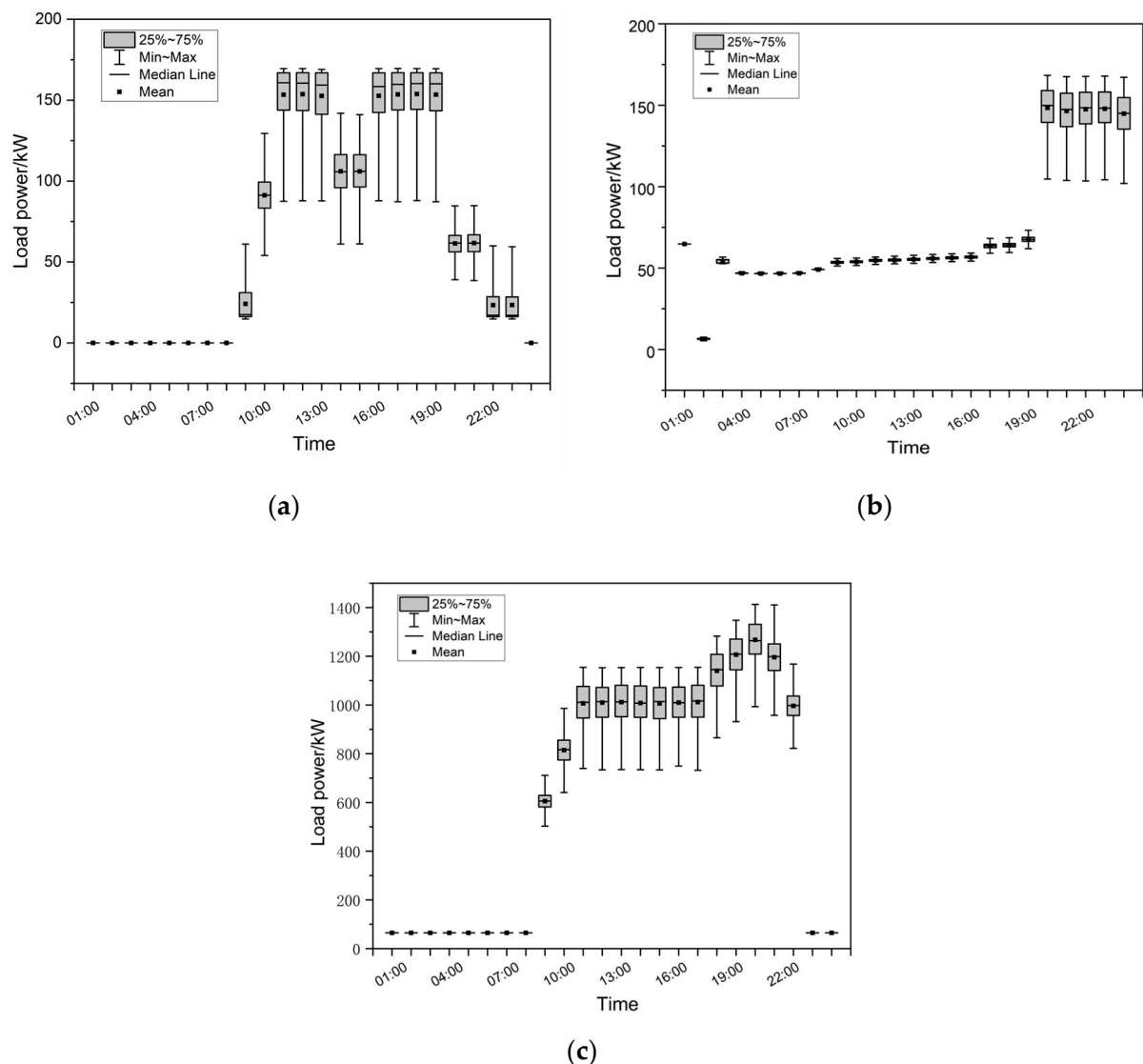


Figure 6. The hourly electricity consumption data for the three buildings after data augmentation on a certain working day: (a) Augmented office building hourly load power; (b) Augmented hotel hourly load power; (c) Augmented mall hourly load power.

After augmentation, the data distribution is more dispersed during the peak energy consumption period, mainly due to an increase in the number of indoor occupants. Most of the augmented data are concentrated within a range of 5–8% relative deviation from

the typical power consumption data; therefore, data augmentation must be introduced to energy use scenarios to make the training dataset as complete as possible.

4.2.2. MAPPO Training and Reward Setting Experiment Results

Using the augmented dataset combined with the ODT-MAPPO model illustrated in Figure 5, MAPPO training for decentralized power dispatch is implemented in Python. The time step for the power dispatch in this study is 1 h. Because the climate conditions of the same month in each year are similar and the monthly climate variations over a year follow certain rules, each training episode corresponds to one month, yielding 12 dispatch strategies for 12 months of the year. In practical operation, strategies can be selected according to a certain month.

The MAPPO algorithm is implemented using PyTorch 2.2.0. Each BIPE serves as an agent, and the parameters of the neural networks are consistent across the three agents. Both actor and critic networks consist of two layers, and the number of neurons in each layer is 64. Rectified linear unit functions are used as activation functions in both hidden layers, and the cumulative reward discount factor is set to $\gamma = 0.99$.

To help BIPEs stably obtain the maximum cumulative reward and thereby converge to the optimal dispatch strategy, experiments must be conducted on the digital platform provided by the ODT to determine appropriate learning rate and reward function settings. Taking January, which falls within the heating season, as an example, experiments are conducted to enhance the local monthly PV power consumption and optimize the grid response.

In an experiment to improve local PV power consumption, if the reward function is directly set as a single-value function of the objective function, a misleading result will be obtained. This result indicates that wasting more surplus PV power leads to a greater immediate reward. To address this issue, the reward function is reformulated by comprehensively considering power generation and electricity consumption as follows:

$$R(t) = -K_1 \sum_i^n (E_{fi}(t) - K_2(S_i(t) - E_i(t))) \quad (17)$$

where K_1 and K_2 are positive hyperparameters.

The MAPPO training results based directly on the objective function and the reward function formulated in Equation (17) are compared, as shown in Figure 7.

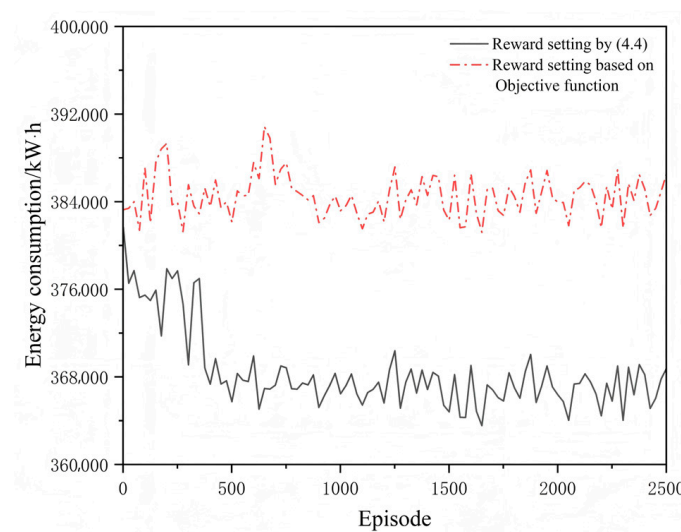


Figure 7. Results of MAPPO training process under different reward function settings.

Figure 7 demonstrates that when the objective function is directly used as the reward function, the training fails to reduce the cumulative power consumption from the public grid; the power consumption remains at approximately 384.2 MWh with no downward trend after more than 2000 training episodes. In contrast, the reward function formulated in Equation (17) steadily reduces the cumulative power consumption during training. Therefore, for an energy dispatch problem with long time steps and a complex energy structure, the reward function must be set by comprehensively considering generation and consumption to stimulate the cluster to reduce the cumulative grid supply electricity. Moreover, the current reward function does not include a penalty for battery cycling degradation. This is because the lithium-ion batteries adopted in this study have a high cycle life, and the 1 h dispatch time step inherently limits the switching frequency. To reduce model complexity, this study does not consider this factor.

Referencing the hyperparameter configurations in Yu et al. [16], the key training parameters are initially set and further fine-tuned on the ODT platform. Through the digital experiment, the total number of training steps is set to 10^6 , and $K_1 = 0.03$, $K_2 = 1$, learning rate $\text{lr} = 5 \times 10^{-4}$ and update coefficient $\varepsilon = 0.01$ are determined. The training results for January are presented in Figure 8, with the reward rising steadily and the cumulative power supply from the public grid in each episode declining steadily. The trends correspond to each other and both reach convergence after approximately 2500 episodes.

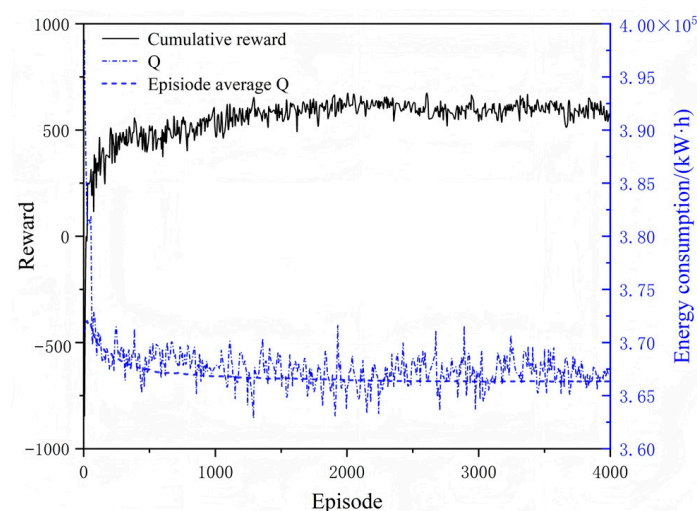


Figure 8. MAPPO power dispatch training return (cumulative reward within an episode) and cumulative power supply from the public grid in January.

The MAPPO training for one month's dispatch strategy requires approximately 48 h. For inference, generating the complete dispatch actions for one month takes approximately 4 min, corresponding to roughly 0.3 s per time step, which confirms the real-time feasibility of the proposed approach.

4.2.3. Comparison of the Dispatch Results Obtained Using Different Control Strategies

To evaluate the dispatch performance of the converged strategy, 10 validation datasets from January are tested based on the ODT. Two common priority strategies (cooperative and non-cooperative self-generating and self-using strategies) are developed, and a typical MAPPO strategy based on typical building energy consumption data without data augmentation is compared as a control group for ODT-MAPPO.

In the self-generating and self-using (SGSU) strategy, the PV generation of each building flows first to the load, then to the battery, and finally to the microgrid; correspondingly, the electricity load is met first by PV generation, then by the battery, and finally by the

microgrid. The difference between the cooperative SGSU (co-SGSU) strategy and the non-cooperative SGSU (noco-SGSU) strategy is whether the surplus PV power of each building can be shared with other buildings in the cluster before flowing into the public grid.

The hourly SOC operation results of the ODT-MAPPO and co-SGSU strategies for four consecutive days in January are shown in Figure 9, where the first two days are holidays and the last two days are working days.

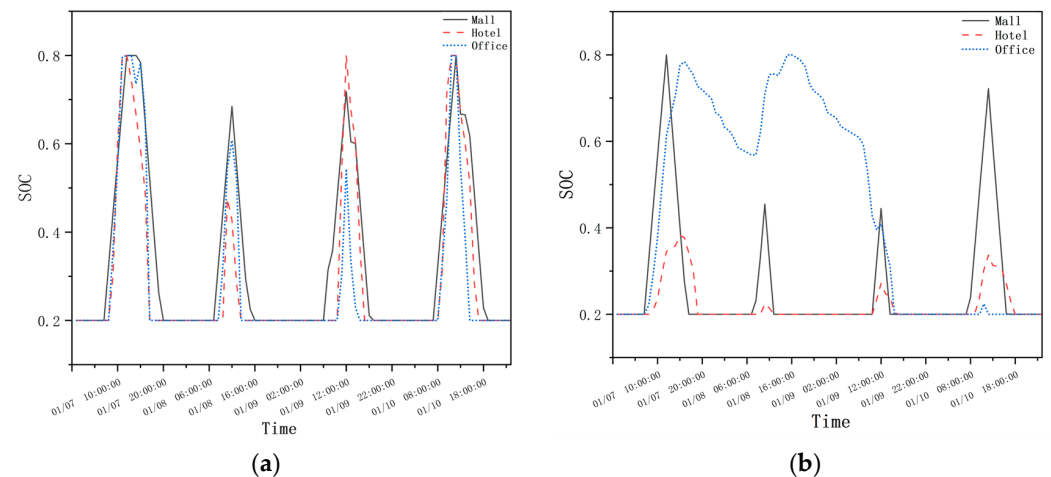


Figure 9. The hourly SOC operation results under different strategies for four consecutive days: (a) Hourly SOC running results of ODT-MAPPO strategy; (b) Hourly SOC running results of co-SGSU strategy.

Under the ODT-MAPPO strategy, each building decides actions according to its own state, such as power generation, electrical load, and energy storage, while accounting for its impact on the overall cluster performance at each time step. Therefore, a high degree of cooperation exists among the buildings in the cluster. Although the co-SGSU prioritizes each building's own needs, during the first two holidays, the PV power generation of office buildings is much higher than the electricity consumption. After the battery is fully charged, the surplus electricity is shared. Owing to the battery self-discharge loss, the SOC of the battery still declines when it does not supply power to the building. Under this strategy, cooperation between buildings is limited; however, the cooperative nature of resource sharing remains.

The dispatch results for improving the local PV power consumption from the ODT-MAPPO, typical MAPPO, and two common priority strategies are listed in Table 1, with the noco-SGSU strategy used as the benchmark. Compared with the noco-SGSU strategy, the three cooperative strategies can effectively reduce the cumulative power supply from the public grid, indicating that cooperation between buildings significantly improves dispatch optimization in the problem of cluster power dispatch. Specifically, the ODT-MAPPO strategy achieves the lowest grid replenishment and the highest local PV power consumption among the four strategies, with a monthly PV power consumption increase of 5870.8 kWh.

Table 1. Dispatch results.

Policy	Grid Replenishment (kWh)	Local PV Power Consumption (kWh)	PV Power Consumption Increase (%)
ODT-MAPPO	−4960.2	+5870.8	+1.78
Typical MAPPO	−2415.6	+3364.4	+0.87
Co-SGSU	−2306.2	+2271.6	+0.83
Noco-SGSU	365,989.0	196,455.7	−

In this study, the randomness of building electricity load caused by occupancy is taken as an example to augment the samples for MAPPO training. A comparison of the dispatch verification results between ODT-MAPPO and typical MAPPO in January shows that the PV power consumption increase rate of ODT-MAPPO is almost twice that of typical MAPPO, indicating that the randomness of occupancy cannot be ignored when obtaining the decentralized dispatch strategy based on the MAPPO power dispatch model. These results suggest that the ODT-MAPPO model can effectively obtain dispatch policies under the specific conditions examined in this study.

Similarly, using the noco-SGSU strategy as a comparison benchmark, the relative reduction rate of cumulative grid replenishment under ODT-MAPPO for each month can be obtained, as listed in Table 2. These results show that the ODT-MAPPO strategy can reduce the monthly grid replenishment of the cluster by 0.55–2.56% throughout the year, with an average increase in PV consumption of 7.28 MWh, demonstrating the effectiveness of the year-round dispatch. In practical operation, each BIPE can select the corresponding ODT-MAPPO policy for a specific month.

Table 2. Monthly ODT-MAPPO dispatch optimization grid replenishment results.

Month	January	February	March	April	May	June
Reduction rate/%	1.36	1.98	1.91	2.09	1.39	0.55
Month	July	August	September	October	November	December
Reduction rate/%	1.41	0.61	0.83	1.49	1.16	2.56

As shown in Table 2, the reduction rate in June is notably lower than in other months, while December achieves the highest. Combined with the PV generation and building electricity consumption data, this difference is primarily attributable to the variation in PV surplus available for cooperative scheduling. In June, building electricity consumption rises significantly due to cooling loads, while PV generation decreases under unfavorable weather conditions, resulting in a reduced PV surplus that limits the capacity for inter-building power coordination. In contrast, December features lower building electricity demand with relatively stable PV output, yielding a larger PV surplus that provides greater flexibility for collaborative dispatch optimization.

To further illustrate the practical significance of the improvement, the benefits of the ODT-MAPPO strategy are quantified using January as an example. In January, the ODT-MAPPO strategy reduces the cumulative grid replenishment by 4960.2 kWh compared with the non-cooperative SGSU strategy. In terms of carbon emissions, using the 2022 Shanghai electricity carbon emission factor of 0.5849 kgCO₂/kWh, the monthly carbon reduction is approximately 2901.2 kgCO₂. Given that the monthly reduction rate ranges from 0.55% to 2.56% throughout the year, the annual cumulative benefits in energy savings and carbon emission mitigation would be considerably more substantial.

4.2.4. Reverse Design of ESS Based on ODT-MAPPO Strategy

When the construction cost is fixed, the rated capacity and rated power of the building ESS should be considered comprehensively according to the dispatch strategy. After the BIPEC-oriented decentralized dispatch strategy is obtained based on the ODT-MAPPO model, the ESS in each BIPE is reversely optimized according to this strategy. Taking the dispatch strategy corresponding to the objective function in Equation (10) as an example, the rated capacity and rated power of the ESS in each BIPE are optimized considering the construction cost. The construction cost C is typically defined as follows:

$$C = C_p P + C_e V_{max} \quad (18)$$

where C_p and C_e correspond to the unit charge and discharge power cost and the unit capacity cost of the energy storage, respectively.

Comprehensively considering the power generation and electricity consumption of the three buildings selected in the case study, two schemes with a relative difference of less than 0.1% in construction costs are selected for each building, as shown in Table 3.

Table 3. ESS scheme parameters.

Building	P (kW)	V_{max} (kWh)	C ($\times 10^4$ CNY)	No.
Office, hotel	260	520	1331.2	1
	250	525	1330	2
Mall	150	1200	2568	3
	500	1000	2560	4

Two ESS schemes are designed for each of the three buildings, forming eight combined cluster schemes. For each scheme combination, the ODT-MAPPO policy is retrained independently following the same training procedure, with the ESS parameters updated accordingly in the simulation-based training environment. The dispatch strategy of the cluster in January is obtained based on the ODT-MAPPO model, and the dispatch effect is verified. The dispatch results are listed in Table 4.

Table 4. Dispatch results of different ESS schemes.

Scheme No.			Grid Replenishment (kWh)	Absolute Difference (kWh)
Office	Hotel	Mall		
1	1	3	361,028.8	-
1	1	4	362,345.8	+1317
1	2	3	360,466.1	−562.7
1	2	4	361,928.1	+899.3
2	1	3	361,152.4	+123.6
2	1	4	360,852.8	−176.0
2	2	3	360,526.9	−501.9
2	2	4	362,212.8	+1184

Taking the scheme selected in this case study as the benchmark, scheme No. 123 minimizes the cumulative grid replenishment for the building cluster in January. Accordingly, the energy storage scheme can be preliminarily optimized as follows: the rated energy storage capacities of the office, hotel and mall are 520, 525, and 1200 kWh, respectively, and the corresponding values of rated electric power are 260, 250, and 150 kW.

The optimization of the ESS configurations, combined with the ODT-MAPPO model, enables the building cluster to achieve a more comprehensive BIPEC dispatch design that jointly addresses both hardware configurations and dispatch strategies.

5. Conclusions

This study proposes a decentralized collaborative dispatch model for BIPECs based on the coupled ODT-MAPPO model, with the objective of minimizing the cumulative power supply from the public grid. The BIPEC system model is analyzed, the power and communication flows of the system are determined, the BIPEC-oriented MAS and MAPPO-based dispatch models are defined, and simulation-assisted ODT technology is introduced for data augmentation to address the problem of insufficient operational data in BIPECs. A case study of a typical public building cluster in Shanghai is conducted, comparing the ODT-MAPPO strategy with noco-SGSU, co-SGSU and typical MAPPO strategies. The main conclusions are as follows:

- (1) The MAPPO algorithm achieves effective training of each CA and converges steadily to the optimal dispatch strategy through optimized action and reward function design. Digital experiments verify that the reward function, which comprehensively accounts for both power generation and electricity consumption, enables the cumulative reward and power supply from the public grid to converge after approximately 2500 episodes.
- (2) The dispatch strategy derived from the ODT-MAPPO model demonstrates coordinated dispatch behavior and practical feasibility in the tested BIPEC scenario. Compared with the non-cooperative self-generating and self-using strategy, the ODT-MAPPO strategy reduces the monthly grid replenishment by 0.55–2.56% throughout the year.
- (3) The simulation-assisted ODT technology enhances the training reliability of MAPPO in BIPEC scenarios that lack operational data, ensuring that each CA obtains more effective dispatch strategies. Compared with the typical MAPPO strategy based on typical energy consumption data, the ODT-MAPPO strategy can further reduce the cumulative power supply from the public grid by 2544.58 kWh in January, and achieves a PV power consumption increase rate approximately twice that of the typical MAPPO strategy.
- (4) The ODT-MAPPO model supports the co-optimization of ESS scheme and dispatch strategy, enabling more comprehensive and effective optimization of dispatch design for BIPECs. Among all evaluated schemes, scheme No. 123 minimizes the cluster's grid replenishment while considering the construction cost. The preliminarily optimized rated ESS capacities for the office, hotel, and mall are 520 kWh, 525 kWh, and 1200 kWh, respectively, with corresponding rated ESS power of 260 kW, 250 kW, and 150 kW.

As foundational research on decentralized flexible power dispatch strategy for building clusters, this study has several limitations. (1) The current framework addresses only single-objective minimization; future work will extend it to multi-objective optimization incorporating economic factors, drawing on hybrid surrogate modeling [46] and ANN integrated optimization methodologies [47]. (2) The fully cooperative MAS structure does not account for mixed cooperation–competition scenarios among buildings with heterogeneous ownership and economic interests. (3) The current ODT framework operates offline with simulated data and does not incorporate advanced occupant behavior models or validation under more challenging conditions; future work will calibrate models using measured building performance data. (4) The comparison baselines are limited to rule-based and standard MAPPO strategies; benchmarking against stronger optimization approaches would better clarify the improvement. (5) The 1 h dispatch time step may smooth out short-term PV fluctuations and fast battery dynamics; finer temporal resolutions could reveal additional optimization opportunities at the cost of increased computational burden. (6) The annual carbon emission reduction and economic benefits have not been comprehensively assessed at the annual scale; future work should further evaluate the comprehensive benefits of the proposed dispatch strategy.

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

Funding: This research was supported by the National Natural Science Foundation of China, grant number 52478102.

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

Conflicts of Interest: Author Haowei Xing was employed by the company The Architectural Design and Research Institute of Zhejiang University Co., Ltd. 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.

Abbreviations

The following abbreviations are used in this manuscript:

AC	Alternating current
BIPE	Building integrated with photovoltaics and energy storage system
BIPEC	Building integrated with photovoltaics and energy storage system cluster
BIPV	Building integrated with photovoltaics
CA	Control agent
DC	Direct current
DoD	Depth of discharge
DT	Digital twin
ODT	Offline digital twin
ESS	Energy storage system
ESSA	Energy storage system agent
LA	Load agent
LIB	Lithium-ion battery
MA	Microgrid agent
MADRL	Multi-agent deep reinforcement learning
MAPPO	Multi-agent proximal policy optimization
MAS	Multi-agent system
MDP	Markov decision process
PV	Photovoltaic
PVA	Photovoltaic power generation agent
RERs	Renewable energy resources
RL	Reinforcement learning
SGSU	Self-generating and self-using

References

- Jin, C.; Lv, Z.; Li, Z.; Sun, K. Green finance, renewable energy and carbon neutrality in OECD countries. *Renew. Energy* **2023**, *211*, 279–284. [\[CrossRef\]](#)
- Paska, J.; Surma, T. Electricity generation from renewable energy sources in Poland. *Renew. Energy* **2014**, *71*, 286–294. [\[CrossRef\]](#)
- Li, Y.; Li, L.; Deng, W.; Zhu, D.; Hong, L. Building integrated photovoltaic (BIPV) development knowledge map: A review of visual analysis using CiteSpace. *Buildings* **2023**, *13*, 389. [\[CrossRef\]](#)
- Hsieh, S.; Omu, A.; Orehounig, K. Comparison of solar thermal systems with storage: From building to neighbourhood scale. *Energy Build.* **2017**, *152*, 359–372. [\[CrossRef\]](#)
- GlobalABC. 2021 *Global Status Report for Buildings and Construction*; GlobalABC: Paris, France, 2021.
- Laustsen, J. *Energy Efficiency Requirements in Building Codes, Energy Efficiency Policies for New Buildings*; International Energy Agency: Paris, France, 2008.
- Tsikalakis, A.G.; Hatziaargyriou, N.D. Centralized control for optimizing microgrids operation. In Proceedings of the IEEE Power and Energy Society General Meeting, San Diego, CA, USA, 24–29 July 2011.
- Olivares, D.E.; Canizares, C.A.; Kazerani, M. A centralized energy management system for isolated microgrids. *IEEE Trans. Smart Grid* **2014**, *5*, 1864–1875. [\[CrossRef\]](#)
- Karavas, C.S.; Kyriakarakos, G.; Arvanitis, K.G.; Papadakis, G. A multi-agent decentralized energy management system based on distributed intelligence for the design and control of autonomous polygeneration microgrids. *Energy Convers. Manag.* **2015**, *103*, 166–179. [\[CrossRef\]](#)

10. Streng, L.; Schultz, P.; Kurths, J.; Raisch, J.; Hellmann, F. A multiplex, multi-timescale model approach for economic and frequency control in power grids. *Chaos* **2020**, *30*, 033138. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Wang, J.; Chen, S.; Li, B.; Qi, B. Coordination and optimization of virtual power plant based on multi-agent system. In Proceedings of the 2022 5th International Conference on Circuits, Systems and Simulation (ICCS), Chengdu, China, 12–14 August 2022.
12. Nguyen, T.T.; Nguyen, N.D.; Nahavandi, S. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybern.* **2020**, *50*, 3826–3839. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Gu, Y.; Paddle, M.; Frimpong, Y.; Khatib, P.P. Decentralized coordination of energy utilization for residential households in the smart grid. *IEEE Trans. Smart Grid* **2013**, *4*, 1341–1350. [\[CrossRef\]](#)
14. 't Hoen, P.J.; Tuyls, K.; Panait, L.; Luke, S.; La Poutré, J.A. An overview of cooperative and competitive multiagent learning. In *Learning and Adaption in Multi-Agent Systems*; Tuyls, K., Hoen, P.J., Verbeeck, K., Sen, S., Eds.; Springer: Berlin, Germany, 2006; Volume 3898, pp. 1–46.
15. Perera, A.T.D.; Kamalaruban, P. Applications of reinforcement learning in energy systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [\[CrossRef\]](#)
16. Yu, C.; Velu, A.; Vinitzky, E.; Gao, J.; Wang, Y.; Bayen, A.; Wu, Y. The surprising effectiveness of PPO in cooperative, multi-agent games. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 24611–24624. [\[CrossRef\]](#)
17. Wang, Y.; Liu, C.H.; Piao, C.; Yuan, Y.; Han, R.; Wang, G.; Tang, J. Human-drone collaborative spatial crowdsourcing by memory-augmented and distributed multi-agent deep reinforcement learning. In Proceedings of the 2022 IEEE 38th International Conference on Data Engineering (ICDE), Kuala Lumpur, Malaysia, 9–12 May 2022.
18. Jia, Y.; Zhang, C.; Huang, Y.; Zhang, W. Lyapunov optimization based mobile edge computing for internet of vehicles systems. *IEEE Trans. Commun.* **2022**, *70*, 7418–7433. [\[CrossRef\]](#)
19. Zhang, Q.; Mahajan, Y.; Chen, I.R.; Ha, D.S.; Cho, J.H. An attack-resilient and energy-adaptive monitoring system for smart farms. In Proceedings of the 2022 IEEE Global Communications Conference (GLOBECOM), Rio de Janeiro, Brazil, 4–8 December 2022.
20. Guo, D.; Tang, L.; Zhang, X.; Liang, Y.C. Joint optimization of handover control and power allocation based on multi-agent deep reinforcement learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 13124–13138. [\[CrossRef\]](#)
21. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M. On-line building energy optimization using deep reinforcement learning. *IEEE Trans. Smart Grid* **2019**, *10*, 3698–3708.
22. Fu, Q.; Han, Z.; Chen, J.; Lu, Y.; Wu, H.; Wang, Y. Applications of reinforcement learning for building energy efficiency control: A review. *J. Build. Eng.* **2022**, *50*, 104165. [\[CrossRef\]](#)
23. Liu, Y.; Zhang, D.; Gooi, H.B. Optimization strategy based on deep reinforcement learning for home energy management. *CSEE J. Power Energy Syst.* **2020**, *6*, 572–582. [\[CrossRef\]](#)
24. Cao, D.; Zhao, J.; Hu, W.; Ding, F.; Huang, Q.; Chen, Z. Attention enabled multi-agent DRL for decentralized volt-VAR control of active distribution system using PV inverters and SVCs. *IEEE Trans. Sustain. Energy* **2021**, *12*, 1582–1592. [\[CrossRef\]](#)
25. Hernandez-Leal, P.; Kartal, B.; Taylor, M.E. A survey and critique of multiagent deep reinforcement learning. *Auton. Agents Multi-Agent Syst.* **2019**, *33*, 750–797. [\[CrossRef\]](#)
26. Boschert, S.; Rosen, R. Digital twin—The simulation aspect. In *Mechatronic Futures*; Hehenberger, P., Bradley, D., Eds.; Springer: Cham, Switzerland, 2016; pp. 59–74.
27. Khan, S.; Alzaabi, A.; Ratnarajah, T.; Arslan, T. Novel statistical time series data augmentation and machine learning based classification of nonobtrusive respiration data for respiration digital twin model. *Comput. Biol. Med.* **2024**, *168*, 107825. [\[PubMed\]](#)
28. Hatziaargyriou, N.; Asano, H.; Iravani, R.; Marnay, C. Microgrids. *IEEE Power Energy Mag.* **2007**, *5*, 78–94. [\[CrossRef\]](#)
29. Zenginis, I.; Vardakas, J.S.; Echave, C.; Morató, M.; Abadal, J.; Verikoukis, C.V. Cooperation in microgrids through power exchange: An optimal sizing and operation approach. *Appl. Energy* **2017**, *203*, 972–981. [\[CrossRef\]](#)
30. Bhavsar, Y.S.; Joshi, P.V.; Akolkar, S.M. Energy management in DC microgrid. In Proceedings of the 2015 International Conference on Energy Systems and Applications (ICESA), Pune, India, 30 October–1 November 2015.
31. Koirala, B.P.; Koliou, E.; Friege, J.; Hakvoort, R.A.; Herder, P.M. Energetic communities for community energy: A review of key issues and trends shaping integrated community energy systems. *Renew. Sustain. Energy Rev.* **2016**, *56*, 722–744. [\[CrossRef\]](#)
32. Li, Y.; Wang, B.; Yang, Z.; Li, J.; Chen, C. Hierarchical stochastic scheduling of multi-community integrated energy systems in uncertain environments via Stackelberg game. *Appl. Energy* **2022**, *308*, 118392. [\[CrossRef\]](#)
33. Zhou, Y.; Wei, Z.; Sun, G.; Cheung, K.W.; Zang, H.; Chen, S. A robust optimization approach for integrated community energy system in energy and ancillary service markets. *Energy* **2018**, *148*, 1–15. [\[CrossRef\]](#)
34. Panda, S.; Mohanty, S.; Rout, P.K.; Sahu, B.K. A conceptual review on transformation of micro-grid to virtual power plant: Issues, modeling, solutions, and future prospects. *Int. J. Energy Res.* **2022**, *46*, 7021–7054.
35. Thavlov, A.; Bindner, H.W. Utilization of flexible demand in a virtual power plant set-up. *IEEE Trans. Smart Grid* **2015**, *6*, 640–647. [\[CrossRef\]](#)
36. Wang, J.; Guo, C.; Yu, C.; Liang, Y. Virtual power plant containing electric vehicles scheduling strategies based on deep reinforcement learning. *Electr. Power Syst. Res.* **2022**, *205*, 107714. [\[CrossRef\]](#)

37. Duraivelu, R.; Elumalai, N. Performance evaluation of a decentralized rooftop solar photovoltaic system with a heat recovery cooling unit. *Environ. Sci. Pollut. Res.* **2021**, *28*, 19351–19366. [[CrossRef](#)]
38. Kuo, H.J.; Hsieh, S.H.; Guo, R.C.; Chan, C.C. A verification study for energy analysis of BIPV buildings with BIM. *Energy Build.* **2016**, *130*, 676–691. [[CrossRef](#)]
39. Chen, T.; An, Y.; Heng, C.K. A review of building-integrated photovoltaics in Singapore: Status, barriers, and prospects. *Sustainability* **2022**, *14*, 10160. [[CrossRef](#)]
40. Merei, G.; Moshövel, J.; Magnor, D.; Sauer, D.U. Optimization of self-consumption and techno-economic analysis of PV-battery systems in commercial applications. *Appl. Energy* **2016**, *168*, 171–178.
41. Tao, F.; Zhang, M.; Liu, Y.; Nee, A.Y.C. Digital twin driven prognostics and health management for complex equipment. *CIRP Ann.* **2018**, *67*, 169–172. [[CrossRef](#)]
42. Qi, Q.; Tao, F.; Hu, T.; Anwer, N.; Liu, A.; Wei, Y.; Wang, L.; Nee, A.Y.C. Enabling technologies and tools for digital twin. *J. Manuf. Syst.* **2021**, *58*, 3–21. [[CrossRef](#)]
43. Kabir, M.R.; Halder, D.; Ray, S. Digital Twins for IoT-Driven Energy Systems: A Survey. *IEEE Access* **2024**, *12*, 177123–177143.
44. Ugwu, C.N.; Ogenyi, F.C.; Ugwu, J.N.; Ugwu, P.O.; Okon, M.B. Integration of AI-driven digital twins for real-time optimization of renewable energy grids. *Front. Energy Res.* **2026**, *14*, 1748233.
45. Macdonald, I.; Strachan, P. Practical application of uncertainty analysis. *Energy Build.* **2001**, *33*, 219–227. [[CrossRef](#)]
46. Zhu, C.; Samuel, O.D.; Patel, G.C.M.; Der, O.; Abbas, M.; Hussain, F.; Ting, T.T. Enhancing CI engine performance and emission control using a hybrid RSM–Rao algorithm for ZnO-doped castor–neem biodiesel blends. *Case Stud. Therm. Eng.* **2025**, *74*, 106841.
47. Khujamberdiev, R.; Cho, H.M. Artificial intelligence in automotives: ANNs’ impact on biodiesel engine performance and emissions. *Energies* **2025**, *18*, 438. [[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.