Interactive Integrated Design Framework for Optimizing the Structure, Capacity, and Operation of Multienergy Systems via Reinforcement Learning

Hui Zhang, Lizhi Zhang, Haozeng Bie, Zhiwei Xu, Guangyao Fan, Bin Jia, and Bo Sun, *Member, IEEE*

***Abstract*—The complex and varied source–load characteristics of multienergy systems (MESs) make it difficult to match supply with demand while maintaining economic efficiency. Previous design methods have not fully considered the coupled relationships between the system structure scheme and the equipment capacity and operation scheme, which has made it impossible for them to design the optimal MES in terms of overall performance. To address this issue, a bilevel interactive integrated design framework (IIDF) is proposed for MESs. A generalized description of the integrated design problem for MESs was formulated that fully considers the selection of devices and their connections. In the first level, and the structure search strategy is constructed based on reinforcement learning to determine the optimal system structure. To accelerate convergence, a search space is designed that narrows the solution domain of candidate devices and connection parameters by integrating domain knowledge of multienergy flow supply and demand balance and heterogeneous energy cascade transformation. In second level, the capacity–operation co–optimization is executed for the determined MES structure from the first level. The performance of IIDF was evaluated against two traditional design methods in three typical scenarios, and the results demonstrated its effectiveness and superiority.**

***Index Terms*—Bilevel framework, coordinated optimization, multienergy system (MES), reinforcement learning (RL).**

Received 13 December 2024; revised 24 February 2025; accepted 10 March 2025. Date of publication 28 April 2025; date of current version 9 July 2025. This work was supported in part by the National Key Research and Development Program of China under Grant 2022YFB4004401, in part by the National Natural Science Foundation of China under Grant 62192753, Grant 62133008, and Grant 62303269, and in part by the National Natural Science Foundation of Shandong Province under Grant ZR2023QF116. Paper no. TII-24-6695. *(Corresponding authors: Bo Sun.)*

Hui Zhang, Zhiwei Xu, Guangyao Fan, Bin Jia, and Bo Sun are with the School of Control Science and Engineering, Shandong University, Jinan 250014, China (e-mail: 202234949@mail.sdu.edu.cn; zwxu@email.sdu.edu.cn; 202320735@mail.sdu.edu.cn; 202120615@mail.sdu.edu.cn; sunbo@sdu.edu.cn).

Lizhi Zhang is with the School of Reconnaissance, Shandong Police College, Jinan 250200, China (e-mail: 201920499@mail.sdu.edu.cn).

Haozeng Bie is with the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai 264209, China (e-mail: 202200800082@mail.sdu.edu.cn).

Digital Object Identifier 10.1109/TII.2025.3556037

NOMENCLATURE

A. Sets and Indices

Ω	Previously defined set of alternative devices.
Ψ	Subset of Ω.
Φ	Set of energy forms.
$\mathbb{N}$	Search space.
i	Index for alternative devices.
d	Index for typical days.
l	Index for decimal number of structure strings.
t	Index for time.

B. Parameters

ζ_i, ϕ_i	Unit investment and maintenance costs of device i.
ϵ_i	Return on investment coefficient of device i.
φ_i	Life span of device i.
ε	Benchmark discount rate.
λ_i	Maximum charge or discharge coefficient of device i.
ξ	Cost coefficients of the MES.
ν_i	Energy conversion efficiency of device i.
η	Learning rate.
θ	Parameters of policy function π.
δ_i	Energy loss coefficient of device i.
Γ_u, Γ_l	Upper and lower bounds of the annual total cost.
m	Number of energy forms.
w	Number of types of heat transfer devices.
y	Number of types of heat/cooling storage devices.
D^d	Number of days of the dth typical day.
L	Length of the structure string.
N	Number of typical days in a year.
$P_{\text{grid},+}$	Price of purchasing electricity.
$P_{\text{grid},-}$	Price of selling electricity.
P_g	Price of gas.
P_{CT}	Carbon tax.
CE^d	Carbon emission on the dth typical day.
T	Time period.
$\triangle t$	Time interval.

C. Variables

$\mathcal{J}_{\text{MES}}$	Total annualized cost of the MES.
α_i	Selection of device i.

1941-0050 © 2025 IEEE. All rights reserved, including rights for text and data mining, and training of artificial intelligence and similar technologies. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

β_i Capacity of device i.
$\overline{\gamma}_i^{a,d}$ Input power of device i.
$\gamma_i^{b,d}$ Output power of device i.
$\gamma_{\text{grid},+}^{\text{e},d}$ Purchasing electricity from the grid.
$\gamma_{\text{grid},-}^{\text{e},d}$ Selling electricity to the grid.
$\psi_{i,+}$ Charging state of device i.
$\psi_{i,-}$ Discharging state of device i.
$A_{1:L}^k$ List of action sequences of search policy.
C_{ic} Investment cost of the MES.
C_{mc} Maintenance cost of the MES.
C_{oc} Operation cost of the MES.
$E[R]$ Expectation of R.
G Strategy gradient.
$L^{\text{e},d}$ Electricity loads on the dth typical day.
$L^{\text{h},d}$ Heat loads on the dth typical day.
$L^{\text{c},d}$ Cooling loads on the dth typical day.
$L^{\text{g},d}$ Gas loads on the dth typical day.
Q_i Energy storage state of device i.
R^k Reward in the kth iteration.
$S_{1:L}^k$ List of state sequences of search policy.
p Probability of sampling each structure string.

D. Functions

π Policy function.
b Baseline function.

I. Introduction

A. Background and Motivation

ENHANCING energy efficiency and the market share of renewable energy sources are considered effective means of achieving carbon neutrality [1]. Multienergy systems (MESs) utilize energy production, conversion, and storage technologies to meet multiple load demands [2] and play an important role in promoting the consumption of renewable energy sources and improving energy efficiency [3].

For diverse application scenarios, rationally planning the system structure and accurately configuring equipment capacity are the fundamental prerequisites for ensuring the efficient, economical, stable, and low-carbon operation of MESs. However, the MES involves a wide variety of devices types and flexible structural configurations, especially in the context of the complex and diverse energy supply–demand characteristics. On the supply side, renewable energy sources are intermittent and fluctuating by nature, and the use of various energy conversion and storage devices results in deeply coupled electricity, gas, and heat energy flows. On the demand side, various scenarios of multiple load demands are possible that all have different characteristics. This makes system design extremely difficult and hinders the application of MESs. Thus, determining the optimal structure, size, and dispatch of MES is always a valuable and challenging topic [4], [5], [6].

B. Literature Review

Many studies have focused on optimizing the capacity configuration of MESs with a given structure. The structure of a MES is relatively fixed and usually comprises energy supply devices that utilize renewable energy sources and a combined cooling, heating, and power system [e.g., power generation units (PGU), absorption chillers (AC), electric chillers (EC), gas boilers (GB)]. But the core power generation device of a MES is the cogeneration unit, which has a fixed or only slightly adjustable thermoelectric output ratio that can easily result in energy waste [7]. Therefore, the flexibility of a MES can be improved by adding electricity storage (ES), cold storage (CS), and gas storage devices to previous structure. On the basis of this system structure, the energy supply devices should be optimized to have sufficient capacity to satisfy the energy demands of the system while minimizing investment costs, energy consumption, and carbon emissions [8]. Preset operational schemes formulate rules for the output of each device (e.g., following the electric load or following the thermal load [9]), which is a simple and feasible approach but has difficulty handling the intermittency and volatility of renewable energy sources and loads. As a result, the obtained capacity is not the optimal solution.

In view of this, some studies further realize the joint optimization of capacity and operation for MESs. Ren et al. [10] pointed out that rule–based operational schemes restrict the working order of devices, so they proposed a two–layer optimization method for the capacity and operation of a MES that greatly improves the economic and energy performance compared with an optimization method that assumes a fixed operational scheme. Han et al. [11] proposed a bilevel optimization method for an island MES where the upper level established a multiobjective programming model that optimizes the capacity in terms of economy and power quality and stability while the lower level constructed a distribution robust optimization model that formulates an operational scheme. Geng et al. [12] proposed a cluster–based multipolicy strategy for optimizing both the capacity and operation of MESs. Deng et al. [13] established a mixed integer nonlinear programming model to find the optimal capacity and operational scheme for a MES considering different energy conversion efficiencies. However, methods that assume a given structure have difficulty adapting to different scenarios, which can degrade the economic performance of the MES.

Recently, some scholars have tried to improve system economic performance and scenario adaptability by considering different structure schemes of MES [14], [15]. Li et al. [16] defined the set of candidate structures for a MES in advance and adopted a genetic algorithm (GA) to optimize the capacity configuration and operation scheme of all candidate structures for different load types. However, the number of candidate structures in the set is very limited. Ameri et al. [17] established a mixed integer linear programming (LP) model for optimizing the structure and capacity of a residential MES that uses binary decision variables to determine whether alternative devices are selected and continuous variables to determine the optimal capacity and operational scheme of the devices. To further obtain more flexible structural schemes, Zhou et al. [18] proposed a MES super–structure comprising multiple candidate energy conversion and storage devices, in which connections between

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

devices are predefined and then optimized the structure, capacity, and operation of the MES. Huang et al. [19] developed a hierarchical model to describe the connections between devices and established a two-stage mixed integer LP model for designing a MES.

The above studies generally determined the system structure according to preset rules, such as hierarchical models, which limited the flexibility of the energy conversion and storage devices and cannot obtain the optimal economic performance of the MES design scheme. For example, the energy interactions between secondary energy conversion devices are limited, and the output energy of energy storage devices is difficult to convert and utilize twice. An ideal solution to optimizing the overall structure of a MES without being restricted by a hierarchical model is to consider the selection and location of all devices as decision variables. Unfortunately, the structure is flexible and changeable as well as deeply coupled with the capacity configuration and operational scheme, which increases the scale and complexity of an integrated design approach undoubtedly.

Numerical programming and evolutionary algorithms have been widely adopted to solve the above optimization problems, such as mixed-integer LP [12], [17], [19], [20], GA [16], and multitask evolutionary optimization algorithms [21]. However, with the increase of the complexity of application scenarios and the variety of alternative devices, the scale of these optimization problems further increases. As a result, these algorithms face the problems of slow convergence and local optimality. As an interactive learning method, reinforcement learning (RL) [22] is suitable for solving optimization problems of complex dynamic systems with uncertainty, and has therefore demonstrated great potential in addressing the increasingly complex MES optimization problems. Aiming at the economy of system operation, several typical RL algorithms have been adopted to optimize the operational schemes of devices of MESs, including Monte Carlo [23], deep deterministic policy gradient [24], [25], [26], Q-learning [27], and actor–critic algorithm [28]. However, these operational schemes may deviate from the actual situations, resulting in the lack of feasibility. Therefore, Chen et al. [29] developed a double deep Q-learning algorithm-based multitimescale optimization technique, building on real-time scheduling, to further improve the performance of the optimization strategy. Furthermore, several studies have specifically concentrated on the advantages of RL in terms of environmental adaptability. Zhou et al. [30] proposed a deep RL approach for MES economic dispatch. The RL agent was trained by the distributed proximal policy optimization algorithm and was capable of addressing economic scheduling challenges across various operational scenarios without recalculation. Although RL has been widely applied for optimizing MESs, there is currently limited research on the design of MESs utilizing RL. When designing MES, the complex coupling relationship among system structure, capacity configuration, and operational scheme must be taken into account, leading to a dramatic increase in the dimensions of state space and action space of RL. Undoubtedly, the solution process of RL will also become extremely time-consuming and computationally intensive. In this case, how to make decisions while meeting the economics and multiple constraints of MES has become a significant technical challenge for RL.

C. Contribution and Paper Organization

To address the above technical challenges, this study provides a new and efficient solution for structure–capacity–operation integrated design of MES. To the best of the authors knowledge, we are among the first to apply RL [22] to design MES. Specifically, the main contributions and innovations of this work are as follows.

1) A universal MES integrated design problem is proposed and described that can be flexibly applied to different scenarios.
2) A RL–based bilevel interactive integrated design framework for MES is proposed to determine the optimal system structure, capacity configuration, and operational scheme.
3) A structure search space is constructed by merging domain knowledge (i.e., the law of multienergy supply–demand balances and the second law of thermodynamics) that greatly reduces the solution space and accelerates iterations.

The rest of this article is organized as follows. Section II describes the universal MES integrated design problem. Section III introduces the proposed framework. Section IV presents case studies. Finally, Section V concludes this article.

II. Problem Description

The integrated design of a MES aims to obtain an economical system scheme based on known device parameters and data of renewable energy sources, energy prices, and user loads. The design is built from scratch, which means that no prior assumptions are made about any devices at the beginning of the design. In general, the design of a MES includes the system structure, capacity configuration, and operational scheme. In this study, both the selection of devices and connection between devices are considered for the system structure and were derived from a previously man-defined set of alternative devices as complete as possible.

The performance of system design is usually measured by calculating the total annualized cost $\mathcal{J}_{\text{MES}}$ [7]

$$\mathcal{J}_{\text{MES}} = C_{\text{ic}} + C_{\text{mc}} + C_{\text{oc}}. \tag{1}$$

The investment cost C_{ic} is defined as

$$C_{\text{ic}} = \sum_{i\in\Omega} \alpha_i \beta_i \zeta_i \epsilon_i \tag{2}$$

$$\epsilon_i = \frac{\varepsilon(1+\varepsilon)^{\varphi_i}}{(1+\varepsilon)^{\varphi_i} - 1} \tag{3}$$

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

where $\Omega = \{\text{grid, PV, WT, PGU, GB, EB, HP, AC, EC, ES, HS, CS}\}$ is the previously defined set[1] of alternative devices and α_i is a 0–1 variable that indicates whether or not alternative device i is selected. β_i is an integer variable that represents the capacity of alternative device i.

The maintenance cost C_{mc} is defined as

$$C_{\text{mc}} = \sum_{i\in\Omega} \alpha_i \beta_i \phi_i. \tag{4}$$

The operational cost C_{oc} is defined as

$$C_{\text{oc}} = \sum_{d=1}^{N} D^d \sum_{t=1}^{24} (P_{\text{grid},+}(t)\gamma^{\text{e},d}_{\text{grid},+}(t) + P_{\text{grid},-}(t)\gamma^{\text{e},d}_{\text{grid},-}(t) + P_{\text{g}} L^{\text{g},d}(t) + P_{\text{CT}} CE^d(t)) \tag{5}$$

where $\gamma^{\text{e},d}_{\text{grid},+}$ and $\gamma^{\text{e},d}_{\text{grid},-}$ are purchasing electricity from the grid and selling electricity to the grid, respectively.

The investment and operational constraints must be satisfied simultaneously. The investment constraints include the selection, connection, and capacity of each device

$$\alpha_i \in \{0,1\} \quad \forall i \in \Omega \tag{6}$$

$$\alpha'_i \in [1,n] \quad \forall i \in \Omega \tag{7}$$

$$\alpha'_i \neq \alpha'_j \quad \forall i \neq j, i \in \Omega, j \in \Omega \tag{8}$$

$$\alpha_i \beta_{i,\min} \leqslant \beta_i \leqslant \alpha_i \beta_{i,\max}, i \in \Omega \tag{9}$$

where α'_i is an integer variable that represents the connection between alternative device i and other selected devices, and, the smaller the value of α'_i, the further forward device i's position is.

Assume that the alternative device i is an energy conversion device with the input and output energy forms a and b, respectively, then the operational constraints can be formulated as follows:

$$\overline{\gamma}^{a,d}_i(t) = \frac{\gamma^{b,d}_i(t)}{\nu_i} \quad \forall t \tag{10}$$

$$\overline{\gamma}^{a,d}_i(t) \leqslant \sum_{j\in\Psi} \gamma^{a,d}_j(t) \quad \forall t \tag{11}$$

$$0 \leqslant \gamma^{b,d}_i(t) \leqslant \alpha_i \beta_i \quad \forall t \tag{12}$$

$$\alpha'_j < \alpha'_i \quad \forall j \in \Psi \tag{13}$$

where $a, b \in \Phi$, $\Phi = \{\text{c, h, e, g}\}$ is a set of energy forms (i.e., cooling, heat, electricity, and gas). ν_i is the energy conversion efficiency from a to b. Ψ, a subset of Ω, is a set of devices whose positions are in front of device i and whose energy output forms are b. $\overline{\gamma}^{a,d}_i(t)$ represents the input power of the alternative device i at time t on the dth typical day. $\gamma^{b,d}_i(t)$ is a continuous variable that represents the output power of the alternative device i at time t on the dth typical day (i.e., the operational scheme of alternative device i).

[1] Generally, this set includes energy conversion devices such as a PGU, GB, EB, heat pump (HP), AC, and EC as well as energy storage devices such as ES, HS, and CS. The set also includes the grid, PV, and WT.

Assume that device i is an energy storage device with the energy form a, then the following constraints must be satisfied:

$$Q_i(t+1) = \left(\gamma^{a,d}_{i,+}(t)\nu_{i,+} - \gamma^{a,d}_{i,-}(t)/\nu_{i,-}\right) \triangle t + (1-\delta_i) Q_i(t) \quad \forall t \tag{14}$$

$$0 \leqslant \overline{\gamma}^{a,d}_i(t) \leqslant \alpha_i \psi_{i,+}(t) \lambda_i \beta_i \quad \forall t \tag{15}$$

$$0 \leqslant \gamma^{a,d}_i(t) \leqslant \alpha_i \psi_{i,-}(t) \lambda_i \beta_i \quad \forall t \tag{16}$$

$$0 \leqslant Q_i(t) \leqslant \alpha_i \beta_i \quad \forall t \tag{17}$$

$$0 \leqslant \psi_{i,+}(t) + \psi_{i,-}(t) \leqslant 1 \tag{18}$$

$$\psi_{i,+}(t), \psi_{i,-}(t) \in \{0,1\} \quad \forall t \tag{19}$$

$$Q_i(0) = Q_i(T) \tag{20}$$

where $\gamma^{a,d}_{i,+}$ and $\gamma^{a,d}_{i,-}$ are continuous variables that represent the input and output power (i.e., operational scheme of alternative device i). $\psi_{i,+}(t)$ and $\psi_{i,-}(t)$ are 0–1 variables that represent the charging or discharging states at time t, (i.e., an addition to the operational scheme for energy storage devices). $\nu_{i,+}$ and $\nu_{i,-}$ are the charging and discharging efficiencies, respectively. The position of energy storage device i also restricts the input source of its energy, and energy storage device i cannot produce or convert energy itself. Therefore, the input power of device i cannot exceed the sum of output powers of all devices whose positions are in front of device i, which requires satisfying (11) and (12).

The energy balance is constrained as follows:

$$L^{\text{e},d}(t) = \sum_{i\in\Omega}\sum_{b\in\Phi} \alpha_i \mathbb{I}(b=\text{e}) \gamma^{\text{b},d}_i(t) - \sum_{i\in\Omega}\sum_{a\in\Phi} \alpha_i \mathbb{I}(a=\text{e}) \overline{\gamma}^{\text{a},d}_i(t) \quad \forall t \tag{21}$$

$$L^{\text{h},d}(t) = \sum_{i\in\Omega}\sum_{b\in\Phi} \alpha_i \mathbb{I}(b=\text{h}) \gamma^{\text{b},d}_i(t) - \sum_{i\in\Omega}\sum_{a\in\Phi} \alpha_i \mathbb{I}(a=\text{h}) \overline{\gamma}^{\text{a},d}_i(t) \quad \forall t \tag{22}$$

$$L^{\text{c},d}(t) = \sum_{i\in\Omega}\sum_{b\in\Phi} \alpha_i \mathbb{I}(b=\text{c}) \gamma^{\text{b},d}_i(t) \quad \forall t \tag{23}$$

$$L^{\text{g},d}(t) = \sum_{i\in\Omega}\sum_{a\in\Phi} \alpha_i \mathbb{I}(b=\text{g}) \overline{\gamma}^{\text{a},d}_i(t) \quad \forall t \tag{24}$$

where $\mathbb{I}(\cdot)$ is an indicating function that has a value of 1 if $\cdot$ is true and 0 otherwise.

In summary, the integrated design problem of a MES for its structure, capacity configuration, and operational scheme can be generalized as follows:

$$(\alpha^*, \alpha'^*, \beta^*, \gamma^*, \psi^*) = \arg\min_{\alpha,\alpha',\beta,\gamma,\psi} \mathcal{J}_{\text{MES}}(\alpha, \alpha', \beta, \gamma, \psi) \quad \text{s.t.} \quad (6-24). \tag{25}$$

According to (25), the integrated design can be described as a complex mixed integer nonlinear programming problem with multiple dimensions and constraints that is extremely difficult

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

Algorithm 1: Proposed Interactive Integrated Design Framework.

Input: Search space $\mathbb{N}$, renewable energy, energy price, and user load data.

Compute:

Step 1: Parameter θ of search policy $\pi_\theta(A|S)$ in the first level is randomly initialized;

Step 2: $\pi_\theta(A|S)$ samples action A from $\mathbb{N}$ with the probability p (i.e., system structure α of the MES);

Step 3: In the second level, aiming at the minimum J_{MES}, the capacity–operation co–optimization is carried out to solve the optimal $\beta^*, \gamma^*, \psi^*$ for the given renewable energy, energy price, user load data, and system structure α;

Step 4: The obtained J_{MES} for system structure α is fed back to the search policy in the first level as the reward R;

Step 5: According to R, the policy gradient G of θ in $\pi_\theta(A|S)$ is estimated;

Step 6: According to G, θ in $\pi_\theta(A|S)$ is updated by gradient ascent algorithm;

Step 7: Repeat steps 2-6 until R converges.

Output:

$\alpha, \beta, \gamma, \psi$ of the last iteration (i.e., the optimal system design scheme of MES), $\alpha^*, \beta^*, \gamma^*, \psi^*$.

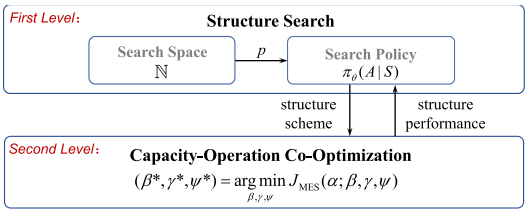


Fig. 1. Schematic illustration of the proposed interactive integrated design framework.

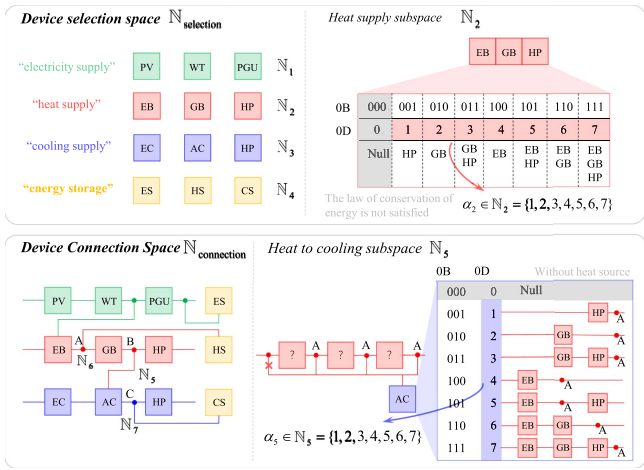


Fig. 2. Diagram of the search space.

to solve. This is because the decision variables are diverse in type and include 0–1, integer, and continuous variables. In addition, different types of decision variables are coupled, which makes representing constraints very complex such as in (11). Finally, there are a large number of nonlinear features caused by the multiplication of variables in the device model and energy balance constraints, such as in (12) and (15).

III. Interactive Integrated Design Framework

To solve the MES integrated design problem described in (25), a bilevel interactive integrated design framework is proposed based on RL.

A. Overview

The technical architecture of the proposed bilevel interactive integrated design framework is depicted in Fig. 1. In the first level, a RL-based search policy is constructed for determining the optimal MES structure. To accelerate convergence, a search space is designed by narrowing the solution domain of candidate devices and connection parameters. In the second level, the capacity–operation co–optimization is executed to determine the optimal capacity configuration and operational scheme of the MES. The optimized target is regarded as a reward feedback to the first level, and the performance of the search policy is improved by the strategy gradient algorithm. The interaction mechanism is shown in Algorithm 1.

B. Structure Search

1) Search Space: In this study, the combined feasible domain for system structure variables (including device selection variables α and connection variables α') is called search space. The dimension of the structure variables is affected by the number of alternative devices, and the system performance of MES will also changes when the connection relationships between different devices change. This means that the search space is very large and contains $2^n \times n!$ MES structures.

Ideally, the larger the search space, the more MES structures can be evaluated. But, a large number of MES structures in the search space have a very poor performance that cannot meet load demands or very high costs. Therefore, existing methods find it difficult to solve (25) or converge toward a solution.

A search space (denoted as $\mathbb{N}$) with domain knowledge is proposed to accelerate convergence, and, the selection and connection variables of all devices in system structure are defined as α, which is represented by a sequence of strings of decimal codes. Two types of domain knowledge, i.e., the law of multienergy supply–demand balances and the second law of thermodynamics, are considered to narrow the $\mathbb{N}$ and greatly reduces the solution space of structure variables.

As shown in Fig. 2, $\mathbb{N}$ is first decomposed into the device selection space $\mathbb{N}_{\text{selection}}$ and connection space $\mathbb{N}_{\text{connection}}$

$$\mathbb{N}_{\text{selection}} \cup \mathbb{N}_{\text{connection}} = \mathbb{N}. \tag{26}$$

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

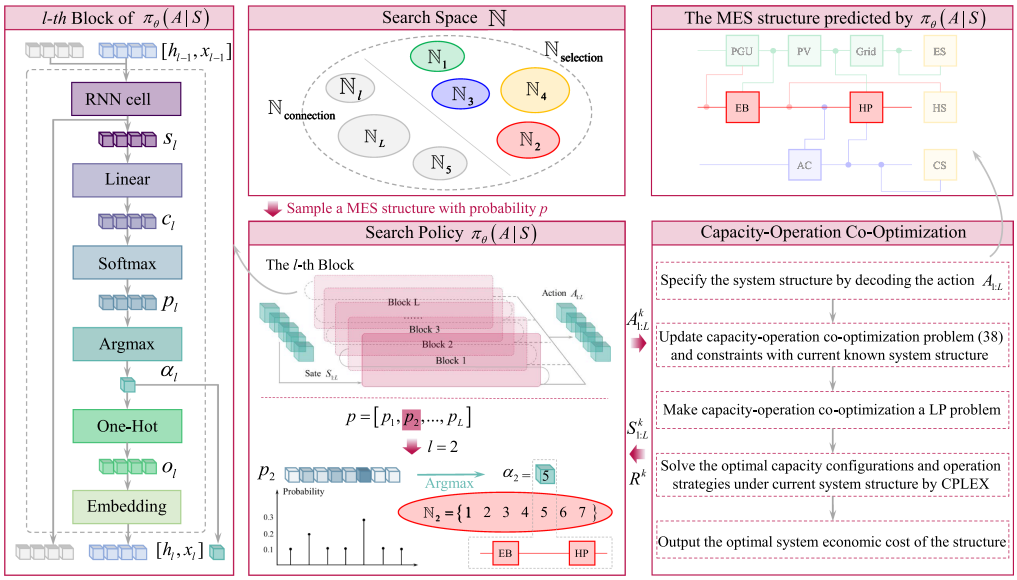


Fig. 3. Detailed schematic of the proposed framework.

Then, the $\mathbb{N}_{\text{selection}}$ is decomposed into different energy supply subspaces. To satisfy the law of multienergy supply–demand balances, at least one energy supply device should be selected from the different energy flow layers, which is implemented by removing a decimal code from the sequence of strings α, as shown in Fig. 2.

Considering that the energy storage devices cannot generate energy independently but are conducive to improving the flexibility of the energy supply, the energy storage devices are also divided into their own energy supply subspace. Assume that the designed MES includes m types of energy, $\mathbb{N}_{\text{selection}}$ will be decomposed into $m+1$ energy supply subspaces

$$\mathbb{N}_1 \cup \mathbb{N}_2 \cup \cdots \cup \mathbb{N}_{m+1} = \mathbb{N}_{\text{selection}} \tag{27}$$

$$\mathbb{N}_p \cap \mathbb{N}_q = \varnothing \quad \forall p \in [1, m+1] \quad \forall q \in [1, m+1], p \neq q. \tag{28}$$

According to the second law of thermodynamics, $\mathbb{N}_{\text{connection}}$ is decomposed as well, as shown in Fig. 2. The key relative connections of devices involved in the utilization of energy cascades are considered, such as the cooling transfer devices (e.g., EC). Similarly, heat transfer devices [e.g., electric boiler (EB)], cooling storage devices (e.g., TES), and heat storage (HS) devices (e.g., CS) all have corresponding device connection subspaces. Assume that there are w types of heat transfer devices and y types of heat/cooling storage energy devices, then $\mathbb{N}_{\text{connection}}$ can be expressed as

$$\mathbb{N}_{m+1+1} \cup \cdots \cup \mathbb{N}_{m+1+w} \cup \cdots \cup \mathbb{N}_{m+1+w+y} = \mathbb{N}_{\text{connection}}. \tag{29}$$

Therefore, the structure strings $\alpha = \{\alpha_l | l = 1, 2, \ldots, L\}$ consists of L decimal numbers, $\alpha_l \in \mathbb{N}_l$, where $L = m + 1 + w + y$.

2) Search Policy: A RL-based search policy for MES structure is proposed. The search policy determines which rules to explore $\mathbb{N}$ by interacting with the second level continuously, and obtains the optimal structure α^*.

First, a policy function based on recurrent neural network (RNN) architecture is designed as agent of search policy, i.e., $\pi_\theta(A|S)$. The action A of $\pi_\theta(A|S)$ can be used to describe the structure strings, i.e., α, and is expressed as

$$A^k_{1:L} = [\alpha^k_1, \alpha^k_2, \ldots, \alpha^k_L] \tag{30}$$

where k represents the kth iteration, $A^k_{1:L}$ means that action A is a list of action sequences of length L, and $\alpha^k_l \in \mathbb{N}_l$. The system structure of a MES can be predicted flexibly by $\pi_\theta(A|S)$. For example, $\pi_\theta(A|S)$ can sample a reasonable selection scheme of heat supply devices from the subspace $\mathbb{N}_2$ with probability p_2, as shown in Figs. 2 and 3.

$\pi_\theta(A|S)$ takes the system structure A evaluated in second level during the last iteration as the kth state S, i.e.,

$$S^k_{1:L} = [\alpha^{k-1}_1, \alpha^{k-1}_2, \ldots, \alpha^{k-1}_L] \tag{31}$$

where the lth state S^k_l has the same connotation and mathematical representation as action α^{k-1}_l. Therefore, $\pi_\theta(A|S)$ is designed according to the RNN–based network to process the modeling requirements from state sequence to action sequence, as shown in Algorithm 2 and Fig. 3.

Then, $\pi_\theta(A|S)$ is trained based on policy gradient algorithm. The kth reward signal R_k, which is related to the system economic cost of the MES, is defined as

$$R^k = \frac{\Gamma_u - J^k_{\text{MES}}}{\Gamma_u - \Gamma_l} \tag{32}$$

where J^k_{MES} is the minimum annual total cost of the kth system structure sampled from $\mathbb{N}$ and Γ_u and Γ_l are the upper and lower bounds of the annual total cost, which are set based on the cost of the separate production system and adjusted according to experience.

Specifically, to search for the most economical system structure, i.e., the optimal α^*, $\pi_\theta(A|S)$ should maximize the expected

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

Algorithm 2: Forward Propagation of $\pi_\theta(A|S)$.

Require:
Current state, i.e., $S^k_{1:L}$.
Compute:
h_0 is initialized with an all-zero vector;
$x_0 \leftarrow S^k_{1:L}$;
for $l = 1, ..., L$ **do**
$h_l \leftarrow \tanh\left(W^h_l\left[x_{l-1}; h_{l-1}\right] + b^h_l\right)$;
$s_l \leftarrow \text{softmax}\left(W^x_l h_l + b^x_l\right)$;
$c_l \leftarrow W^s_l s_l + b^s_l$;
$p_l \leftarrow \text{softmax}\left(c_l\right)$;
$\alpha_l \leftarrow \text{argmax}\left(p_l\right)$;
$o_l \leftarrow \text{one} - \text{hot}\left(\alpha_l\right)$;
$x_l \leftarrow \text{embedding}\left(o_l\right)$.
end
Return:
Next action, i.e., $A^k_{1:L} \leftarrow [\alpha_1, \alpha_2, ..., \alpha_L]$.

reward

$$G(\theta) = E_{P(A_{1:L};\theta)}[R] \tag{33}$$

where E is expectation. Then, the parameter θ is updated by the gradient ascent algorithm

$$\theta_{k+1} = \theta_k + \eta \nabla_\theta G(\theta) \tag{34}$$

where η is learning rate and $\nabla_\theta G(\theta)$ is the strategy gradient of $G(\theta)$ that can be estimated by

$$\nabla_\theta G(\theta) = \frac{1}{k}\sum_{i=1}^{k}\sum_{l=1}^{L}\frac{\partial \log \pi\left(A_l|A_{(l-1):1};\theta\right)}{\partial\theta}\cdot R^i. \tag{35}$$

To further reduce the variance of the strategy gradient estimate, the strategy gradient can be unbiased estimated as

$$\nabla_\theta G(\theta) = \frac{1}{k}\sum_{i=1}^{k}\sum_{l=1}^{L}\frac{\partial \log \pi\left(A_l|A_{(l-1):1};\theta\right)}{\partial\theta}\cdot\left(R^i - b\right) \tag{36}$$

where the baseline function b is the exponential moving average of the best economic target of the previous structures.

C. Capacity–Operation Co–Optimization

The capacity configuration and operational scheme of each selected device for current system structure α^k from the search policy in first level are co-optimized. A common economic optimization objective for a given system structure is adopted [7], [9]

$$(\beta^*, \gamma^*, \psi^*) = \arg\min_{\beta,\gamma,\psi} \quad \mathcal{J}_{\text{MES}}\left(\alpha^k; \beta, \gamma, \psi\right)$$
$$\text{s.t.} \quad (9-12, 14-24). \tag{37}$$

Although the structure α^k is known, the charging and discharge constraints of the energy storage devices are still nonlinear. Therefore, the Big-M method is adopted to equivalent linearize the constraints and make the capacity–operation co–optimization a LP problem

$$0 \le \overline{\gamma}^{a,d}_i(t) \le \alpha_i\lambda_i\gamma_i, 0 \le \gamma^{a,d}_i(t) \le \alpha_i\lambda_i\gamma_i \quad \forall i \quad \forall t, \tag{38}$$

TABLE I
TIME-OF-USE ELECTRICITY PRICES

Time period	Purchasing	Selling
10:00-13:00;18:00-23:00	0.925 CNY/kWh	0.4 CNY/kWh
7:00-10:00;13:00-18:00	0.614 CNY/kWh	0.3 CNY/kWh
23:00-7:00	0.402 CNY/kWh	0.2 CNY/kWh

TABLE II
EFFICIENCY AND COST PARAMETERS OF ALTERNATIVE EQUIPMENT

Device	Efficiency	Investment cost	Maintenance cost	Lifespan
PV	-	3000 CNY/kW	40 CNY/kW	25 years
WT	-	3500 CNY/kW	37 CNY/kW	20 years
PGU	0.35[1] 0.55[2]	2000 CNY/kW	45 CNY/kW	20 years
EB	0.9	800 CNY/kW	50 CNY/kW	20 years
GB	0.9	330 CNY/kW	38 CNY/kW	20 years
HP	3.0	3000 CNY/kW	50 CNY/kW	20 years
EC	3.0	1000 CNY/kW	35 CNY/kW	20 years
AC	0.9	1200 CNY/kW	30 CNY/kW	20 years
ES	0.95	1250 CNY/kWh	100 CNY/kWh	10 years
HS	0.95	130 CNY/kWh	5 CNY/kWh	12 years
CS	0.95	130 CNY/kWh	5 CNY/kWh	12 years

[1] 0.35 is electrical efficiency.
[2] 0.55 is thermal efficiency.

$$0 \le \overline{\gamma}^{a,d}_i(t) \le \alpha_i\psi_{i,+}(t)M, 0 \le \gamma^{a,d}_i(t) \le \alpha_i\psi_{i,-}(t)M \quad \forall i \quad \forall t \tag{39}$$

where the subscript i represents energy storage device i and M is a larger constant. The optimal capacity configuration and operational scheme of the MES, i.e., $\beta^*, \gamma^*, \psi^*$, can be easily solved by using widely used software tools, such as GAMS, MOSEK, and CPLEX, for quick solution [31], and, the optimal system economic cost $\mathcal{J}_{\text{MES}}$ of the current system structure α^k is fed back to the search policy in the first level as the reward R.

IV. Case Study

A. Setup

1) Settings: The effectiveness of the proposed framework was evaluated by using typical data for the renewable power generation and building loads from a northern Chinese city to develop three scenarios: school (case 1), residential area (case 2), and industrial park (case 3). The load demands had obvious seasonal differences. The K-means algorithm was used to cluster load demands and renewable energy data throughout the year. The electricity, heat, and cooling demand patterns for five typical days in each case are shown in Fig. 4. The time–of–use electricity prices were taken from Wu et al. [21] and shown in Table I. The price of natural gas was 2.9 CNY/Nm^3, and the unit carbon tax was 0.3 CNY/kg [32]. The key economic and technical parameters of the alternative devices were taken from Zhang et al. [20] and shown in Table II. Parameters related to the MES were taken from Ke et al. [33]. The trends of photovoltaics (PV) and wind turbines (WT) in each case are described in Fig. 5: case 1 did not include PV or WT while case 2 included only PV.

2) Comparison Methods: Different comparison methods are used to verify the advancement and effectiveness of the proposed framework, and are described as follows.

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

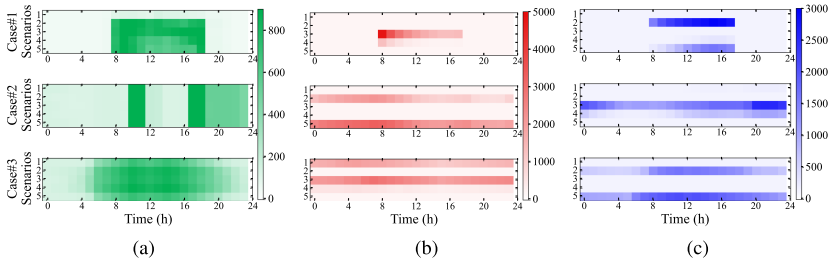


Fig. 4. Load data for five typical days of the year in each case. (a) Electricity loads. (b) Heat loads. (c) Cooling loads.

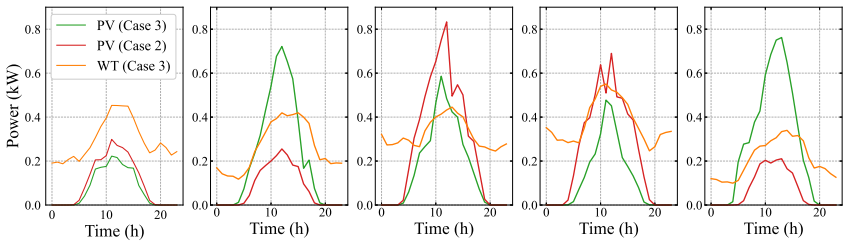


Fig. 5. Renewable energy generation data for five typical days in cases 2 (residential area) and 3 (industrial park).

TABLE III
DESIGN RESULTS OF DIFFERENT METHODS IN CASES 1–3

Cases		Cost ($\times 10^4$ CNY)	PV (kW)	WT (kW)	PGU (kW)	EB (kW)	GB (kW)	HP (kW)	EC (kW)	AC (kW)	ES (kWh)	HS (kWh)	CS (kWh)
1	*IIDF*	**270.48**	/	/	1001	/	1066	/	1182	1416	/	2902	521
	TID	271.16	/	/	1004	/	1061	/	995	1420	/	2903	1729
	TID-S	271.87	/	/	987	/	1088	/	1371	1433	/	2902	/
2	*IIDF*	**478.75**	400	/	899	16	/	1618	/	622	11	1082	464
	TID	481.43	400	/	892	41	/	1551	/	654	123	758	645
	TID-S	551.58	400	/	756	/	2547	/	1477	917	/	981	/
3	*IIDF*	**527.62**	300	300	754	/	67	1138	10	756	/	981	63
	TID	528.02	300	300	747	/	116	1120	/	787	/	856	61
	TID-S	596.29	300	300	557	/	1712	/	1091	865	/	588	/

1) *Interactive integrated design framework (IIDF)* is the method proposed in this study. In *IIDF*, the policy function was set to $L = 7$ and $\eta = 0.001$. During training, the maximum number of iterations is set to 300 and the Adam optimizer is adopted.
2) *TID* means the traditional integrated design method using the hierarchical model in [19]. In *TID*, GA is used to design the system structure of MES, and the capacity and operation are optimized as an LP problem like *IIDF*. The GA was set to a maximum of 50 generations each with a population size of 50. The optimal value does not change significantly within 10 consecutive generations, and stop iteration.
3) *TID-S* means the *TID* with a given structure that refers to [34], which solves the LP problem directly.

All methods adopt the MATLAB toolbox YALMIP with the CPLEX solver to conduct LP problem. And, all the comparison experiments were carried out on a desktop computer with an i7-8700 Intel processor, equipped with 8 GB of RAM.

B. Comparative Analysis

Table III presents the MES designs of the three methods in cases 1–3. In general, the different methods incorporated PV or WT in cases they were available and maximized their capacity to exploit their advantages in terms of economics and carbon emission reduction. Because of the obvious characteristics of the cooling and heat loads, the PGU and AC were the key components to realize energy cascades and were configured with larger capacities in all cases. Regarding the energy storage devices, HS and CS were better options than ES because they had lower initial costs, which increased the flexibility of the electric and heat outputs of the MES. These results indicate that the three methods all produced reasonable system designs.

In all cases, *TID* and *IIDF* designed MESs that performed better than the one designed by *TID-S*. This means that a MES designed according to experience is not necessarily the most economical despite it being reasonable and feasible subjectively. *IIDF* had the lowest $\mathcal{J}_{\text{MES}}$ in all cases, but the differences in cost between the three methods differed according to the case. Notably, in case 2, *IIDF* reduced $\mathcal{J}_{\text{MES}}$ compared with *TID* and *TID-S* by 26 800 CNY (0.56%) and 728 300 CNY (13.20%), respectively.

Although *TID* and *IIDF* selected the same devices, the MES designed by *IIDF* performed better than the MES designed by *TID*. This is because the connections between the devices differed, which affected the capacity configuration and operational

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

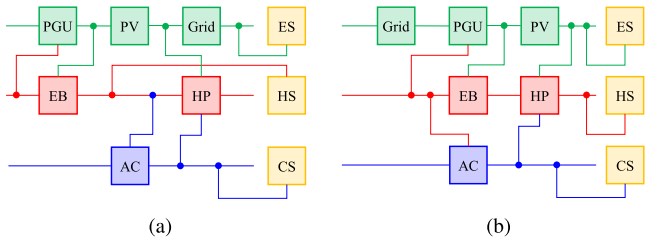


Fig. 6. Design results in case 2: (a) *IIDF* and (b) *TID*.

TABLE IV
OPERATIONAL COSTS FOR ALL METHODS

Scenario		1 (CNY)	2 (CNY)	3 (CNY)	4 (CNY)	5 (CNY)
case 1	*IIDF*	**759**	**12677**	13554	**7905**	**8076**
	TID	**759**	12744	13556	**7905**	8120
	TID-S	**759**	12703	**13547**	**7905**	**8076**
case 2	*IIDF*	6376	**11442**	**13026**	**6417**	**18605**
	TID	**6328**	11475	13226	6525	18699
	TID-S	6401	14963	12680	6457	26455
case 3	*IIDF*	**11546**	10985	**17731**	**8265**	**12023**
	TID	11548	10984	17780	8275	12040
	TID-S	15837	**10981**	24800	8334	12040

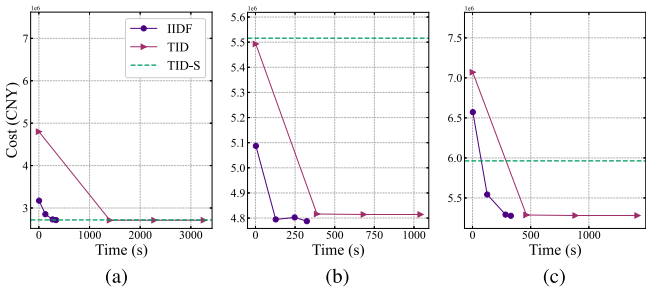


Fig. 7. Convergence of different methods for cases 1–3: (a) Case 1, (b) Case 2, and (c) Case 3.

scheme. For example, in case 2 the *IIDF* placed the HS in front of the AC while the *TID* placed the HS behind all heat supply devices, as shown in Fig. 6. Therefore, the stored heat energy cannot be further converted into energy by the *TID*, which affected the operational flexibility and economic cost. Overall, *TID* could only design a structure with relatively fixed connections between devices due to the limitations of the preset hierarchical model. For example, all energy storage devices could only be placed at the end of each energy flow. In contrast, *IIDF* could flexibly search for an arbitrary and reasonable structure for the MES that improved its performance.

Table IV presents the operational costs for all methods, demonstrating that the MESs designed by IIDF exhibit the best operational performance and economic efficiency.

In three different cases, the proposed *IIDF* required only 338.11 s, 324.07 s, and 329.50 s to solve the optimal design schemes of the MESs, whereas the *TID* method required 3255.82 s, 1041.74 s, and 1417.83 s, as shown in Fig. 7. The computation time of the proposed *IIDF* is reduced by approximately 89.62%, 68.89%, and 76.76% compared to the *TID* method, respectively. The main reason is that the proposed *IIDF* determines the best system structure in a relatively small search

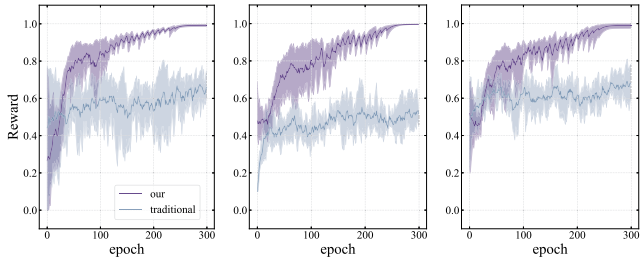


Fig. 8. Convergence of rewards in the ablation experiment for cases 1–3: (a) Case 1, (b) Case 2, and (c) Case 3.

TABLE V
RESULTS OF $\mathcal{J}_{\text{MES}}$ IN THE ABLATION EXPERIMENT

Search Space	Case 1	Case 2	Case 3
our (CNY)	**2,704,800**	**4,787,500**	**5,276,200**
traditional (CNY)	3,151,070	5,292,873	5,829,170

space compared with *TID*. This means that the proposed *IIDF* can accurately search for the best system design in terms of performance and economy in a very short time and demonstrated better computational efficiency than *TID*.

C. Ablation Experiment

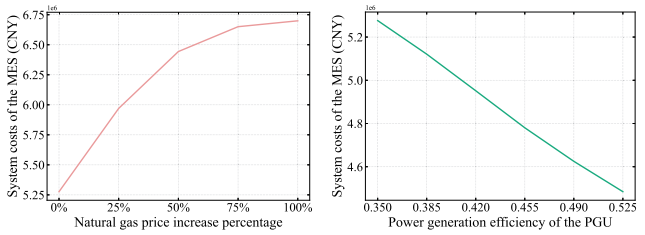
The ablation experiment was conducted to gather evidence of the necessity of fusing domain knowledge and quantifying its contribution to the performance of the proposed *IIDF*. As shown in Fig. 8, the performance of *IIDF* decreased substantially after removing two domain knowledge components from search space. It is obvious that the *IIDF* with domain knowledge can achieve rapid convergence, while the *IIDF* without domain knowledge cannot find the optimal MES scheme in 300 iterations despite a converging trend. Compared to ablation *IIDF*, *IIDF* with domain knowledge reduced $\mathcal{J}_{\text{MES}}$ in all cases by 14.16%, 9.51% and 9.48%, respectively, as shown in Table V. Thus, the domain knowledge was necessary for *IIDF* to design MESs.

D. Sensitivity Analysis

Taking the industrial park scenario (i.e., Case 3) as an example, sensitivity experiments and analyses were conducted to quantitatively investigate the impact of key technical and economic parameters on the economic performance of the MES.

The operational strategies of the MES devices are predominantly determined by energy prices. The increase in natural gas procurement costs may emerge as a critical factor negatively impacting the economic efficiency of the MES, especially in light of rising energy demand in the future. As the price of natural gas continues to rise, the annualized total cost of the MES (i.e., $\mathcal{J}_{\text{MES}}$) will persistently increase, as shown in Fig. 9(a). When the natural gas price rises to 75% of the current price (i.e., 5.075 CNY/Nm^3), the $\mathcal{J}_{\text{MES}}$ will increase by 20.66%. As natural gas is extensively used as an energy carrier for the PGU, rising prices cause a notable decline in PGU capacity. When the natural gas price doubles, the capacity of the PGU decreases to

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.



(a) (b)

Fig. 9. Sensitivity analysis: (a) natural gas price and (b) investment cost of PGU.

merely 77 kW. Correspondingly, the drastic reduction in waste heat recovery leads to an increase in the HP capacity to 2241 kW.

With technological advancements, the efficiency of some key equipment is expected to improve. This study specifically examines the impact of increased power generation efficiency of the PGU on the economic performance and design scheme of the MES. When the PGU's power generation efficiency increases from 0.350 to 0.525, its corresponding optimal capacity also rises from 754 to 834 kW. It should be noted that as the power generation efficiency improves, the waste heat recovery efficiency of the PGU decreases accordingly. This results in a reduction of the optimal capacity of the AC from 756 to 463 kW. However, from a systemic perspective, the economic performance of the MES still improves with the increase in PGU's power generation efficiency. When the PGU's power generation efficiency increases by 50%, the annualized total cost of the MES decreases by 15%, as shown in Fig. 9(b). In conclusion, while technological progress may benefit the economic performance of the MES, its impact is significantly less pronounced compared to that of energy prices.

V. Conclusion

This study proposed a RL–based bilevel interactive integrated design framework (i.e., *IIDF*) for the structure, capacity and operation of MESs. We constructed a global feasible domain for MES structure by adding equipment position variables, and it is not limited by the preset hierarchical model. In first level of *IIDF*, The search space is proposed to narrow the constructed feasible domain and ensure the reasonable searched structure by fusing domain knowledge. On this basis, the RL–based search strategy is proposed to determine the optimal MES structure. Moreover, the strategy function is cleverly designed based on RNN, which can flexibly generate MES structure from the search space. In the second level, the capacity–operation co–optimization is executed for the determined MES structure. The optimized target is regarded as a reward feedback to the first level, and the performance of the search policy is improved by the strategy gradient algorithm. The case study has demonstrated that *IIDF* can make full use of prior knowledge, such as multienergy flow supply and demand balance, and realize the optimal design of MES under different energy demand scenarios. In addition, *IIDF* has significant computational competitiveness, which is obvious in comparison with the traditional methods. Future research aims to enhance the adaptability of *IIDF* to different regional scales, especially the design of regional MES.

References

[1] M. S. Misaghian, G. Tardioli, A. G. Cabrera, I. Salerno, D. Flynn, and R. Kerrigan, "Assessment of carbon-aware flexibility measures from data centres using machine learning," *IEEE Trans. Ind. Appl.*, vol. 59, no. 1, pp. 70–80, Jan./Feb. 2023.

[2] S. Chen, Z. Wei, G. Sun, W. Wei, and D. Wang, "Convex hull based robust security region for electricity-gas integrated energy systems," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1740–1748, May 2019.

[3] N. Liu, L. Tan, H. Sun, Z. Zhou, and B. Guo, "Bilevel heat-electricity energy sharing for integrated energy systems with energy hubs and prosumers," *IEEE Trans. Ind. Inform.*, vol. 18, no. 6, pp. 3754–3765, Jun. 2022.

[4] H. Han, H. Zhang, J. Yang, and H. Su, "Distributed model predictive consensus control for stable operation of integrated energy system," *IEEE Trans. Smart Grid*, vol. 15, no. 1, pp. 381–393, Jan. 2024.

[5] M. Mohammadi, Y. Noorollahi, B. Mohammadi-Ivatloo, and H. Yousefi, "Energy hub: From a model to a concept-a review," *Renew. Sustain. Energ. Rev.*, vol. 80, pp. 1512–1527, 2017.

[6] S. Klyapovskiy, S. You, H. Cai, and H. W. Bindner, "Integrated planning of a large-scale heat pump in view of heat and power networks," *IEEE Trans. Ind. Appl.*, vol. 55, no. 1, pp. 5–15, Jan./Feb. 2019.

[7] F. Li, B. Sun, C. Zhang, and C. Liu, "A hybrid optimization-based scheduling strategy for combined cooling, heating, and power system with thermal energy storage," *Energy*, vol. 188, 2019, Art. no. 115948.

[8] N. Zhao, W. Gu, Z. Zheng, and T. Ma, "Multi-objective bi-level planning of the integrated energy system considering uncertain user loads and carbon emission during the equipment manufacturing process," *Renew. Energy*, vol. 216, 2023, Art. no. 119070.

[9] F. Fang, Q. H. Wang, and Y. Shi, "A novel optimal operational strategy for the CCHP system based on two operating modes," *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 1032–1041, Feb. 2011.

[10] X.-Y. Ren, Z.-H. Wang, and L.-L. Li, "Multi-objective optimization and evaluation of hybrid combined cooling, heating and power system considering thermal energy storage," *J. Energy Storage*, vol. 86, 2024, Art. no. 111214.

[11] F. Han, J. Zeng, J. Lin, C. Gao, and Z. Ma, "A novel two-layer nested optimization method for a zero-carbon island integrated energy system, incorporating tidal current power generation," *Renew. Energy*, vol. 218, 2023, Art. no. 119381.

[12] S. Geng, M. Vrakopoulou, and I. A. Hiskens, "Optimal capacity design and operation of energy hub systems," *Proc. IEEE*, vol. 108, no. 9, pp. 1475–1495, Sep. 2020.

[13] Y. Deng et al., "A novel operation strategy based on black hole algorithm to optimize combined cooling, heating, and power-ground source heat pump system," *Energy*, vol. 229, 2021, Art. no. 120637.

[14] M. Mohammadi, Y. Noorollahi, B. Mohammadi-Ivatloo, and H. Yousefi, "Energy hub: From a model to a concept–a review," *Renew. Sustain. Energ. Rev.*, vol. 80, pp. 1512–1527, 2017.

[15] Y. Xu, C. Yan, H. Liu, J. Wang, Z. Yang, and Y. Jiang, "Smart energy systems: A critical review on design and operation optimization," *Sustain. Cities Soc.*, vol. 62, 2020, Art. no. 102369.

[16] Y. Li, C. Liu, L. Zhang, and B. Sun, "A partition optimization design method for a regional integrated energy system based on a clustering algorithm," *Energy*, vol. 219, 2021, Art. no. 119562.

[17] M. Ameri and Z. Besharati, "Optimal design and operation of district heating and cooling networks with CCHP systems in a residential complex," *Energy Build.*, vol. 110, pp. 135–148, 2016.

[18] Z. Liu, M. Zeng, H. Zhou, and J. Gao, "A planning method of regional integrated energy system based on the energy hub zoning model," *IEEE Access*, vol. 9, pp. 32161–32170, 2021.

[19] W. Huang, N. Zhang, J. Yang, Y. Wang, and C. Kang, "Optimal configuration planning of multi-energy systems considering distributed renewable energy," *IEEE Trans. Smart Grid*, vol. 10, no. 2, pp. 1452–1464, Feb. 2019.

[20] Z. Zhang et al., "Combining agent-based residential demand modeling with design optimization for integrated energy systems planning and operation," *Appl. Energy*, vol. 263, 2020, Art. no. 114623.

[21] T. Wu, S. Bu, X. Wei, G. Wang, and B. Zhou, "Multitasking multi-objective operation optimization of integrated energy system considering biogas-solar-wind renewables," *Energy Convers. Manag.*, vol. 229, 2021, Art. no. 113736.

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.

[22] X. Liu, H. Yan, W. Zhou, N. Wang, and Y. Wang, "Event-triggered optimal tracking control for underactuated surface vessels via neural reinforcement learning," *IEEE Trans. Ind. Inform.*, vol. 20, no. 11, pp. 12837–12847, Nov. 2024.
[23] F. Kienzle, P. Ahčin, and G. Andersson, "Valuing investments in multi-energy conversion, storage, and demand-side management systems under uncertainty," *IEEE Trans. Sustain. Energy*, vol. 2, no. 2, pp. 194–202, Feb. 2011.
[24] Y. Ye, D. Qiu, X. Wu, G. Strbac, and J. Ward, "Model-free real-time autonomous control for a residential multi-energy system using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3068–3082, Apr. 2020.
[25] B. Zhang et al., "Dynamic energy conversion and management strategy for an integrated electricity and natural gas system with renewable energy: Deep reinforcement learning approach," *Energy Convers. Manag.*, vol. 220, 2020, Art. no. 113063.
[26] M. Chen, Z. Shen, L. Wang, and G. Zhang, "Intelligent energy scheduling in renewable integrated microgrid with bidirectional electricity-to-hydrogen conversion," *IEEE Trans. Netw. Sci. Eng.*, vol. 9, no. 4, pp. 2212–2223, Apr. 2022.
[27] M. Razghandi, H. Zhou, M. Erol-Kantarci, and D. Turgut, "Smart home energy management: VAE-GAN synthetic dataset generator and Q-learning," *IEEE Trans. Smart Grid*, vol. 15, no. 2, pp. 1562–1573, Feb. 2024.
[28] A. R. Sayed, X. Zhang, G. Wang, J. Qiu, and C. Wang, "Online operational decision-making for integrated electric-gas systems with safe reinforcement learning," *IEEE Trans. Power Syst.*, vol. 39, no. 2, pp. 2893–2906, Feb. 2024.
[29] J. Chen et al., "Multi-timescale reward-based DRL energy management for regenerative braking energy storage system," *IEEE Trans. Transport. Electrific.*, early access, Jan. 10, 2025, doi: 10.1109/TTE.2025.3528255.
[30] S. Zhou et al., "Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach," *Int. J. Electr. Power Energy Syst.*, vol. 120, 2020, Art. no. 106016.
[31] H. Xiao, W. Pei, Z. Dong, and L. Kong, "Bi-level planning for integrated energy systems incorporating demand response and energy storage under uncertain environments using novel metamodel," *CSEE J. Power Energy Syst*, vol. 4, no. 2, pp. 155–167, 2018.
[32] L. Kang et al., "Effects of load following operational strategy on cchp system with an auxiliary ground source heat pump considering carbon tax and electricity feed in tariff," *Appl. Energy*, vol. 194, pp. 454–466, 2017.
[33] Y. Ke, H. Tang, M. Liu, Q. Meng, and Y. Xiao, "Optimal sizing for wind-photovoltaic-hydrogen storage integrated energy system under intuitionistic fuzzy environment," *Int. J. Hydrogen Energy*, vol. 48, no. 88, pp. 34193–34209, 2023.
[34] J. Wang, Z. J. Zhai, Y. Jing, and C. Zhang, "Particle swarm optimization for redundant building cooling heating and power system," *Appl. Energy*, vol. 87, no. 12, pp. 3668–3679, 2010.

Hui Zhang received the B.S. degree in internet of things engineering from Shandong Jianzhu University, Jinan, China, in 2022. He is currently working toward the Ph.D. degree in control engineering with the School of Control Science and Engineering, Shandong University, Jinan.

His current research interests include hydrogen safety, machine learning, and computational intelligence.

Lizhi Zhang received the B.S. degree in automation from Qingdao University, Qingdao, China, in 2016, and the M.S. degree in control engineering and the Ph.D. degree in control theory and control engineering from Shandong University, Jinan, China, in 2019 and 2024, respectively.

He is currently with the Shandong Police College, Jinan. His research interests include integrated energy system planning and optimization.

Haozeng Bie is currently working toward the B.S. degree in computer science and technology with the School of Mechanical, Electrical and Information Engineering, Shandong University, Weihai, China.

His current research interests include software engineering, computer organization principles, and algorithm design.

Zhiwei Xu received the B.S. degree in mathematics and applied mathematics from Northwest Normal University, Lanzhou, China, in 2015, and the M.S. degree in control science and engineering from Central South University, China, in 2018, and the Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2022, respectively.

He is currently a Postdoctoral Research Fellow with the School of Control Science and Engineering, Shandong University, Jinan, China. His current research interests include data-driven optimization, model-based learning control, reinforcement learning, distributed optimization, and their applications to cooperative wind farm control.

Guangyao Fan received the M.S. degree in energy and power from North China Electric Power University, Baoding, China, in 2023. He is currently working toward the Ph.D. degree in control theory and control engineering with Shandong University, Jinan, China.

His current research interests include modeling, planning, and optimization of hydrogen energy systems.

Bin Jia received the M.S. degree in power electronics and electric drive in 2020 from Shandong University, Jinan, China, where he is currently working toward the Ph.D. degree in power electronics and electric drive.

His current research interests include optimization of integrated energy systems and deep reinforcement learning.

Bo Sun (Member, IEEE) was born in Shandong Province, China, in 1982. He received the B.S. degree in automation from Shandong University, Jinan, China, in 2004, and the Ph.D. degree in control theory and control engineering from Shandong University, Jinan, China, in 2009.

In 2010, he joined Shandong University, where he is currently a Professor with the School of Control Science and Engineering. His research interests include the optimal control of engineering and the optimization of integrated energy systems.

Authorized licensed use limited to: SHANDONG UNIVERSITY. Downloaded on July 12,2025 at 02:43:45 UTC from IEEE Xplore. Restrictions apply.