

A Hybrid Data-driven Approach Integrating Temporal Fusion Transformer and Soft Actor-critic Algorithm for Optimal Scheduling of Building Integrated Energy Systems

Ze Hu, *Student Member, IEEE*, Peijun Zheng, *Member, IEEE*, Ka Wing Chan, Siqi Bu, *Senior Member, IEEE*, Ziqing Zhu, *Member, IEEE*, Xiang Wei, and Yosuke Nakanishi, *Member, IEEE*

Abstract—Building integrated energy systems (BIESs) are pivotal for enhancing energy efficiency by accounting for a significant proportion of global energy consumption. Two key barriers that reduce the BIES operational efficiency mainly lie in the renewable generation uncertainty and operational non-convexity of combined heat and power (CHP) units. To this end, this paper proposes a soft actor-critic (SAC) algorithm to solve the scheduling problem of BIES, which overcomes the model non-convexity and shows advantages in robustness and generalization. This paper also adopts a temporal fusion transformer (TFT) to enhance the optimal solution for the SAC algorithm by forecasting the renewable generation and energy demand. The TFT can effectively capture the complex temporal patterns and dependencies that span multiple steps. Furthermore, its forecasting results are interpretable due to the employment of a self-attention layer so as to assist in more trustworthy decision-making in the SAC algorithm. The proposed hybrid data-driven approach integrating TFT and SAC algorithm, i.e., TFT-SAC approach, is trained and tested on a real-world dataset to validate its superior performance in reducing the energy cost and computational time compared with the benchmark approaches. The generalization performance for the scheduling policy, as well as the sensitivity analysis, are examined in the case studies.

Index Terms—Building integrated energy system (BIES), hybrid data-driven approach, time-series forecast, optimal scheduling, soft actor-critic (SAC), temporal fusion transformer (TFT).

I. INTRODUCTION

THE rapid development in industry and urban areas has led to significant changes in energy systems, resulting

in high renewable penetration and challenges for sustainable development. With buildings accounting for about 40% of global energy consumption, it is crucial to enhance the efficiency of building energy systems for meeting rising energy demands and supporting sustainability [1]. Energy integration provides a way to improve the operation efficiency in the distribution-level energy system [2]. By coordinating multiple energies including power, gas, and heat, the building integrated energy systems (BIESs) can achieve sufficient renewable usage and obtain more abundant flexibility. Therefore, an effective energy management in BIES becomes vital for improving the operational flexibility and maximizing the renewable energy use in the whole energy system.

However, the optimal operation of BIES is hindered by two key challenges: ① the high operational risk due to the intermittent and uncertain nature of photovoltaic (PV) generation and energy demand [3], and ② intractable optimization caused by non-convexity of combined heat and power (CHP) unit [4]. For the former, the PV generation and energy demand uncertainties have been shown to bring significant profit losses and endanger the system stability by leading to energy shortage or renewable curtailment [5]. This problem even gets more severe in large buildings with high peak demands or high solar capacities. The accurate forecast of PV generation and energy demand is thus crucial for smart scheduling of energy devices (e.g., energy storage) to avoid profit losses and system blackouts. As for the latter, the CHP unit is well known for providing flexibility in power and heat in a feasible operation region (FOR), which is non-convex in practice and makes the optimization non-tractable. The FOR convexification is a widely adopted solution but sacrifices considerable operational flexibility [6]. The optimal scheduling of CHP unit remains an open question in the optimal operation of BIESs. Moreover, the variable renewable/demand forecast and non-convex operation optimization are not independent of each other, e.g., the flexible scheduling of CHP unit can provide compensation for the renewable uncertainty. This indicates a deep correlation between the forecast and non-convex scheduling in BIESs.

BIESs have been extensively studied, particularly in the

Manuscript received: August 19, 2024; revised: November 14, 2024; accepted: January 2, 2025. Date of CrossCheck: December 30, 2024. Date of online publication: January 28, 2025.

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>).

Z. Hu, K. W. Chan, S. Bu, Z. Zhu, and X. Wei are with the Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China (e-mail: zzzed.hu@connect.polyu.hk; siqi.bu@polyu.edu.hk; kevin.kw.chan@polyu.edu.hk; ziqing-yancy.zhu@polyu.edu.hk; weixiang0610@outlook.com).

P. Zheng (corresponding author) and Y. Nakanishi are with the Graduate School of Environment and Energy Engineering, Waseda University, Tokyo, Japan (e-mail: zhengpeijun@akane.waseda.jp; nakanishi-yosuke@waseda.jp).

DOI: 10.35833/MPCE.2024.000909



areas of scheduling [7], [8] and expansion planning [9], [10]. Much of the existing research has focused on developing model-based frameworks for the optimal operation in multi-carrier energy systems. These optimization problems generally rely on precise models and estimated exogenous factors such as weather-dependent renewable generation and energy loads. To address uncertainties, approaches like robust optimization (RO) and stochastic optimization (SO) have been used, where RO models the uncertainties as bounded sets, and SO uses a set of scenarios to represent the uncertainties.

While these conventional approaches are effective in managing the scheduling of multi-carrier energy systems, they face challenges in handling highly nonlinear units, particularly in competitive markets. Stochastic programming (SP) becomes inefficient as the number of scenarios increases, and RO often yields overly conservative results by focusing on the worst-case scenarios. Besides, both SP and RO suffer from the curse of dimensionality, where the increased actions, decision variables, and constraints lead to exponentially growing computational requirements, limiting their scalability for real-world applications involving multiple devices and uncertainties [11].

Reinforcement learning (RL) presents an innovative alternative that effectively addresses the above limitations by providing a means to tackling dynamic and sequential decision-making challenges [12]. Unlike classical optimization approaches, which are typically static and require the modeling of joint uncertainty distributions across time periods, RL models the optimization problems as Markov decision processes (MDPs). This makes RL well-suited for navigating complex, nonlinear, and uncertain environments, allowing for real-time adaptation and learning. Moreover, the training process of RL agents is grounded in dynamic programming (DP) to make real-time sequential decisions and consider long-term returns [11]. It can efficiently utilize increasing data from the environment, adapting to various state conditions and capturing system uncertainties [13].

Furthermore, by incorporating deep neural network (DNN), deep reinforcement learning (DRL) algorithms like deep deterministic policy gradient (DDPG) and twin delayed DDPG (TD3) can generate continuous actions and estimate the non-convex value functions. DRL algorithms outperform traditional RL algorithms and mathematical programming in solving optimization problems, offering lower computational burden and better applicability in real-world scenarios [14]. Therefore, DRL algorithms have been adopted and considered effective tools for most optimization and decision-making problems including scheduling of integrated energy systems [15], [16], optimal power flow calculation [17], [18], and voltage control [19], [20], etc.

In the context of scheduling problems of BIESs, DRL algorithms receive available information to make operational decisions. The scheduling is based on day-ahead/hour-ahead prediction for required variables including renewable generation, energy demand, etc. Although some DRL algorithms can learn from the current state to make decisions, there is no explicit forecasting procedure in the design of DRL algorithms, resulting in a poor ability to deal with future uncer-

tainties. Integrating decision-making with forecast for a holistic operational tool is a neutral idea to improve the operational efficiency. Recently, some literature has tended to integrate decision-making with forecast as a holistic data-driven tool for scheduling of integrated energy systems. For instance, [15] adopts a long short-term memory (LSTM) method to extract temporal features and assist the decision-making of DRL algorithms in integrated energy management. Reference [12] combines a convolutional neural network (CNN) and bidirectional LSTM (BLSTM) to forecast PV generation in an energy hub by analyzing sky images. The predicted value is then imported into the DDPG for further scheduling decision-making. Although these approaches have shown good performance, the LSTM struggles with capturing complex temporal patterns and dependencies that span multiple time steps effectively [21], and related research is still limited.

The efficient scheduling of a BIES with handling non-convexity and uncertainties presents three major challenges. ① Traditional optimization approaches face significant difficulties in solving the operational optimization problem of BIES due to the inherent non-convexity of the devices. Moreover, as the system size increases, these approaches often become computationally prohibitive. ② Existing research on scheduling problems of BIES seldom integrates renewable energy forecasts with decision-making processes using data-driven approaches. Consequently, such comprehensive approaches remain underdeveloped and lack adaptability for specific BIES applications. ③ Many studies employ DRL algorithms in conjunction with black-box forecasting tools, raising concerns about the model transparency and reliability. The opacity of these algorithms can lead to significant profit losses [22], thereby limiting the real-world applicability of data-driven approaches.

To this end, our research addresses these gaps by integrating the TFT for accurate forecast with the SAC algorithm for robust operation. The main contributions of this paper are as follows.

1) This paper presents a detailed decision-making model for BIES, including micro-CHP unit, battery energy storage systems (BESSs), PV panels, and gas boilers (GBs). The non-convex scheduling problem is formulated into an optimization problem and then reformulated into an MDP for the application of RL algorithms.

2) This paper proposes a hybrid data-driven approach integrating TFT and SAC algorithm, i.e., TFT-SAC approach, to tackle the non-convex operational optimization problem in BIES. The TFT is used to forecast the renewable generation and energy demand based on historical data, and the obtained forecasts are then utilized by the SAC algorithm to solve the scheduling problems. Unlike conventional black-box forecasting methods, the TFT provides interpretability through the attention mechanism, enhancing the trustworthiness of forecasting results for decision-making. Furthermore, the SAC algorithm, trained to maximize the policy entropy, can learn an operational strategy with superior robustness and generalization capabilities.

3) The proposed TFT-SAC approach is trained and tested on a real-world dataset to validate its superior performance

in reducing the energy cost and computational time compared with the benchmark approaches. The generalization performance for the learned scheduling policy and the sensi-

tivity analysis are examined in various scenarios.

A comprehensive comparison between the proposed TFT-SAC approach and other approaches is presented in Table I.

TABLE I
COMPARISON BETWEEN PROPOSED TFT-SAC APPROACH AND OTHER APPROACHES

Reference	Non-convex model	Forecast		Optimization			Solution algorithm
		Model	Explainability	Robustness	Generalization	Computational efficiency	
[23]		GRU-BLSTM		✓			RSO
[24]		LSTM					RO
[25]		ANN					Deterministic
[26]					✓	✓	TD3
[12]	✓	CNN-BLSTM			✓	✓	DDPG
[15]		LSTM		✓	✓	✓	SAC
This paper	✓	TFT	✓	✓	✓	✓	SAC

Note: ANN, GRU, and RSO are short for artificial neural network, gated recurrent unit, and robust stochastic optimization, respectively.

The remainder of this paper is organized as follows. Section II covers the system description, device modeling, optimization problem, and MDP. Section III introduces the proposed hybrid data-driven approach integrating TFT and SAC algorithm. Section IV validates the proposed TFT-SAC approach with simulations, and Section V concludes this paper.

II. PROBLEM FORMULATION

A. System Description

This study focuses on a modern BIES that encompasses grid-connected electric systems and independent heating systems, as illustrated in Fig. 1. In practice, such systems can be found in university campuses, residential complexes, and industrial parks.

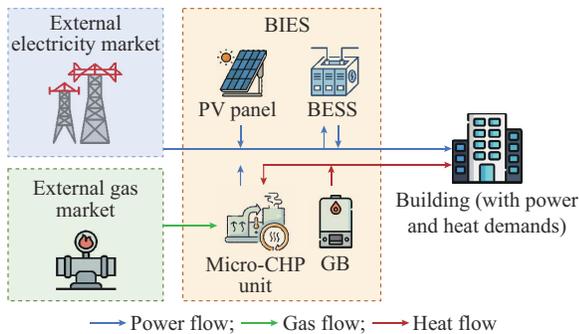


Fig. 1. Illustration of BIES.

As shown in Fig. 1, the BIES operates to meet multiple energy demands using both internal energy devices and external energy resources. Specifically, the electric system, which comprises PV panels, micro-CHP unit, and BESSs, is grid-connected to satisfy the power demands of building. Typically, BIESs purchase electricity from the external electricity market when the demand exceeds renewable generation and may sell electricity when renewable generation is surplus. The BESS enhances the operational flexibility and adds complexity to the decision-making process. PV and

BESS, as components of DC systems, are connected to the building and power grid through electronic interfaces. The maximum power point tracking (MPPT) is used to control the inverter between the DC and AC systems, maximizing energy extraction from PV panels despite fluctuating solar conditions. For the purposes of this paper, the dynamics inside the power converters are neglected, as the focus is on optimizing the hourly operational strategy.

Additionally, independent heating systems, consisting of micro-CHP units and GBs, are commonly deployed in building complexes, campuses, and industrial parks, particularly in regions with high heat demands. These localized heating systems reduce the significant transmission losses associated with centralized heating. The BIES model also assumes a connection to an external natural gas market as the fuel source for the micro-CHP units. Detailed models of these devices are provided as follows.

B. Device Modeling

1) Micro-CHP Unit Modeling

The micro-CHP unit is a crucial component of BIESs, functioning as a single-input multi-output energy converter. It is highly efficient in converting natural gas to power and heat, as a key element in enhancing the energy efficiency of BIES. Typically, the micro-CHP unit is modeled with constant energy conversion efficiencies for both power and heat. However, the generation of power and heat by micro-CHP unit is interdependent, resulting in an FOR. In this paper, we employ a non-convex operational model for micro-CHP unit. The non-convex FOR of this model is depicted in Fig. 2, bounded by the curve $ABCDEF$. This FOR is considered to comprise two convex subregions, labeled as I and II.

The mathematical representation of the FOR for the micro-CHP unit is given by (1), as detailed in [12].

$$P_{\text{CHP},e}^t - P_{\text{CHP},e}^B - \frac{P_{\text{CHP},e}^B - P_{\text{CHP},e}^C}{P_{\text{CHP},h}^B - P_{\text{CHP},h}^C} (P_{\text{CHP},h}^t - P_{\text{CHP},h}^B) \leq 0 \quad \forall t \in T \quad (1a)$$

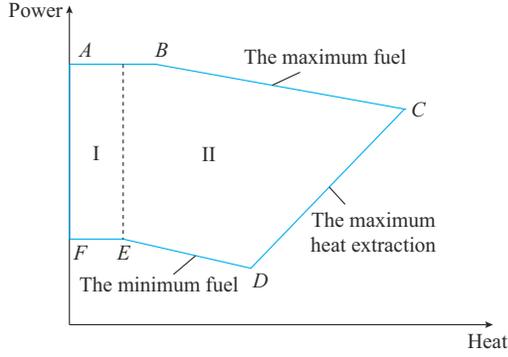


Fig. 2. FOR of micro-CHP unit.

$$P_{\text{CHP},e}^t - P_{\text{CHP},e}^C - \frac{P_{\text{CHP},e}^C - P_{\text{CHP},e}^D}{P_{\text{CHP},h}^C - P_{\text{CHP},h}^D} (P_{\text{CHP},h}^t - P_{\text{CHP},h}^C) \leq 0 \quad \forall t \in T \quad (1b)$$

$$-\left(1 - \bar{X}_{\text{CHP}}^t\right) \Gamma \leq P_{\text{CHP},e}^t - P_{\text{CHP},e}^E - \frac{P_{\text{CHP},e}^E - P_{\text{CHP},e}^F}{P_{\text{CHP},h}^E - P_{\text{CHP},h}^F} \left(P_{\text{CHP},h}^t - P_{\text{CHP},h}^E\right) \quad \forall t \in T \quad (1c)$$

$$-\left(1 - \underline{X}_{\text{CHP}}^t\right) \Gamma \leq P_{\text{CHP},e}^t - P_{\text{CHP},e}^D - \frac{P_{\text{CHP},e}^D - P_{\text{CHP},e}^E}{P_{\text{CHP},h}^D - P_{\text{CHP},h}^E} \left(P_{\text{CHP},h}^t - P_{\text{CHP},h}^D\right) \quad \forall t \in T \quad (1d)$$

$$\bar{X}_{\text{CHP}}^t + \underline{X}_{\text{CHP}}^t = I_{\text{CHP}}^t \quad \forall t \in T \quad (1e)$$

$$-\left(1 - \underline{X}_{\text{CHP}}^t\right) \Gamma \leq P_{\text{CHP},h}^t - P_{\text{CHP},h}^E \leq \left(1 - \bar{X}_{\text{CHP}}^t\right) \Gamma \quad \forall t \in T \quad (1f)$$

$$0 \leq P_{\text{CHP},e}^t \leq P_{\text{CHP},e}^A I_{\text{CHP}}^t \quad \forall t \in T \quad (1g)$$

$$0 \leq P_{\text{CHP},h}^t \leq P_{\text{CHP},h}^A I_{\text{CHP}}^t \quad \forall t \in T \quad (1h)$$

where $P_{\text{CHP},e}^t$ and $P_{\text{CHP},h}^t$ are the output power and heat of micro-CHP unit at time t , respectively; $P_{\text{CHP},e}^A$ and $P_{\text{CHP},h}^A$ are the generated power and heat of micro-CHP unit at point A, and those at other points B, C, D, E, and F are similarly defined; Γ is a sufficiently large number used to assist in the model description; I_{CHP}^t is the commitment status of the micro-CHP unit; $T = \{1, 2, \dots, 24\}$ is the set of operational hours; and \bar{X}_{CHP}^t and $\underline{X}_{\text{CHP}}^t$ are the operating statuses in the convex subregions I and II, respectively. If the micro-CHP unit operates in the convex subregion I, $\bar{X}_{\text{CHP}}^t = 1$ and $\underline{X}_{\text{CHP}}^t = 0$; otherwise, $\underline{X}_{\text{CHP}}^t = 1$ and $\bar{X}_{\text{CHP}}^t = 0$.

The total operation cost of micro-CHP unit C_{CHP}^t at time t is expressed as:

$$C_{\text{CHP}}^t (P_{\text{CHP},e}^t, P_{\text{CHP},h}^t) = \bar{\alpha}_{\text{CHP}} (P_{\text{CHP},e}^t)^2 + \bar{\beta}_{\text{CHP}} P_{\text{CHP},e}^t + \bar{\gamma}_{\text{CHP}} + \underline{\alpha}_{\text{CHP}} (P_{\text{CHP},h}^t)^2 + \underline{\beta}_{\text{CHP}} P_{\text{CHP},h}^t + \underline{\gamma}_{\text{CHP}} P_{\text{CHP},e}^t P_{\text{CHP},h}^t \quad (2)$$

where $\bar{\alpha}_{\text{CHP}}$, $\underline{\alpha}_{\text{CHP}}$, $\bar{\beta}_{\text{CHP}}$, $\underline{\beta}_{\text{CHP}}$, $\bar{\gamma}_{\text{CHP}}$, and $\underline{\gamma}_{\text{CHP}}$ are the cost coefficients.

2) BESS Modeling

The BESS is conceptualized as a battery capable of charging and discharging with distinct efficiencies. The operational strategy of BESS is designed with a granularity of one hour, corresponding to one time slot. This means that all

charging and discharging activities of BESS within a time period are aggregated into a single operation. Consequently, the BESS can either charge or discharge in any given time slot, but not both simultaneously [27].

$$E_{\text{BESS}}^t = (1 - \beta) E_{\text{BESS}}^{t-1} + P_{\text{BESS},c}^t \eta_{\text{BESS},c} - P_{\text{BESS},d}^t \quad (3a)$$

$$0 \leq P_{\text{BESS},c}^t \leq S_{\text{BESS},c}^t P_{\text{BESS},c,\text{max}} \quad (3b)$$

$$0 \leq P_{\text{BESS},d}^t \leq S_{\text{BESS},d}^t P_{\text{BESS},d,\text{max}} \quad (3c)$$

$$S_{\text{BESS},c}^t + S_{\text{BESS},d}^t \leq 1 \quad (3d)$$

$$E_{\text{BESS},\text{min}} \leq E_{\text{BESS}}^t \leq E_{\text{BESS},\text{max}} \quad (3e)$$

where E_{BESS}^t is the state of charge (SoC) of BESS at time t ; β and $\eta_{\text{BESS},c}$ are the predetermined loss factor and charging efficiency, respectively; $P_{\text{BESS},c}^t$ and $P_{\text{BESS},d}^t$ are the charging power and discharging power of BESS at time t , respectively; $S_{\text{BESS},c}^t$ and $S_{\text{BESS},d}^t$ are the charging state and discharging state of BESS at time t , respectively; and the subscripts max and min represent the maximum and minimum values of corresponding variables, respectively.

The SoC of BESS is calculated in (3a). The charging power and discharging power of BESS are constrained by (3b) and (3c), respectively. Constraint (3d) is employed to determine the charging or discharging state of BESS. The total capacity of BESS is constrained by (3e).

3) GB Modeling

The GB is modelled as an energy device transforming natural gas to heat with a fixed rate. The model of GB can be described as:

$$P_{\text{GB},h}^t = \eta_{\text{GB}} P_{\text{GB},g}^t \quad (4a)$$

$$P_{\text{GB},g,\text{min}} \leq P_{\text{GB},g}^t \leq P_{\text{GB},g,\text{max}} \quad (4b)$$

$$P_{\text{GB},h,\text{min}} \leq P_{\text{GB},h}^t \leq P_{\text{GB},h,\text{max}} \quad (4c)$$

where η_{GB} is the natural gas conversion efficiency; $P_{\text{GB},g}^t$ is the consumed natural gas of GB at time t ; and $P_{\text{GB},h}^t$ is the generated heat of GB at time t .

C. Optimization Problem

Considering all the models of devices in BIES presented above, the primary objective of BIES is to minimize the total cost of system operation. Specifically, the operational cost encompasses several components, including the cost of purchasing electricity and gas from the external markets (EMs), the degradation cost of BESSs, and the penalty incurred for unfulfilled energy demand. Consequently, the optimization problem for BIES operator can be formulated as:

$$\min_j C_b = \sum_{t \in T} \left[x_{w,e}^t (P_{\text{CHP},e}^t + P_{\text{BESS},c}^t - P_{\text{BESS},d}^t - P_{\text{PV},e}^t) + x_{w,g}^t P_{\text{GB},g}^t + C_{\text{CHP}}^t \right] \quad (5a)$$

s.t.

$$P_e^t = P_{w,e}^t + P_{\text{PV},e}^t + P_{\text{CHP},e}^t + P_{\text{BESS},d}^t - P_{\text{BESS},c}^t \quad (5b)$$

$$P_h^t = P_{\text{CHP},h}^t + P_{\text{GB},h}^t \quad (5c)$$

$$P_{w,g}^t = P_{\text{CHP},g}^t \quad (5d)$$

where $P_{w,e}^t$ and $P_{w,g}^t$ are the power purchased from wholesale electricity and natural gas markets, respectively; $x_{w,e}^t$ and $x_{w,g}^t$

$$\mathcal{L}(\Omega, W) = \sum_{y_t \in \Omega} \sum_{q \in \Omega} \sum_{\tau=1}^{\tau_{\max}} \frac{QL(y_t, \hat{y}(q, t - \tau_{\max}, \tau_{\max}), q)}{M_{\tau_{\max}}} \quad (9)$$

where $\mathcal{L}(\Omega, W)$ is the quantile loss of single time series at the average prediction point, Ω is the domain of training data containing $M_{\tau_{\max}}$ samples, and W is the weight of TFT model; y_t is the actual data; \hat{y} is the prediction data; τ_{\max} is the maximum step; and the function $QL(\cdot)$ can be expressed as:

$$QL(y_t, \hat{y}, q) = q(y_t - \hat{y})_+ + (1 - q)(\hat{y} - y_t)_+ \quad (10)$$

where QL includes predicted values corresponding to different quantiles (e.g., 0.1, 0.5, and 0.9); and $(\cdot)_+ = \max(0, \cdot)$. To ensure consistency in prediction dimensions across different prediction points, the regularization is applied as:

$$q_{risk} = \frac{2 \sum_{y_t \in \tilde{\Omega}} \sum_{\tau=1}^{\tau_{\max}} QL(y_t, \hat{y}(q, t - \tau, \tau), q)}{\sum_{y_t \in \tilde{\Omega}} \sum_{\tau=1}^{\tau_{\max}} |y_t|} \quad (11)$$

where $\tilde{\Omega}$ is the domain of test samples; and q_{risk} is the normalized quantile losses across the entire forecasting horizon.

2) Gating Mechanism

In the time-series forecast, especially with multiple regression, identifying relevant variables and the extent of non-linear processing is challenging. The TFT model uses gated residual networks (GRNs) for adaptive non-linear processing:

$$GRN_{\omega} = LayerNorm(\mathbf{a} + GLU_{\omega}(\boldsymbol{\eta}_1)) \quad (12)$$

$$\boldsymbol{\eta}_1 = \mathbf{W}_{1,\omega} \boldsymbol{\eta}_2 + \mathbf{b}_{1,\omega} \quad (13)$$

$$\boldsymbol{\eta}_2 = ELU(\mathbf{W}_{2,\omega} \mathbf{a} + \mathbf{W}_{3,\omega} \mathbf{c} + \mathbf{b}_{2,\omega}) \quad (14)$$

$$GLU_{\omega}(\boldsymbol{\eta}_1) = \sigma(\mathbf{W}_{4,\omega} \boldsymbol{\eta}_1 + \mathbf{b}_{4,\omega}) \odot (\mathbf{W}_{5,\omega} \boldsymbol{\eta}_1 + \mathbf{b}_{5,\omega}) \quad (15)$$

where $LayerNorm(\cdot)$ is the layer normalization function; $\mathbf{a} + GLU_{\omega}(\boldsymbol{\eta}_1)$ represents the linear and nonlinear contributions, with GLU_{ω} controlling the degree of nonlinearity, and \mathbf{a} is the vector of primary inputs to GRN; \mathbf{c} is an optional context vector; $ELU(\cdot)$ is the activation function of exponential linear unit; $\sigma(\cdot)$ is the sigmoid activation function; $\mathbf{W}_{1,\omega}$, $\mathbf{W}_{2,\omega}$, $\mathbf{W}_{3,\omega}$, $\mathbf{W}_{4,\omega}$, and $\mathbf{W}_{5,\omega}$ are the weight sharing indices; and $\mathbf{b}_{1,\omega}$, $\mathbf{b}_{2,\omega}$, $\mathbf{b}_{4,\omega}$, and $\mathbf{b}_{5,\omega}$ are the bias sharing indices. The GRN layer is controlled by the GLU layer, which may skip the layer entirely if GLU outputs are close to 0.

3) VSN

The VSN is a key component of the TFT that improves the performance by selecting important features and filtering out noises. It assigns weights to features, which are used to combine the processed inputs:

$$v_{x_i} = Softmax\left(GRN_{v_x}(\boldsymbol{\Xi}_t, \mathbf{c}_s)\right) \quad (16)$$

where v_{x_i} is the weight corresponding to features; $\boldsymbol{\Xi}_t$ is the flattened vector; and \mathbf{c}_s is obtained from the static covariate encoder. The processed features are weighted by their corresponding variable selection weights and then combined.

4) Temporal Self-attention Layer

The TFT model employs a temporal self-attention layer that plays a key role in capturing long-term dependencies in time-series data. This layer not only improves the model ability to understand complex temporal relationships but also enhances the interpretability of forecasts. The self-attention layer used here is a masked and interpretable multi-head attention layer combined with a gating mechanism to selectively control information flow.

1) Self-attention mechanism

The core concept behind the temporal self-attention layer is to calculate the relevance, or ‘‘attention’’, of different time steps to each other, enabling the TFT model to focus on important events or sequences within the data. This is done using the following equation for attention:

$$Attention(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = A(\mathbf{Q}, \mathbf{K}) \mathbf{V} \quad (17)$$

where \mathbf{V} is the value of input based on the similarity between the query vector \mathbf{Q} and key vector \mathbf{K} ; and $A(\cdot)$ is a normalization function that determines the attention weights of value \mathbf{V} . The scaled dot-product mechanism for calculating attention is defined as:

$$A(\mathbf{Q}, \mathbf{K}) = Softmax\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_{attn}}}\right) \quad (18)$$

where d_{attn} is the dimension of attention layer.

2) Multi-head self-attention mechanism

Multi-head self-attention mechanism enhances the power of the self-attention mechanism by allowing the model to jointly focus on information from different representation subspaces at different positions. Instead of using a single set of queries, keys, and values, the multi-head self-attention mechanism splits them into multiple sets, each of which is processed independently. Each head computes attention separately, and the results are then concatenated and linearly transformed to produce the final output. By having multiple heads, the TFT model can capture a richer set of relationships and nuances in the data compared with a self-attention mechanism, which are presented as:

$$MultiHead(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = [\mathbf{H}_1 \ \mathbf{H}_2 \ \dots \ \mathbf{H}_{m_H}] \mathbf{W}_H \quad (19)$$

$$\mathbf{H}_h = Attention\left(\mathbf{Q}\mathbf{W}_Q^{(h)}, \mathbf{K}\mathbf{W}_K^{(h)}, \mathbf{V}\mathbf{W}_V^{(h)}\right) \quad (20)$$

where $\mathbf{W}_Q^{(h)} \in \mathbb{R}^{d_{model} \times d_{attn}}$, $\mathbf{W}_K^{(h)} \in \mathbb{R}^{d_{model} \times d_{attn}}$, and $\mathbf{W}_V^{(h)} \in \mathbb{R}^{d_{model} \times d_v}$ are the head-specific weights for queries, keys, and values, respectively, and d_{model} and d_v are the dimensions of model and weight, respectively; and $\mathbf{W}_H \in \mathbb{R}^{(m_H d_v) \times d_{model}}$ linearly combines outputs concatenated from all heads \mathbf{H}_h ($h = 1, 2, \dots, m_H$), and m_H is the number of heads.

3) Interpretability enhancement

One of the main issues with traditional multi-head attention mechanism is that each head uses different value vectors, making it difficult to directly determine the feature importance from the attention weights. By modifying the mechanism to share the same value vector across all heads, the TFT model can produce a unified set of attention weights, thereby improving interpretability:

$$IMH(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \tilde{\mathbf{H}} \tilde{\mathbf{W}}_H \quad (21)$$

$$\tilde{H} = \tilde{A}(\mathbf{Q}, \mathbf{K}) \mathbf{V} \mathbf{W}_V = \left(\frac{1}{m_H} \sum_{h=1}^{m_H} A(\mathbf{Q} \mathbf{W}_Q^{(h)}, \mathbf{K} \mathbf{W}_K^{(h)}) \right) \mathbf{V} \mathbf{W}_V = \frac{1}{m_H} \sum_{h=1}^{m_H} \text{Attention}(\mathbf{Q} \mathbf{W}_Q^{(h)}, \mathbf{K} \mathbf{W}_K^{(h)}, \mathbf{V} \mathbf{W}_V) \quad (22)$$

where $\mathbf{IMH}(\cdot)$ is the interpretable multi-head; $\tilde{\mathbf{W}}_H \in \mathbb{R}^{d_{\text{model}} \times d_{\text{att}}}$ denotes the final linear mapping used across \mathbf{W}_H ; and $\mathbf{W}_V \in \mathbb{R}^{d_{\text{model}} \times d_V}$ is the value weight shared across all heads. Compared with $A(\mathbf{Q}, \mathbf{K})$ in (18), this modification allows each attention head to share the same set of values $\tilde{A}(\mathbf{Q}, \mathbf{K})$, resulting in a single and interpretable set of attention scores that can be analyzed to determine feature importance [29].

B. SAC Algorithm

In this subsection, we describe the SAC algorithm, which is a state-of-the-art maximum-entropy-based off-policy DRL algorithm, to solve the optimization problem of BIES. Typical DRL algorithms generally suffer from limited robustness in real-world applications due to ineffective exploration [30]. In contrast, the SAC algorithm uses entropy as a regularization term in the objective function to enhance the adaptability and generalization performance.

1) Algorithm Description

As a DRL algorithm with an actor-critic structure, the SAC algorithm outperforms most algorithms, e.g., DDPG, in convergence performance. SAC algorithm maximizes both accumulative rewards and policy entropy. The entropy function $H(\cdot)$ is defined in (23), where $\pi(\cdot | s_t)$ is the strategy condition to the state s_t . The state value function $V_r^\mu(s)$ and state-action value function $Q_r^\mu(s, a)$ are presented in (24) and (25), respectively, where the temperature parameter α determines the relative importance of the entropy term against the reward, and thus controls the stochasticity of the optimal policy [30].

$$H(\pi(\cdot | s_t)) = - \sum_a \pi(a | s_t) \ln \pi(a | s_t) \quad (23)$$

$$V_r^\mu(s) = \mathbb{E} \left(\sum_{t \in T} \gamma^t (R_t + \alpha H(\pi(\cdot | s_t))) \middle| s_0 = s \right) \quad (24)$$

$$Q_r^\mu(s, a) = \mathbb{E} \left(\sum_{t \in T} \gamma^t \left(R_t + \alpha \sum_{t \in T} H(\pi(\cdot | s_t)) \right) \middle| s_0 = s, a_0 = a \right) \quad (25)$$

At the same time, the state value function can be presented as (26) according to (23) and (24). Equation (24) allows us to derive the solution of the policy as (27).

$$V_r^\mu(s_t) = \mathbb{E}(Q_r^\mu(s, a)) + \alpha H(\pi(\cdot | s_t)) \quad (26)$$

$$\pi^*(\cdot | s_t) = \arg \max_{\pi \in \Delta} V_r^\mu(s) = \frac{e^{Q_h^\pi(s, \cdot)/\alpha}}{\sum_a e^{Q_h^\pi(s, a)/\alpha}} \quad (27)$$

where $\Delta = \{\pi | \pi \geq 0, \mathbf{1} \cdot \pi = 1\}$, guaranteeing that π is a valid probabilistic distribution on the action space; and $Q_h^\pi(s, \cdot)$ is the Q function for taking all each action in the state s . When the Q value converges to the optima, the optimal policy achieves the optimal state value function. Therefore, the updating of Q -value function can be realized by using the

closed-form solution in an off-policy scheme.

2) Algorithm Implementation

The SAC algorithm adopts an actor-critic structure with DNNs to estimate the policy (actor) and Q -value functions (critic). The actor network is represented by the policy function $\mu(s | \theta^\mu)$ parameterized by θ^μ . The critic employs clipped double Q networks Q_1 and Q_2 and their target networks Q'_1 and Q'_2 . Therefore, the target y_t for the Q value is expressed as (28). Then, the L2 loss is used to update the Q -network in (29) for $j = \{1, 2\}$.

$$y_t = r_t + \gamma \left(\min_{j \in \{1, 2\}} Q_j(s_{t+1}, \tilde{a}_{t+1} | \theta^{Q_j}) - \alpha \lg \pi_\theta(\tilde{a}_{t+1} | s_{t+1}) \right) \quad (28)$$

$$\nabla_{\theta^Q} L = \frac{1}{N} \sum_{n \in \mathcal{N}} \left(y_t - Q_j(s, a | \theta^{Q_j}) \right)^2 \quad (29)$$

where \tilde{a}_{t+1} is the action under the current policy in the next state s_{t+1} ; $\mathcal{N} = \{1, 2, \dots, N\}$ is the set of mini batches indexed by n ; and π_θ is the executed policy.

To train these networks, the agent randomly samples tuples (s_j, a_j, r_j, s_{j+1}) from the ERB to form a mini batch n for experience replay learning. The online critic networks are updated by one step of gradient descent to the mean square error (MSE) θ^Q in (29), while the actor network is updated by one step of gradient ascent using (30). To stabilize the training, the target network parameters are soft updated with (31).

$$\nabla_{\theta^Q} L = \nabla_{\theta^Q} \frac{1}{N} \left(\min_{j \in \{1, 2\}} Q_j(s_t, \tilde{a}_t(s)) - \alpha \lg \pi_\theta(a_t | s_t) \right) \quad (30)$$

$$\theta^{Q'} \leftarrow \rho \theta^{Q'} + (1 - \rho) \theta^{Q'} \quad (31)$$

where $\tilde{a}_t(s)$ is a sample from $\pi_\theta(\cdot | s_t)$; and ρ is the soft update parameter.

C. Discussions

The use of the proposed TFT-SAC approach is unique and effective for the dynamic operation and control of BIES. This combination offers several advantages but also has potential shortcomings compared with other traditional approaches.

1) Integrated forecasting and operation: the TFT provides accurate and data-driven forecasts of PV generation and energy demand, which allows the SAC algorithm to make informed decisions. This integration reduces uncertainty in the decision-making process, leading to more reliable system operations.

2) Offline training and efficient online operation: the proposed TFT-SAC approach allows for offline training using historical data, enabling the development of a robust policy before deployment. Once trained, the algorithm operates in real time with minimal computational overhead, which is a significant advantage over approaches like SO or RO that require repeated recalculation.

3) Handling non-convexity: the operation of BIES involves non-convex constraints such as the FOR. The SAC algorithm, leveraging DNNs, can effectively learn non-convex optimal operating policies due to the powerful representation capabilities of DNNs. In comparison, traditional mathemati-

cal programming approaches such as mixed-integer linear programming (MILP) address non-convexity by linearizing nonlinear relationships and explicitly formulating integer constraints, facing scalability and computational challenges particularly in large and dynamic systems like BIES. Heuristic algorithms can explore complex optimization landscapes and are often more flexible than mathematical programming. However, they may suffer from high computational demands, especially in large-scale systems, and may converge to local optima rather than finding the global solution.

4) Training complexity: the proposed TFT-SAC approach requires extensive offline training, which can be computationally expensive and time-consuming, particularly for large datasets. The performance highly relies to a high-quality training dataset, which is typically hard to acquire in the real world.

5) Dependence on forecasting accuracy: the effectiveness of SAC algorithm in making optimal decisions depends heavily on the forecasting accuracy provided by TFT. If the forecasts are inaccurate due to unexpected external factors, the quality of the operational decisions may be compromised.

Overall, the proposed TFT-SAC approach provides an effective solution for BIES operation. The integrated forecast and optimized structure, capability to handle non-convexity, and efficient implementation make this approach a compelling alternative to traditional approaches, despite some challenges related to training complexity and dependence on forecasting accuracy.

IV. CASE STUDY

A. Simulation Setup

To validate the effectiveness of the proposed TFT-SAC approach, we conduct case studies using data from a real building located in Zhenjiang, China. The BIES under study comprises devices like a micro-CHP unit, PV panels, BESSs, and GBs to meet both heat and power demands.

The micro-CHP unit, with a rated output of 25.3 kW, is designed to satisfy the heat demand of the building while partially covering its power demand. The PV system includes 610 PV panels, each with a capacity of 280 W, resulting in a theoretical maximum output of 170.8 kW. However, due to practical limitations, the actual capacity is 153 kW. The BESS consists of 24 LiFePO₄ batteries, each with a storage capacity of 5.12 kWh, providing a maximum output of 72 kW. This setup enables the BESS to support peak power demand for up to 4 hours. Detailed information of micro-CHP unit and BESS is shown in Supplementary Material A.

The proposed TFT-SAC approach is implemented in Python, and the neural networks are developed using PyTorch [31]. To achieve the optimal performance, the neural network parameters and hyperparameters are carefully chosen based on empirical values and adjusted throughout the training process. The complete configuration details for SAC algorithm are presented in Tables II and III. While hyperparameter setting of TFT for forecasts of energy demand and PV generation is shown in Table IV.

TABLE II
NEURAL NETWORK ARCHITECTURE SETTING OF SAC ALGORITHM

Neural network	Number of hidden layers	Number of neurons	Learning rate	Soft update parameter	Optimizer
Actor	3	[512, 32]	1×10^{-4}	11×10^{-2}	Adam
Critic	2	[512, 32]	1×10^{-3}	11×10^{-2}	Adam

TABLE III
HYPERPARAMETER SETTING OF SAC ALGORITHM

Training parameter	Number
Replay buffer size	1×10^6
Replay start size	128
Batch size	128
Discount factor	0.99

TABLE IV
HYPERPARAMETER SETTING OF TFT FOR FORECASTS OF ENERGY DEMAND AND PV GENERATION

Parameter	Forecast of energy demand	Forecast of PV generation
Learning rate	1×10^{-4}	3.5×10^{-3}
Grad clip value	0.1	0.9
Patience	10	2
Batch size	16	16
Drop out	0.2	0.1
Time step	168	24
Hidden size	128	32
Number of LSTM layers	6	4
Number of attention heads	6	3
Loss function	Quantile loss	Quantile loss

B. Computational Performance of Different Algorithms

This subsection compares the SAC algorithm with baseline algorithms TD3 and DDPG. Each algorithm is trained for 10000 episodes on sampled days from the training set. Figure 4 shows the episodic reward evolution of different algorithms during offline training process. Considering the fluctuations in state features, the data have been smoothed using a 100-episode moving average method. This is because the oscillations caused by the exogenous state features cannot be addressed by the operational strategies even if the policy is optimal.

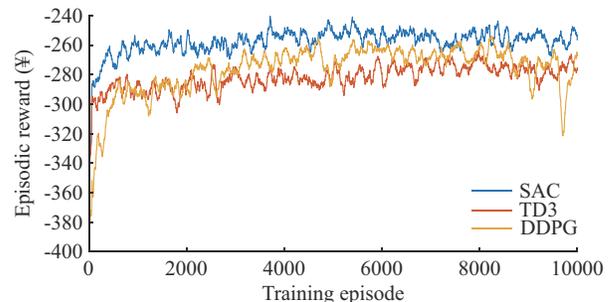


Fig. 4. Episodic reward evolution of different algorithms during offline training process.

Figure 4 shows that initially, the learning curves of different algorithms are similar due to randomly selected energy schedules and Gaussian noises. Early on, rewards are low for all algorithms. As training progresses, rewards increase as agents learn and refine their policies. The reward of SAC algorithm grows the fastest initially, followed by TD3 and then DDPG. Around 2000 iterations, the reward of DDPG increases sharply, surpassing TD3 but is still lower than SAC algorithm, which is close to converging. DDPG and TD3 converge around 5000 iterations. The SAC algorithm achieves a significantly higher final reward compared with DDPG and TD3, with the reward of DDPG slightly higher than that of TD3. This indicates the superior offline training performance of SAC algorithm.

To evaluate the performance of the proposed TFT-SAC approach, we use the trained actor network parameters to generate operational strategies for the BIES over 50 test days. We compare this forecasting-combined RL approach with benchmark approaches: typical RL approaches (TD3, DDPG, and SAC) and another forecasting-combined RL approach (LSTM-SAC). Figure 5 compares the cumulative costs for energy consumption with different approaches over 50 test days. The results indicate that the cumulative costs with typical RL approaches are significantly higher than those with forecasting-combined RL approaches. The cost gap increases with more training episodes, highlighting the differences between different approaches. For forecasting-combined approaches, the cumulative costs are similar, showing that combining forecasting with RL is effective. Notably, the proposed TFT-SAC approach achieves lower costs than LSTM-SAC, demonstrating its superior performance. However, the difference between the proposed TFT-SAC approach and LSTM-SAC is small compared with their differences between typical RL approaches, suggesting limited room for improvement in current forecasting-combined RL approaches.

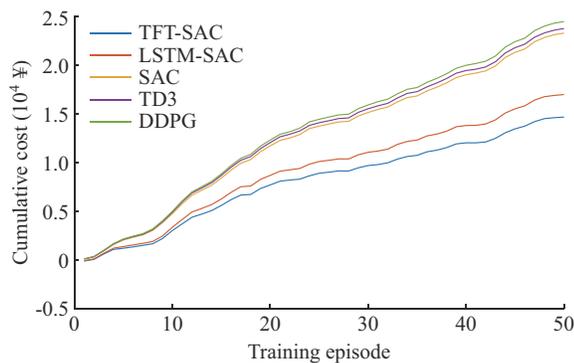


Fig. 5. Cumulative cost for energy consumption with different approaches over 50 test days.

C. Forecasting Performance Analysis

As can be known from Table V, the TFT model outperforms the LSTM model across three performance metrics, i.e., mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination R^2 in forecasts of both PV generation and energy demand.

TABLE V
PERFORMANCE METRICS OF TFT AND LSTM MODELS

Forecast object	Model	MAE	RMSE	R^2
PV generation	LSTM	3.66	12.23	0.8402
	TFT	5.22	11.24	0.8721
Energy demand	LSTM	3.37	4.60	0.9407
	TFT	2.20	3.26	0.9670

Figures 6 and 7 show that the forecasting curves of TFT model closely fit the target curves, demonstrating its effectiveness in capturing time-series patterns. The TFT model particularly excels in PV generation forecasting, accurately capturing peaks and valleys, which is crucial for energy forecast. In summary, the TFT model shows superior forecasting accuracy and pattern recognition compared with the LSTM model, which is crucial for energy management in BIES, guiding energy allocation, optimizing resource utilization, and improving overall energy efficiency.

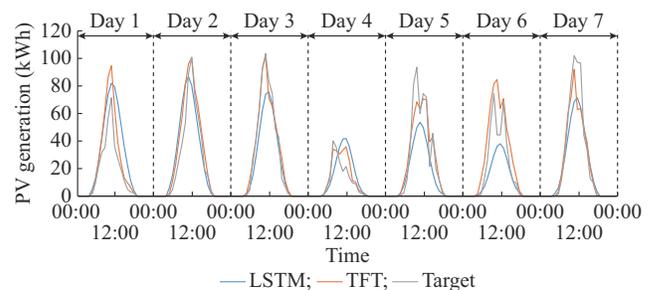


Fig. 6. Performance of LSTM and TFT models in PV generation forecasting.

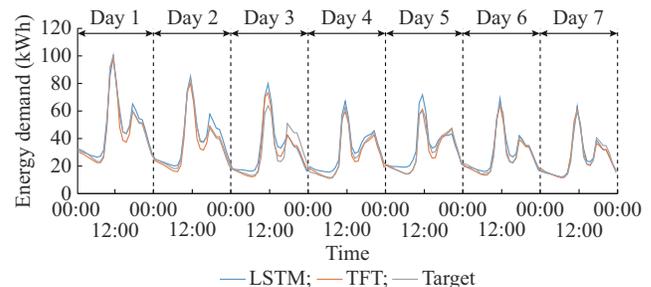


Fig. 7. Performance of LSTM and TFT models in energy demand forecasting.

The meteorological data include net solar irradiation (NSI), solar irradiation (SI), ultraviolet (UV), outdoor air temperature (OAT), rainfall (RF), relative humidity (RH), temperature-humidity-wind (THW), and surface air temperature (SAT). Figure 8 illustrates the relative importance of different features in the TFT model for PV generation forecasting. In the encoder, SI appears as the most significant factor, indicating that direct sunlight intensity plays a crucial role in PV generation forecasting. Meanwhile, in the decoder, longitude emerges as the most important feature, highlighting the importance of geographical positioning in the forecasting process. This is intuitive because the position affects the angle of sunlight and daylight duration, which ultimately impacts PV generation.

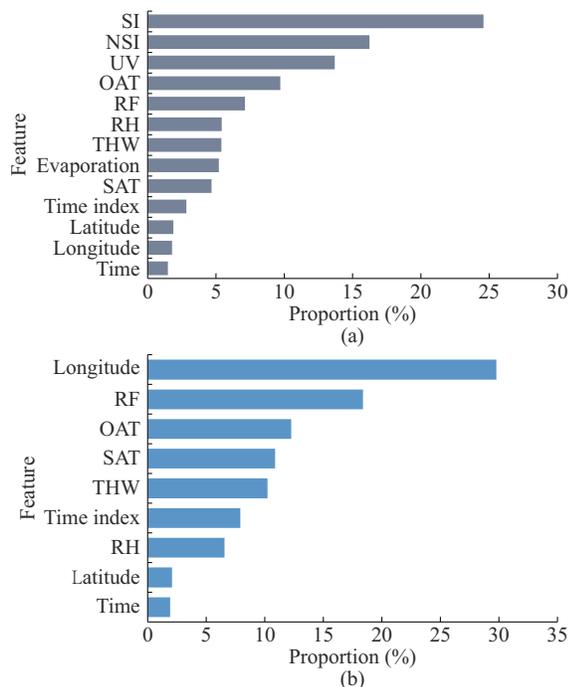


Fig. 8. Relative importance of different features in TFT model for forecasting PV generation. (a) Encoder. (b) Decoder.

Figure 9 depicts relative importance of different features in TFT model for forecasting energy demand. Unlike PV generation, which predominantly relies on weather-related factors, the energy demand is highly influenced by calendar-based information. Features such as hour of the day, workday status, and specific time-based attributes are ranked highly, reflecting the relationship between user behavior and energy usage. These calendar-related features indicate the impact of typical human activities and routines such as work schedules and holidays on energy demand.

The importance ranking reveals that the TFT model considers both weather conditions and temporal attributes to accurately forecast energy demands. This is crucial because user activities are often influenced by the time of day or specific events on the calendar, and these behavioral patterns significantly affect energy usage in buildings. The model attention to these aspects shows its ability to learn from diverse data sources and focuses on the most impactful features during the training process, resulting in a more reliable forecast.

Figures 10 and 11 illustrate the attention distribution of TFT model over the past 7 days (indexed by -7 to -1) during the forecasting process. Figure 10 shows that the attention of TFT model is concentrated on the recent past, especially the previous day, reflecting the strong daily cyclic patterns of PV generation. Minor peaks indicate the consideration of earlier time steps, which have lower weights due to the influence of short-term environmental factors like SI.

Figure 11 shows a smooth distribution across various historical time steps with a gradual increase. This suggests that the TFT model considers a range of past data, reflecting that the high complexity and irregularity of energy demands are influenced by factors like user behavior, daily activities, and weather conditions.

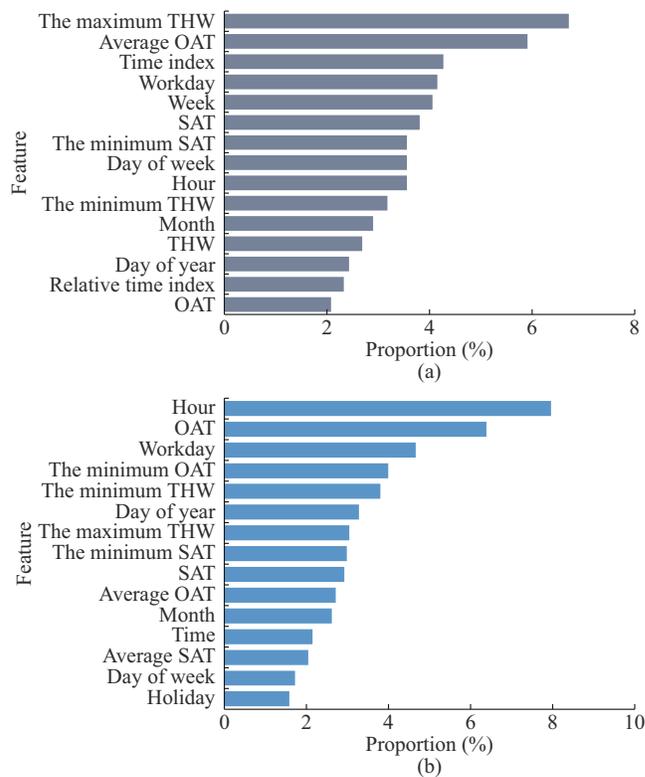


Fig. 9. Relative importance of different features in TFT model for forecasting energy demand. (a) Encoder. (b) Decoder.

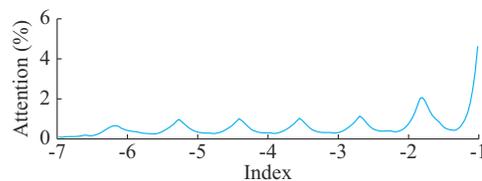


Fig. 10. Attention of TFT model over past 7 days for forecasting PV generation.

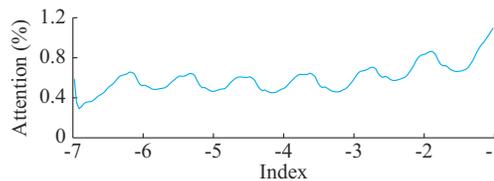


Fig. 11. Attention of TFT model over past 7 days for forecasting energy demand.

In comparison, the TFT model for PV generation forecasting focuses on recent time steps due to daily cyclic patterns, while that for forecasting energy demands has a broad attention span over the entire historical cycle, balancing long-term trends and short-term impacts. The gradual increase in attention weights indicates the emphasis on recent information for imminent forecasts.

The uniform attention distribution for energy demand suggests its cyclical patterns are less pronounced or more complex than that of PV generation. This highlights the importance of extracting information from multiple time scales for accurate forecasts and underscores the need for effective en-

ergy management strategies to optimize BIES operational efficiency.

In summary, the TFT model provides accurate and interpretable forecasts for both PV generation and energy demand, supporting the RL algorithm in formulating efficient scheduling strategies.

D. Generalization Performance

To validate the generalization performance, different approaches are tested over a test set that shows different statistical characteristics compared with the training set. The test set is represented by several typical weeks labeled W-1 to W-4 for comparative analysis. These typical weeks include scenarios with extreme PV generation or energy demand. Table VI presents the daily operational costs of BIES across different weeks. The results clearly demonstrate that forecasting-combined RL approaches achieve significantly lower operational costs compared with typical RL approaches, underscoring the effectiveness of combining the forecast and decision-making. Furthermore, the average operational cost of proposed TFT-SAC approach is lower than that of LSTM-SAC, indicating that the proposed TFT-SAC approach outperforms all the comparable approaches across different scenarios, thereby demonstrating its strong generalization capabilities. Although the daily cost improvements may appear marginal, the cumulative benefits of the proposed TFT-SAC approach over extended operation could result in substantial additional profits.

TABLE VI
COMPARISON OF DAILY AVERAGE OPERATIONAL COST OF BIES ACROSS DIFFERENT WEEKS

Week	Daily average operational cost (¥)				
	DDPG	TD3	SAC	LSTM-SAC	TFT-SAC
W-1	500.14	499.30	490.19	328.02	325.79
W-2	361.75	361.20	347.92	232.76	231.60
W-3	450.34	449.66	431.40	318.91	311.03
W-4	733.25	732.44	715.75	521.30	520.99

E. Robust Operation

To compare the robustness of the proposed TFT-SAC approach with other RL approaches, we introduce independent Gaussian noises to real PV generation and energy demand to represent uncertain scenarios. The average daily operational costs of BIES at different noise levels are presented in Table VII.

TABLE VII
COMPARISON OF DAILY AVERAGE OPERATIONAL COST OF BIES AT DIFFERENT NOISE LEVELS

Noise level N	Daily average operational cost (¥)				
	DDPG	TD3	SAC	LSTM-SAC	TFT-SAC
0.01	596.07	557.56	557.49	505.12	490.04
0.02	596.38	558.24	558.18	505.82	491.88
0.03	597.37	559.02	558.96	506.62	494.91
0.04	599.80	559.85	559.78	507.47	495.13
0.05	603.91	560.73	560.66	508.38	495.17

Across all noise levels, the typical RL approaches incur significantly higher operational costs than forecasting-combined RL approaches, with cost differences ranging from ¥60 to ¥100. Among all the tested approaches, the proposed TFT-SAC approach demonstrates the lowest average operational costs, indicating superior robustness. However, the cost variations between the proposed TFT-SAC approach and LSTM-SAC remain small in the range of ¥10 and ¥20. In contrast, the cost difference of the proposed TFT-SAC approach with $N=0.01$ and $N=0.05$ is approximately ¥5, and that of TD3, SAC, and LSTM-SAC is ¥3. This larger cost variation suggests that proposed TFT-SAC approach is more sensitive to forecasting accuracy than other approaches, even though it consistently achieves the lowest average operational costs among all approaches.

F. Operational Analysis

To evaluate the generalization of the optimal energy management policy learned by the proposed TFT-SAC approach, we apply two typical scenarios: a summer day (August 27) and a winter day (December 25). Figures 12 and 13 show the power and heat profiles on the two typical days, respectively.

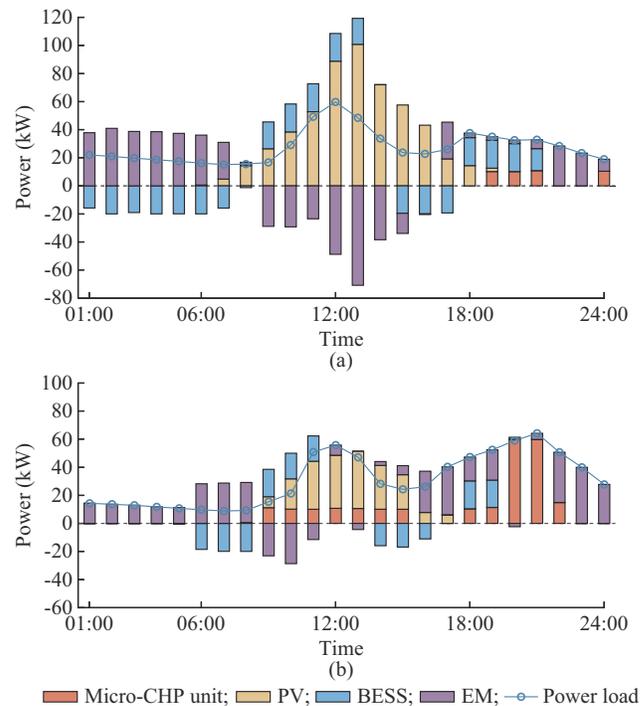


Fig. 12. Power generation and consumption of BIES. (a) A summer day. (b) A winter day.

Both scenarios share common trends. Initially, from 00:00 to 08:00, the BIES purchases electricity due to zero PV generation and low SoC of BESS. BESS charges at low prices for future demands. From 09:00 to 15:00, PV generation and BESS discharging could meet most power demands, with excess power sold with high electricity prices. From 18:00 to 24:00, the BIES does not sell electricity, and micro-CHP unit becomes the primary power source due to high demand.

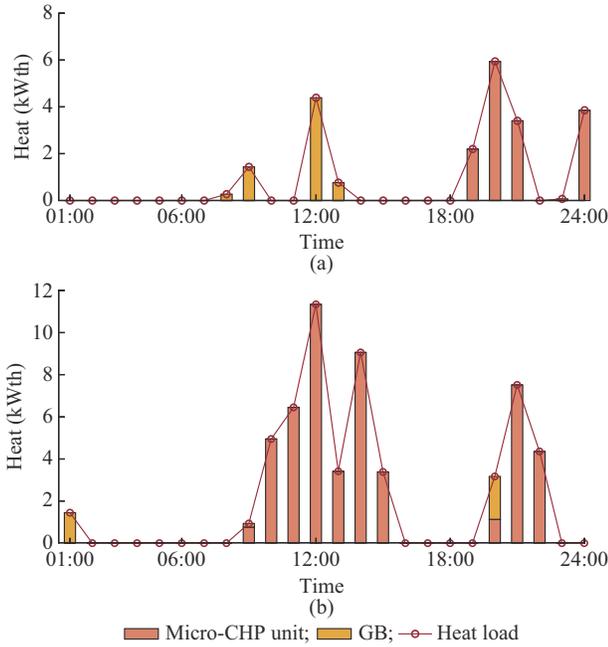


Fig. 13. Heat generation and consumption of BIES. (a) A summer day. (b) A typical day.

Nevertheless, there are some evident differences between the two typical days. On the winter day, the micro-CHP unit operates from 09:00 to 15:00 to meet high heat demands and support the power demands due to low PV generation. On the summer day, the micro-CHP unit is inactive as PV and BESS can meet the demands and the excess power is sold. The policy effectively uses micro-CHP unit in winter and BESS in summer, charging at low prices and discharging at high prices to maximize the economic benefits.

Finally, it can be concluded that the proposed TFT-SAC approach can learn an effective policy and can generalize to variable state information on different test days. Also, the flexibility of BIES is investigated on two typical winter and summer days. Specifically, the summer day has a higher PV generation and lower heat demand, so it has a higher energy export and makes use of more flexibility of BIES. Due to the lower PV generation and higher heat demand, the winter day has a higher power import and a higher utilization of the micro-CHP unit, which also provides lots of flexibility to BIES.

G. Sensitivity Analysis

In this subsection, a detailed sensitivity analysis is conducted to evaluate the impact of changes in key factors on the operation and performance of BIES. Specifically, we analyze the sensitivities of the episodic reward to variations in electricity price, PV generation, power demand, and heat demand, as shown in Fig. 14.

The sensitivity analysis is performed by varying each parameter independently from 90% to 110% of the initial configured value, with a granularity of 5%. This range is selected to represent potential fluctuations in market and operational conditions, and the granularity is chosen to provide a balanced level of detail without excessive computational overhead.

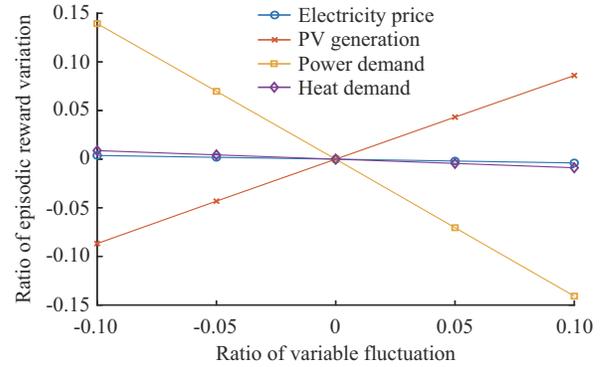


Fig. 14. Sensitivity analysis on several factor.

The results in Fig. 14 indicate that the episodic rewards of BIES are negatively correlated with electricity price, which is expected given that higher electricity prices increase the cost of electricity purchase. There is a positive correlation between PV generation and episodic reward, as the increased PV generation reduces the need for power from EM and allows for more excess power to be sold back to EM. Both power and heat demands negatively impact the rewards, with power demand having a particularly significant effect. This can be attributed to the fact that meeting higher demands requires more energy procurement, which incurs additional costs.

Interestingly, the power demand has a greater effect on the episodic reward compared with PV generation. This is because the total daily PV generation is lower than the total power demand. As a result, any decrease in power demand has a larger marginal impact on profitability, either through reduced procurement or allowing more energy to be sold during peak periods.

In terms of scheduling policies, the changes in power demand and PV generation lead to noticeable shifts in action prioritization. For instance, the increased PV generation results in more frequent utilization of BESS for energy arbitrage, while fluctuations in electricity price affect decisions regarding energy procurement timing. These findings emphasize the importance of accurate forecasts for PV generation and energy demand in effectively optimizing the operational strategies of BIES.

V. CONCLUSION

In conclusion, this paper develops a novel hybrid data-driven approach, i. e., TFT-SAC approach, for the optimal scheduling in BIES. Specifically, the TFT model enhances the forecasting accuracy and transparency through attention mechanisms and the VSN, enhancing interpretability and trustworthy of forecasting results. The integration of SAC algorithm for optimization further strengthens this framework by ensuring more effective exploration during training, leading to stronger robustness and generalization capabilities. Simulation results demonstrate the superior performance of the proposed TFT-SAC approach compared with the existing approaches. The interpretability of the TFT model and the generalization performance of SAC algorithm are analyzed. The sensitivity analysis of reward on several key factors in BIES is also made.

REFERENCES

- [1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy and Buildings*, vol. 128, pp. 198-213, Sept. 2016.
- [2] W. Wu, P. Li, B. Wang *et al.*, "Integrated distribution management system: architecture, functions, and application in China," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 2, pp. 245-258, Mar. 2022.
- [3] H. Qiu, V. Veerasamy, C. Ning *et al.*, "Two-stage robust optimization for assessment of PV hosting capacity based on decision-dependent uncertainty," *Journal of Modern Power Systems and Clean Energy*, vol. 12, no. 6, pp. 2091-2096, Nov. 2024.
- [4] X. Huang, Z. Xu, Y. Sun *et al.*, "Heat and power load dispatching considering energy storage of district heating system and electric boilers," *Journal of Modern Power Systems and Clean Energy*, vol. 6, no. 5, pp. 992-1003, Nov. 2018.
- [5] C. Huang, H. Zhang, L. Wang *et al.*, "Mixed deep reinforcement learning considering discrete-continuous hybrid action space for smart home energy management," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 3, pp. 743-754, May 2022.
- [6] H. Zhao, B. Wang, X. Wang *et al.*, "Active dynamic aggregation model for distributed integrated energy system as virtual power plant," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 5, pp. 831-840, Sept. 2020.
- [7] M. Sechilariu, B. Wang, and F. Locment, "Building integrated photovoltaic system with energy storage and smart grid communication," *IEEE Transactions on Industrial Electronics*, vol. 60, no. 4, pp. 1607-1618, Apr. 2013.
- [8] Y. Li, C. Wang, G. Li *et al.*, "Improving operational flexibility of integrated energy system with uncertain renewable generations considering thermal inertia of buildings," *Energy Conversion and Management*, vol. 207, p. 112526, Mar. 2020.
- [9] R. Jing, M. Wang, Z. Zhang *et al.*, "Comparative study of posteriori decision-making methods when designing building integrated energy systems with multi-objectives," *Energy and Buildings*, vol. 194, pp. 123-139, Jul. 2019.
- [10] Y. Zhang, P. E. Campana, A. Lundblad *et al.*, "Planning and operation of an integrated energy system in a Swedish building," *Energy Conversion and Management*, vol. 199, p. 111920, Nov. 2019.
- [11] Z. Zhu, Z. Hu, K. W. Chan *et al.*, "Reinforcement learning in deregulated energy market: a comprehensive review," *Applied Energy*, vol. 329, p. 120212, Jan. 2023.
- [12] A. Dolatabadi, H. Abdeltawab, and Y. A. I. Mohamed, "A novel model-free deep reinforcement learning framework for energy management of a PV integrated energy hub," *IEEE Transactions on Power Systems*, vol. 38, no. 5, pp. 4840-4852, Sept. 2023.
- [13] D. Qiu, Z. Dong, X. Zhang *et al.*, "Safe reinforcement learning for real-time automatic control in a smart energy-hub," *Applied Energy*, vol. 309, p. 118403, Mar. 2022.
- [14] Z. Zhu, K. W. Chan, S. Xia *et al.*, "Optimal bi-level bidding and dispatching strategy between active distribution network and virtual alliances using distributed robust multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2833-2843, Jul. 2022.
- [15] Y. Zhou, Z. Ma, J. Zhang *et al.*, "Data-driven stochastic energy management of multi energy system using deep reinforcement learning," *Energy*, vol. 261, p. 125187, Dec. 2022.
- [16] Z. Hu, K. W. Chan, Z. Zhu *et al.*, "Techno-economic modeling and safe operational optimization of multi-network constrained integrated community energy systems," *Advances in Applied Energy*, vol. 15, p. 100183, Sept. 2024.
- [17] Y. Zhou, B. Zhang, C. Xu *et al.*, "A data-driven method for fast AC optimal power flow solutions via deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1128-1139, Nov. 2020.
- [18] D. Cao, W. Hu, X. Xu *et al.*, "Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 5, pp. 1101-1110, Sept. 2021.
- [19] Q. Ma and C. Deng, "Simplified deep reinforcement learning based volt-var control of topologically variable power system," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 5, pp. 1396-1404, Sept. 2023.
- [20] Y. Wang, M. Mao, L. Chang *et al.*, "Intelligent voltage control method in active distribution networks based on averaged weighted double deep Q-network algorithm," *Journal of Modern Power Systems and Clean Energy*, vol. 11, no. 1, pp. 132-143, Jan. 2023.
- [21] B. Lim, S. Ö. Arık, N. Loeff *et al.*, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," *International Journal of Forecasting*, vol. 37, no. 4, pp. 1748-1764, Oct. 2021.
- [22] W. J. von Eschenbach, "Transparency and the black box problem: why we do not trust AI," *Philosophy & Technology*, vol. 34, no. 4, pp. 1607-1622, Sept. 2021.
- [23] T. M. Alabi, L. Lu, and Z. Yang, "Data-driven optimal scheduling of multi-energy system virtual power plant (MEVPP) incorporating carbon capture system (CCS), electric vehicle flexibility, and clean energy marketer (CEM) strategy," *Applied Energy*, vol. 314, p. 118997, May 2022.
- [24] S. Zhou, D. He, Z. Zhang *et al.*, "A data-driven scheduling approach for hydrogen penetrated energy system using LSTM network," *Sustainability*, vol. 11, no. 23, p. 6784, Dec. 2019.
- [25] A. Kämper, R. Delorme, L. Leenders *et al.*, "Boosting operational optimization of multi-energy systems by artificial neural nets," *Computers & Chemical Engineering*, vol. 173, p. 108208, May 2023.
- [26] Y. Xu, W. Gao, Y. Li *et al.*, "Operational optimization for the grid-connected residential photovoltaic-battery system using model-based reinforcement learning," *Journal of Building Engineering*, vol. 73, p. 106774, Aug. 2023.
- [27] G. Pan, W. Gu, Y. Lu *et al.*, "Optimal planning for electricity-hydrogen integrated energy system considering power to hydrogen and heat and seasonal storage," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2662-2676, Oct. 2020.
- [28] R. Wen, K. Torkkola, B. Narayanaswamy *et al.* (2017, Nov.). A multi-horizon quantile recurrent forecast. [Online]. Available: <https://arxiv.org/abs/1711.11053>
- [29] A. Vaswani, N. Shazeer, N. Parmar *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, vol. 30, pp. 1-10, Aug. 2017.
- [30] T. Haarnoja, A. Zhou, K. Hartikainen *et al.* (2018, Jan.). Soft actor-critic algorithms and applications. [Online]. Available: <https://arxiv.org/abs/1812.05905>.
- [31] A. Paszke, S. Gross, F. Massa *et al.*, "Pytorch: an imperative style, high-performance deep learning library," *Advances in Neural Information Processing Systems*, vol. 32, pp. 1-12, Dec. 2019.

Ze Hu received the B.Eng degree in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2021. He is now pursuing the Ph.D. degree at The Hong Kong Polytechnic University, Hong Kong, China. His research interests include applications of machine learning and game theory in energy system operation and electricity market.

Peijun Zheng received the M.S. degree from North China Electric Power University, Beijing, China, in 2021, and the Ph.D. degree from Waseda University, Tokyo, Japan, in 2024. He is currently working as a Postdoctoral in Tsinghua University, Beijing, China. His research interests include smart building, building energy saving, optimization, interpretable machine learning, and proton exchange membrane fuel cell.

Ka Wing Chan received the B.Sc. (Hons.) and Ph.D. degrees in electronic and electrical engineering from the University of Bath, Bath, U.K., in 1988 and 1992, respectively. He is currently an Associate Professor with the Department of Electrical Engineering, The Hong Kong Polytechnic University, Hong Kong, China. His research interests include power system stability, analysis and control, power grid integration, security, resilience and optimization, and demand response management.

Siqi Bu received the Ph.D. degree from the Electric Power and Energy Research Cluster, The Queen's University of Belfast, Belfast, U.K., where he continued his postdoctoral research work before entering industry. Then, he was with National Grid U.K., Warwick, U.K., serving as an experienced U. K. National Transmission System Planner and Operator. He is an Associate Professor and Associate Head with Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China, and also a Chartered Engineer with U.K. Royal Engineering Council, London, U.K.. He is a Fellow of IET. His research interests include power system stability, operation and economics considering renewable energy integration, smart grid application, and transport electrification.

Ziqing Zhu received the M.S. degree in electrical power system engineering from the University of Manchester, U.K., in 2019, and the Ph.D. degree

from The Hong Kong Polytechnic University, Hong Kong, China, in 2023, where he is currently serving as an Assistant Professor (Research). He is serving as the Early Career Researcher Associate Editor of IET Smart Grid. His research interests include application of artificial intelligence and machine learning in power system operation and electricity market.

Xiang Wei received the B.Eng. degree in automation from Hubei Normal University, Huangshi, China, in 2019, and the M.Sc. degree in control engineering with Shenzhen University, Shenzhen, China, in 2022. He is now pursuing the Ph.D. degree in The Hong Kong Polytechnic University, Hong Kong, China. His research interests include smart grid, integrated energy system planning, and renewable energy.

Yosuke Nakanishi received the B.S. and M.S. degrees in electrical engineering from Waseda University, Tokyo, Japan, in 1978 and 1980, respectively, and the Ph.D. degree from Tokyo Metropolitan University, Tokyo, Japan, in 1996. He is currently a Professor with the Graduate School of Environment and Energy Engineering, Waseda University. He received the Prize Paper Award from IEEE Power Engineering Education Committee in 1991. He is a Senior Member of the IEE of Japan and a member of CIGRE. He is also a Convenor of Investigation Committee on Grid Technologies for large amount of Wind Power, IEEJ, and a Member of Investigation Committee on history of power system analysis, IEEJ. His research interests include simulation and analysis of power system and distribution power system.