IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 30 May 2024, accepted 11 June 2024, date of publication 19 June 2024, date of current version 18 July 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3416706

RESEARCH ARTICLE

Predictive Energy Management for Microgrid Using Multi-Agent Deep Deterministic Policy Gradient With Random Sampling

NIPHON KAEWDORNHAN[1] **AND RONGRIT CHATTHAWORN**[1,2]

[1]Department of Electrical Engineering, Faculty of Engineering, Khon Kaen University, Khon Kaen 40002, Thailand
[2]Center for Alternative Energy Research and Development, Khon Kaen University, Khon Kaen 40002, Thailand

Corresponding author: Rongrit Chatthaworn (rongch@kku.ac.th)

This work was supported in part by the Research and Graduate Studies, Khon Kaen University; and in part by the National Research Council of Thailand (NRCT) under Contract N41A661146.

ABSTRACT In a MicroGrid (MG) equipped with a Battery Energy Storage System (BESS), an Energy Management System (EMS) plays a crucial role in predictive controlling BESS operations for optimal power flow among uncertainties from renewable energy resources and heavy loads, such as solar photovoltaic systems and electric vehicles, respectively. State-of-the-art EMS designs have integrated Deep Reinforcement Learning (DRL) for EMS development and Probabilistic Power Flow (PPF) for preventing the violation of power system constraints while accounting for all uncertainties. However, using PPF to handle uncertainties alongside training a single-agent DRL provides the optimal solution for addressing all uncertain scenarios, but not the best solution for each scenario. Moreover, employing a single-agent DRL yields a low performance in predictive controlling of BESS operations. To address these challenges, a multi-agent DRL based on Deep Deterministic Policy Gradient (DDPG) is proposed. This method divides the roles of each agent for predicting 24-hour-ahead actions in BESS control based on changing 24-hour-ahead MG behavior every hour. Furthermore, MG parameters are randomly sampled to retain MG uncertainties instead of relying on PPF for uncertainty mitigation. Consequently, multi-agent DDPG with random sampling can directly learn from the MG environment and provide the best solution for each scenario. Simulation results demonstrate that the proposed method can reduce training computational time by 92.84%, provide a higher value of the summed mean of 24-hours-ahead reward by 1.50% to 28.37%, and achieve a lower mean daily total related cost by 9.22% compared to applying the state-of-the-art method.

INDEX TERMS Battery energy storage, deep reinforcement learning, electric vehicle, microgrid, photovoltaics, predictive energy management.

LIST OF ABBREVIATION AND NOMENCLATURE

BAS	Best Action Sequence.
BESS	Battery Energy Storage System.
DAP	Day-Ahead Profile.
DDPG	Deep Deterministic Policy Gradient.
DPF	Deterministic Power Flow.
DRL	Deep Reinforcement Learning.
EMS	Energy Management System.
EV	Electric Vehicle.
FIT	Feed-In-Tariff rate.
HAMG	Hour-Ahead MG behavior.
HST	Hottest-Spot Temperature.
MG	MicroGrid.
MGO	Microgrid Operator.
MRS	Mean of the Reward Sequence.
PDF	Probability Density Function.
PEM	Predictive Energy Management.
PF	Power Flow.
PPF	Probabilistic Power Flow.
pu	per unit.

The associate editor coordinating the review of this manuscript and approving it for publication was Vitor Monteiro.

RER	Renewable Energy Resource.
SoC	State-of-Charge of BESS.
solar PV	solar Photovoltaic.
TOU	Time-Of-Use rate.
a_t	Action of the agent.
B_{nm}	Susceptance (pu) of the line connected between the bus n and m.
C_t^{BESS}	BESS degradation cost ($) at hour t.
C_{Cap}^{BESS}	Capital cost ($) of BESS.
$C_{SoC_t^{BESS}}^{BESS}$	BESS degradation cost ($) at SoC_t.
C_t^{CO2}	Carbon emission cost ($) at hour t.
$C_t^{Exchanged}$	Exchanged energy cost ($) at hour t.
$CO2_t^{rate}$	Carbon capture rate ($/kWh) at hour t.
d_i^{EV}	Traveling distance (km) of the EV i.
E_{Cap}^{BESS}	Capital BESS capacity (kWh).
E_t^{BESS}	BESS energy (kWh) at hour t.
$E_{i,cap}^{EV}$	Battery capacity (kWh) of the EV i.
FIT_t	FIT rate ($/kWh) at hour t.
G_{nm}	Conductance (pu) of the line connected between the bus n and m.
$HST_{\max}^{Tr}$	Maximum HST of the transformer (°C).
HST_t^{Tr}	HST of the transformer (°C) at hour t.
$I_{l,t}$	Current magnitude (kA) of the line l at hour t.
k_i^{EV}	Type of the EV i.
m, n	Positions of the bus in the MG.
N	Number of buses in the MG.
P_{rated}^{BESS}	Rated BESS power (kW).
P_t^{BESS}	BESS power (kW) at hour t.
P_t^{MG}	MG's Active power (kW) at hour t.
P_n^g, P_n^d	Generated active power (pu) and Demand active power (pu) at the bus n.
P_{rated}^{PV}	Solar PV's rated output power (kW).
P_t^{PV}	Solar PV's power (kW) at hour t.
Q_n^g, Q_n^d	Generated reactive power (pu) and Demand reactive power (pu) at the bus n.
r_{decay}	Decay rate of the agent training.
R_t^{Cell}	Radiation ratio of the solar radiation at hour t.
R_t	Reward of the agent at hour t.
S_t, S_{t+1}	Current and Next state of the agent.
S_t^{Tr}	Transformer loading (kVA) at hour t.
$SoC_{\max}^{BESS}$	Maximum SoC of the BESS.
$SoC_{\min}^{BESS}$	Minimum SoC of the BESS.
SoC_t^{BESS}	SoC of BESS at hour t.
$SoC_{i,arr}^{EV}$	EV's SoC at the arrival time.
$SoC_{i,dep}^{EV}$	EV's SoC at the departure time.
$t_{i,Ch}^{EV}$	Charging time of the EV i.
$t_{i,dep}^{EV}$, $t_{i,arr}^{EV}$	Departure and Arrival times of the EV i.
$t_{i,StartCh}^{EV}$	Starting charging time of the EV i.
$t_{i,stay}^{EV}$	The hour of the EV's stay at the house.
$t_{i,StopCh}^{EV}$	Stopping charging time of the EV i.
T_t^{Amb}	Ambient temperature (°C) at hour t.
T_{STC}^{Cell}	Temperature (°C) on the solar cell at the Standard Test Condition.
T_t^{Cell}	Temperature (°C) on the solar cell at hour t.
T^{Nor}	Solar cell temperature at the nominal operating (°C).
TOU_t	TOU rate ($/kWh) at hour t.
$V_{n,t}$	Voltage magnitude (pu) of the bus n at hour t.
x_t^{BESS}	BESS control signal at hour t.
$y_{t,x}$	Constraint parameter x at hour t.
$y_x^{\max}$, $y_x^{\min}$	Maximum and Minimum values of the constraint parameter x.
ε_i^{EV}	Consumption rate (kWh/km) of EV i.
δ_{nm}	Angle bus voltage between the bus n and m.
η_{ch}^{BESS}, η_{dis}^{BESS}	Charging and Discharging coefficients of the BESS.
η_g	Overall generation efficiency of the solar PV system.
ρ_t	Penalty value at hour t.
μ	Self-discharge coefficient of BESS.
α_0, α_1, α_2	Curve-fitting coefficients of the BESS.
α_P	Solar PV's output power coefficient related to the temperature (kW/°C).
φ_t^{BESS}	Number of the BESS life cycle at hour t.
χ_t, χ_{cer}, χ_{std}	Solar radiation (W/m^2) at hour t, the one at a certain radiation point, and the one at the Standard Test Condition.

I. INTRODUCTION

Carbon emission is a pervasive challenge that has been emphasized by numerous governments. A significant contributor to this predicament is the widespread use of Internal Combustion Engine (ICE) vehicles in urban areas. Along with the usage of ICE vehicles, most electricity generation relies on fossil fuels. Notably, fossil fuels, particularly natural gas and coal, account for approximately 60% of all electricity generation [1], contributing to a substantial 25% of overall GreenHouse Gas (GHG) emissions [2]. A strategic approach to confine carbon and GHG emissions involves promoting electricity generation from RERs, specifically solar PV system, and the adoption of EVs as a substitute for traditional ICE vehicles [3].

However, the integration of EVs and solar PVs in the distribution system leads to the fluctuation of the power flow due to the uncertain power generation of solar PVs and disorderly EV charging. This results in violating operating system constraints, such as transformer overloading, undesired peak loads, and violations of bus voltage and line current [4], [5]. To address these issues, installing a BESS along with a robust EMS is the most effective way to handle power fluctuations in the distribution system integrated by EVs and solar PVs. Therefore, the concept of using BESS is applied

to small power systems to achieve self-sufficiency, enabling them to manage power fluctuations. This concept is known as a MG [4].

The design of EMS within MGs can be categorized into two primary approaches, including a rule-based EMS and an optimization-based EMS [6]. Several research studies have applied rule-based EMS in a single MG and multiple MGs. For example, Chakraborty et al. [7] developed a complex rule-based EMS for grid-connected MGs with solar PVs and BESSs, emphasizing reliability, real-time operation, and cost optimization. Their research focused on maximizing MGO profit through BESS control while minimizing real-time energy costs and ensuring power quality regulation. Kyriakou et al. [8] employed a fuzzy logic system-based rule-based EMS to control EV charging/discharging and determine optimal power set-points for all MGs. The objective was to minimize daily operation costs under fluctuating electricity prices and related constraints. Moreover, Kurukuru et al. [9], Jafari et al. [10] and Teo et al. [11] presented a fuzzy logic control-based rule-based EMS to manage energy in the MGs. However, the rule-based EMS has been demonstrated to present its capability to handle the complexity inherent in nonlinear control with low computational time for real-time control. While it can lessen the computational time by employing the if-else concept or fuzzy membership function, there is a potential risk of becoming confined to a local optimum.

In recent years, there has been significant research focus on optimization-based EMS development using DRL. The DRL can construct a well-trained Deep Neural Network to obtain the best action for MG energy management. DRL training relies on trial-and-error and does not depend on the population concept, leading to lower computational times compared to the application of metaheuristics [12], [13], [14]. The metaheuristics are discovered in state-of-the-art research related to EMS design in [15], [16], [17], [18], [19], [20], [21], and [22]. The DRL was applied in several research works related to MG energy management. For a single-agent DRL, Goh et al. [23] proposed using a single-agent DRL for dispatching the BESS installed in a MG. This research work utilized a single-agent based on Double Deep Q-learning Network and Policy Gradient with a novel reward function, termed a multistage reward function. Harrold et al. [24] proposed a single-agent DRL based on Rainbow Deep Q-Networks, a type of DRL, for dispatching BESS installed in the MG. Simulation results showed that Rainbow Deep Q-Networks outperforms actor-critic and linear programming methods, resulting in increased revenue due to solar PV generation, MG load, and Real-Time Pricing. Yu et al. [25] introduced a single-agent DRL based on DDPG for controlling BESS and heating, ventilation, and air conditioning in smart homes, leading to reduced energy costs for smart homes. Kolodziejczyk et al. [26] proposed a single-agent based on adaptive DRL using a Q-learning algorithm combined with a dense deep neural network for real-time control of BESS integrated with solar PV in the MG. Kaewdornhan and Chatthaworn [14] presented a single-agent DRL based on DDPG to obtain the optimal solution for BESS control while minimizing costs associated with exchanged energy, carbon emissions, and BESS degradation.

For a multi-agent DRL, Foruzan et al. [27] applied multi-agent DRL based on Q-learning to formulate an optimal energy management strategy without relying on prior information. Each supplier or customer is modeled as a single-agent to interact and share information for the collective benefit of all agents. Guo et al. [12] had compared different types of DRL, including DDPG, Deep Q-Network, and Dueling Deep Q-learning Network, for EMS development as a multi-agent DRL for power exchange management between multiple MGs connected in a distribution system. The numerical outcomes indicated that a multi-agent DDPG could provide the lowest overall costs. Kaewdornhan et al. [13], [28] proposed a multi-agent DRL based on DDPG for applying an optimal energy management in multiple MGs considering the uncertainties of solar PV generation, home demand, and EV charging demand. Salari et al. [29], [30] introduced a multi-agent DRL based on fuzzy Q-learning method to optimize EV charging integrated with solar PV generation for multiple homes, resulting in cost savings. Monfaredi et al. [31] proposed a multi-agent based on various DRL types to optimize the energy management of multiple energy carrier MGs in grid-connected mode. Each agent represents a single MG, collaborating with others to minimize total emissions and operating costs. Huang et al. [32] applied a multi-agent DRL based on various DRL types to meet the requirements of the manufacturing system while reducing operational costs for multiple MGs. All energy supplies, represented as MGs, are treated as independent agents to minimize associated costs.

In previous studies focusing on a single-agent DRL for MG's EMS design and optimal energy management, certain limitations were identified. These studies typically did not include power flow calculations while considering uncertainties associated with RERs and EVs, raising concerns about potential violations of MG system constraints. However, in [14], a novel approach was introduced using a single-agent DRL with the PPF to address power system constraints with a high confidence level and handle uncertainties related to RERs and EVs, offering an optimal solution for BESS control. Although this optimal solution can apply to all scenarios, it may not be the best solution for each scenario. Additionally, previous studies had not developed comprehensive models of EV charging behavior that are utilized in the training process of DRL. Meanwhile, research on multi-agent DRL has focused on its application across multiple MGs to facilitate interaction and coordination, effectively managing energy in interconnected MG systems. However, the use of multi-agent DRL within a single MG operator for BESS control or to address predictive energy management challenges specific to a single MG has not been extensively explored in the existing literature.

To address current challenges in the field, this paper proposes a multi-agent DRL based on DDPG (referred to as multi-agent DDPG), which has been identified as the most effective DRL for MG's EMS design in studies [12], [13], [14], [28], to control BESS installed in a single MG. By the multi-agent concept, we employ 24 agents (corresponding to the number of hours in a day), each capable of predicting BESS control actions 24 hours in advance. Furthermore, a comprehensive EV charging model is developed to facilitate the training of the multi-agent DDPG. The training process for the multi-agent DDPG directly utilizes random sampling for scenario generation instead of employing PPF. This enables the multi-agent DDPG to effectively learn uncertainty scenarios directly from random sampling, leading to best solutions for BESS control. The MG in this study represents a low-voltage distribution system comprising 27 houses, each equipped with rooftop solar PV panels and a single EV. The objective is to minimize exchanged energy, BESS degradation, and carbon emission costs while considering operational constraints of the power system. The main contributions of this paper are summarized as follows:

1) This paper introduces a novel EMS for a low-voltage distribution system developed as a MG. The novel EMS is constructed by implementing multi-agent DDPG for day-ahead energy management (PEM) in a single MGO, which has not been explored in current state-of-the-art research. The MG incorporates solar PV rooftops and EVs, defined as alternative elements in accordance with the Thailand government's plan. Furthermore, this research considers the real behavior of EV usage generated by using a comprehensive EV model that takes into account departure time, arrival time, traveling distance, EV type, SoC of the EV's battery, rated power charging, and the TOU rate, which has not been explored in previous research works, alongside the PEM optimization.

2) The novel MG behavior is proposed and used for training the multi-agent DDPG of a single MGO. The 24-HAMG is generated by random sampling and changes every hour, resulting in 24 sets of 24-HAMG in a day. To enhance the performance of the MGO's decision-making in BESS control, each agent in the multi-agent DDPG (24 agents) receives each 24-HAMG generated to predict the 24-hour-action in BESS control. By dividing roles among agents, each agent can focus more on discovering the best policy in BESS control when importing each set of 24-HAMG. Therefore, although the 24-HAMG changes every hour, implementing the multi-agent DDPG can enhance the decision-making performance of the MGO.

3) The proposed method, termed multi-agent DDPG (24 agents) with random sampling, is employed to address the PEM problem. This method demonstrates the adaptability of DDPG agents in handling uncertainties of 24-HAMG directly without the need for PPF and offers the best solutions for dispatching BESS. Implementing the PPF consumes a high training time and does not provide the best solution for each 24-HAMG. Therefore, the proposed method is compared with a single-agent DDPG utilizing PPF to demonstrate better performance for PEM task in the MG.

This paper is organized as follows. Section II represents the PEM architecture. Section III presents the PEM problem formulation, while Section IV demonstrates the DDPG problem formulation. Section V shows the methodology of this work. Simulation results are presented in Section VI, while the discussions are shown in Section VII. Finally, Section VIII demonstrates the conclusions of this work.

II. PREDICTIVE ENERGY MANAGEMENT ARCHITECTURE

The energy management task can be categorized into two types: Real-Time Energy Management and PEM. This paper focuses on the PEM task, emphasizing the use of DDPG to construct a well-trained Deep Neural Network. There are two PEM frameworks discussed: a conventional PEM framework referenced in [13] and [14], and a proposed PEM framework presented in this study.

A. CONVENTIONAL PEM FRAMEWORK

For the state-of-the-art studies, a single-agent DDPG with PPF is applied to provide the optimal solution for BESS control with a high confidence level regarding elements and operation system constraints. The operation of this approach is depicted in Fig. 1. Firstly, the agent undergoes thorough training using the DDPG optimization process. The agent receives comprehensive information from the main grid, including TOU rates and carbon emission rates, to predict the optimal action for each hour. Next, the action is transmitted to the BESS to determine the charge/discharge BESS power. If the discharge power from the BESS is insufficient to meet the MG demand, the MG will supplement with power from the transformer along with the BESS power. In certain scenarios where the MG experiences surplus power due to high solar PV generation, the BESS will charge using part of the surplus power and inject the remaining surplus power into the transformer. Afterward, the PPF process will operate to calculate the mean and standard deviation (Std.) of the system parameters, such as the maximum/minimum bus voltage, and the maximum line current. The scenarios for PPF calculation are generated using the Nataf transformation theory, which utilizes PDFs of all random variables in the analysis. Each scenario is utilized to estimate the net power in each house for preparation in the DPF based on unbalance PF calculations. After the DPF calculations are completed, the desired system parameters are stored and utilized to estimate their mean and standard deviation using a point estimation method. Consequently, the number of DPF calculations corresponds to the number of generated scenarios, a process known as the PPF loop. Next, the estimated system parameters (Mean and Std.) are sent to calculate the reward with a high confidence level. The related parameters are transferred to the agent again in the next step. Finally, the agent updates the hour to provide the next action. The above process continues until the final hour of the day is reached. By applying the PPF, the Mean and Std of desired parameters are evaluated

to provide feedback to the agent. These parameters are calculated to encompass all possible uncertainties of random variables within the MG. To this end, direct learning with MG uncertainty is not feasible for the agent. Therefore, the agent must learn to provide optimal solutions for BESS control that cover the Mean and Std of these parameters. The optimal solution provided can be applied across all uncertain scenarios encountered in the PEM task.

B. PROPOSED PEM FRAMEWORK

In the conventional PEM framework, the PPF restricts the agent's ability to learn about MG uncertainty directly. Therefore, to enable direct learning with MG uncertainty, a random sampling process is applied to generate scenarios. In the absence of a tool to mitigate MG uncertainty, transitioning from a single-agent to a multi-agent (24 agents) framework enhances learning in MG uncertainty and improves the performance of predictive BESS operation control. The proposed PEM framework is shown in Fig. 2.

From Fig. 2, firstly, 24 agents are constructed based on the DDPG structure, labeled as the 0^{th}, 1^{st}, ..., 23^{rd}, respectively. The selected agent receives comprehensive 24-hour-ahean information, such as TOU and carbon rates from the main grid, which prepare them to determine and execute the best course of action. Then, the agent will predict the best action sequence (24-hour-ahead action) when recognizing the 24-HAMG. For the 24-HAMG generation, the historical data of the MG parameters, including solar radiation, ambient temperature, appliance load, and EV charging demand, are fitted to PDFs. Next, the PDF of each MG parameter is utilized in random sampling to generate 24 sets of 24-hour-ahead profiles for each MG parameter. Each set has the initial hour beginning from 0^{th} to 23^{rd} hour. The 24-hour-ahead profiles of all MG parameters, starting at the same initial hour, are stored together as a set for simulating MG behavior. This set is referred to as a single 24-HAMG. Thus, the MG has 24 sets of 24-HAMGs. These sets are called a single DAP. In this work, several DAPs are applied in training and testing process of the agents. For a single DAP, the 24 agents (multi-agent), labeled from 0^{th} to 23^{rd}, will be selected to predict the action sequence for BESS control when each 24-HAMG is imported. Each agent is selected based on the initial hour of the imported 24-HAMG, and each agent is responsible for predicting the BAS when provided with the imported 24-HAMG.

When the agent is selected to predict the BAS. In this work, the 24-HAMG changes 24 times a day (every hour), defined as the worst case in the PEM task. Thus, the first action of the BAS is applied to the BESS model to evaluate the BESS power, similar to the conventional PEM framework. Then, the system parameters and transformer loading are estimated by the DPF with unbalanced PF, while the BESS parameters are calculated. Next, The related parameters are used to calculate the reward. Subsequently, the related parameters are transferred to the agent to prepare for the next BAS prediction, along with updating the 24-HAMG. This process continues

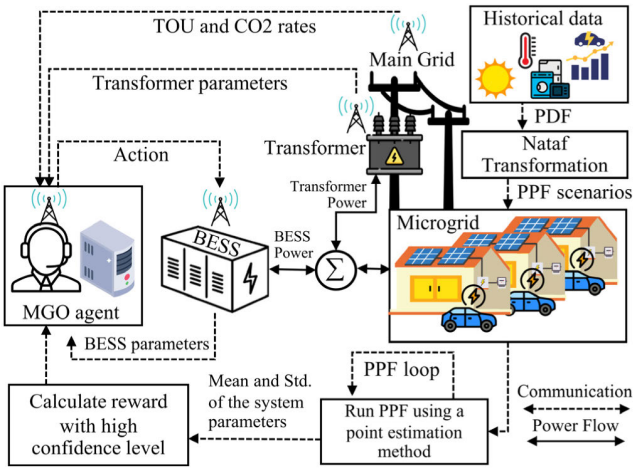

FIGURE 1. Conventional PEM framework.

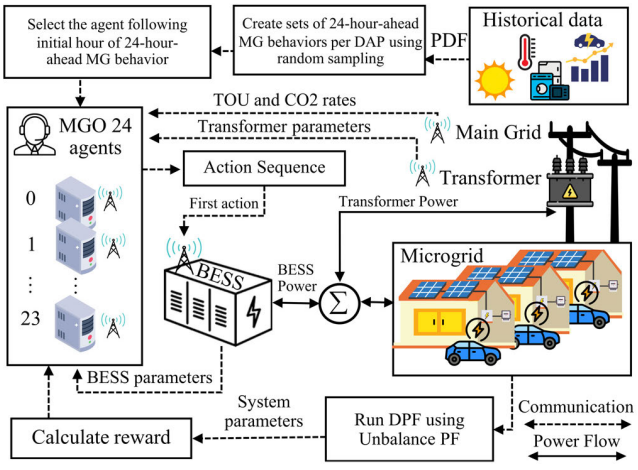

FIGURE 2. Proposed PEM framework.

until the 24-HAMG reaches the final set, and it operates similarly in every generated DAP.

Therefore, multi-agent DDPG constructed can be applied in the PEM task of the MG through the proposed PEM framework. Moreover, the multi-agent can directly learn about the uncertain DAP generated without the need for PPF, leading to lower computational requirements and providing the best solution for each uncetian scenario.

III. PEM PROBLEM FORMULATION

To construct the PEM optimization problem, the considered objective functions and related constraints are formulated. Moreover, the daily optimization problem is represented in this section.

A. OBJECTIVE FUNCTIONS

Three objective functions are minimized in this paper, including the exchanged energy cost, BESS degradation cost, and the carbon emission cost.

1) EXCHANGED ENERGY COST

Since the MG is set as a grid-connected mode, the energy exchanged between the MG and the main grid can reflect to the electricity cost/revenue. The equation of this objective function can be formulated as follows [13]:

$$C_t^{Exchanged} = \begin{cases} TOU_t \cdot P_t^{MG}; & P_t^{MG} \geq 0 \\ FIT_t \cdot P_t^{MG}; & P_t^{MG} < 0 \end{cases} \quad (1)$$

There are two conditions in this objective function: if P_t^{MG} is greater than or equal to 0 kW, the MG is receiving power from the main grid, leading to a positive cost occurrence. In contrast, if P_t^{MG} is less than 0 kW, the MG is injecting power into the main grid, resulting in revenue occurrence from the injected energy, which, in the context of cost, is a negative value.

2) BESS DEGRADATION COST

The BESS degradation cost is an essential factor that can reflect the proper operation of the BESS. Adding this objective function to the optimization process can prevent the disorganized control of the BESS by the agent DDPG, leading to a decreasing loss-of-life of the BESS. The BESS degradation cost can be represented as follows [33]:

$$C_t^{BESS} = \begin{cases} C^{BESS}_{SoC_t^{BESS}} - C^{BESS}_{SoC_{t-1}^{BESS}}; & P_t^{BESS} > 0 \\ 0; & P_t^{BESS} \leq 0 \end{cases} \quad (2)$$

$$C^{BESS}_{SoC_t^{BESS}} = \frac{C_{Cap}^{BESS}}{\varphi_t^{BESS}} \quad (3)$$

$$\varphi_t^{BESS} = \alpha_0 \times (1 - SoC_t^{BESS})^{-\alpha_1} \times \exp\left(\alpha_2 \times SoC_t^{BESS}\right) \quad (4)$$

From the equation (2), two conditions are used to estimate the BESS degradation cost: if P_t^{BESS} is more than 0 kW, the BESS is discharging, leading to a positive cost occurrence. In contract, if P_t^{BESS} is less than or equal to 0 kW, charging model of the BESS is operating, resulting in the cost equal to 0. Therefore, the BESS degradation cost is calculated when the BESS mode is in discharge mode.

3) CARBON EMISSION COST

Due to the increasing carbon emissions in the electricity generation sector, the carbon emission cost is applied in several research studies related to MG energy management. This cost can prevent disorganized-drawn power from the main grid. In this paper, the carbon emission cost will be calculated when the MG absorbs power from the main grid because most electricity generation from the main grid uses fossil fuels, which can produce carbon emissions. The objective function can be calculated as follows [13]:

$$C_t^{CO2} = \begin{cases} \mathrm{CO2}_t^{rate} \cdot P_t^{MG}; & P_t^{MG} \geq 0 \\ 0; & P_t^{MG} < 0 \end{cases} \quad (5)$$

There are two conditions for C_t^{CO2} calculation: if the P_t^{MG} is greater than or equal to 0 kW, the MG is absorbing power from the main grid, leading to a positive cost occurrence, otherwise, the cost will equal 0.

B. CONSTRAINTS

To provide solutions within a feasible space, related constraints are added to the optimization problem. In this work, constraints related to the system operation, BESS operation, and transformer operation are considered.

1) SYSTEM OPERATION CONSTRAINT

To control power flow in the MG properly, the agent has to provide the solution to balance the power within the MG. Thus, the power balance equations in the context of active and reactive power balance are taken in the optimization process. The power balance equations can be formulated as follows [13]:

$$P_n^g - P_n^d = \sum_{m=1}^{N} V_n V_m \left(G_{nm} \cos \delta_{nm} + B_{nm} \sin \delta_{nm}\right) \quad (6)$$

$$Q_n^g - Q_n^d = \sum_{m=1}^{N} V_n V_m \left(G_{nm} \sin \delta_{nm} + B_{nm} \cos \delta_{nm}\right) \quad (7)$$

The bus voltage and line current limits are concerned in this work to prevent voltage and current violations. These limits can be performed as follows [34]:

$$0.9 \text{ pu} \leq V_{n,t} \leq 1.1 \text{ pu} \quad (8)$$

$$\left|I_{l,t}\right| \leq 0.35 \text{ kA} \quad (9)$$

Since the MG is tested in Thailand, the standard operation of the Provincial Electricity Authority (PEA) supervising the low-voltage distribution system is employed. Thus, $V_{n,t}$ is limited at [0.9, 1.1] pu and $I_{l,t}$ is limited which does not exceed 0.35 kA [34].

2) BESS OPERATION CONSTRAINT

To meet the BESS operation security, the BESS is controlled under its operation limit. There are the fundamental limits that are set as the BESS operation limits: the SoC and BESS power. These variables can be formulated as follows [13]:

$$SoC_t^{BESS} = (1 - \mu)\, SoC_{t-1}^{BESS} - \frac{E_t^{BESS}}{E_{Cap}^{BESS}} \quad (10)$$

$$E_t^{BESS} = \begin{cases} P_t^{BESS} \eta_{ch}^{BESS}; & P_t^{BESS} \leq 0 \\ P_t^{BESS} / \eta_{dis}^{BESS}; & P_t^{BESS} > 0 \end{cases} \quad (11)$$

The E_t^{BESS} is calculated by using equation (11), which has two conditions: if the P_t^{BESS} is less than or equal to 0 kW (charging), the first condition will be used to calculate E_t^{BESS}. In contrast, if the P_t^{BESS} is more than 0 kW (discharging), the second condition will be applied to calculate E_t^{BESS}.

In the context of the BESS operation limits, the SoC_t^{BESS} and P_t^{BESS} are limited within the BESS operation security as follows [13]:

$$SoC_{\min}^{BESS} \leq SoC_t^{BESS} \leq SoC_{\max}^{BESS} \quad (12)$$

$$\left|P_t^{BESS}\right| \le P_{rated}^{BESS} \tag{13}$$

3) TRANSFORMER OPERATION CONSTRAINT

A distribution transformer is the essential element in the MG and has to work in the proper operation zone to reduce the transformer's loss of life. The HST is a critical parameter in forecasting the transformer's loss of life [35]. It is determined based on the transformer loading (kVA) and the ambient temperature (°C). When the transformer is worked to overloading and the ambient temperature rises, there is a notable increase in the HST. Therefore, relying solely on transformer loading (kVA) defined as the transformer constraint may not suffice to ensure the transformer's loss of life. Consequently, HST emerges as a more reliable indicator for safeguarding the transformer's loss of life. The HST can be mathematically formulated as follows:

$$HST_t^{Tr} = f\left(S_t^{Tr}, T_t^{Amb}\right) \tag{14}$$

$$HST_t^{Tr} \le HST_{\max}^{Tr} \tag{15}$$

The $f(\cdot)$ is the HST's calculation function, it is shown in [35].

C. DAILY OPTIMIZATION PROBLEM

When the agent recognizes 24-HAMG imported, the agent forecasts the 24-hour-ahead actions for BESS control to minimize the total cost within the same timeframe, akin to solving a daily optimization problem. The formulation for minimizing the daily total cost is presented as follows:

$$\min_{P_{0\to23}^{BESS}} \sum_{t=0}^{23}\left(C_t^{Exchanged} + C_t^{BESS} + C_t^{CO2}\right) \tag{16}$$

where related constraints, including equations (6)-(9) and equations (12)-(15), are considered alongside the minimization process.

IV. DDPG PROBLEM FORMULATION

In this paper, conventional and proposed methods based on DDPG optimization are employed to solve the PEM problem. Therefore, these problem are mapped into the DDPG problem formulation, which are presented in this section.

A. CONVENTIONAL PROBLEM FORMULATION

In the conventional problem, a single-agent DDPG interacts with the MG environment on an hourly basis, with all system parameters provided by the PPF. This approach yields an optimal solution for each hour, which can be applied to the PEM task. To implement this concept, the daily optimization problem is transferred into an hourly DDPG problem. Subsequently, the sum of all hourly costs yields the daily total cost, mirroring the structure of a daily optimization problem. The hourly DDPG problem involves four key variables: current state, action, reward, and next state. These variables can be formulated as follows:

$$S_{t,PPF} = S_{PPF}(t) = \begin{bmatrix} S_{t,PPF}^{Tr} \\ P_{t,PPF}^{MG} \\ P_{t,PPF}^{BESS} \\ SoC_{t,PPF}^{BESS} \\ TOU_t \\ FIT_t \\ CO2_t^{rate} \\ t \end{bmatrix} \tag{17}$$

$$a_{t,PPF} = \left[x_{t,PPF}^{BESS}\right]; x_{t,PPF}^{BESS} \in [-1, 1] \tag{18}$$

$$P_{t,PPF}^{BESS} = a_{t,PPF} \cdot P_{rated}^{BESS} \tag{19}$$

$$R_{t,PPF} = -\left(C_{t,PPF}^{Exchanged} + C_{t,PPF}^{BESS} + C_{t,PPF}^{CO2} + \rho_{t,PPF}\right) \tag{20}$$

$$\rho_{t,PPF} = \omega \sum_{x=1}^{N_y} \ln\left(\frac{\left|y_x^{\max} - y_{t,x}^{PPF}\right| + \left|y_x^{\min} - y_{t,x}^{PPF}\right|}{y_x^{\max} - y_x^{\min}}\right) \tag{21}$$

$$S_{t+1,PPF} = S_{PPF}(t+1) \tag{22}$$

where the current state, action, reward, and next state at hour t are represented as $S_{t,PPF}$, $a_{t,PPF}$, $R_{t,PPF}$, and $S_{t+1,PPF}$, respectively. The parameters determined as the current state and the next state are normalized within the boundary $[-1, 1]$. Moreover, the variable subscripted by PPF are the variable estimated from the PPF process. In the context of the $a_{t,PPF}$, it is set within $[-1,1]$, which can be shown that if the $a_{t,PPF}$ is provided within $[-1,0)$, the BESS is charged with $P_{t,PPF}^{BESS}$ calculated by equation (19). Otherwise, the BESS is discharged with $P_{t,PPF}^{BESS}$. For the $R_{t,PPF}$, the negative sum of the relevant objective functions is mapped to the $R_{t,PPF}$, incorporating any associated the penalty $\rho_{t,PPF}$.

The $\rho_{t,PPF}$ can be calculated by using equation (21), which is defined as a violating index in the optimization task. If the violation occurs due to improper BESS control, the $\rho_{t,PPF}$ will become the large positive value with the penalty coefficient ω, resulting in the reward having the negative infinity value. Otherwise, the $\rho_{t,PPF}$ will be 0. Therefore, the agent will try to control the BESS for dealing with the $\rho_{t,PPF}$ along with the related cost minimization, leading to providing the highest reward value. Furthermore, $y_{x,i}^{PPF}$ is the related constraint parameter x at hour t, evaluated by the PPF.

B. PROPOSED PROBLEM FORMULATION

In the proposed problem,a multi-agents is allowed in direct learning with the MG environment without employing the PPF for MG uncertainty mitigation. The 24-HAMGs are generated through random sampling from PDFs of random variables. Additionally, each agent plays a distinct role in predicting the 24-hour-ahead actions for BESS control based on the initial hour of the imported 24-HAMG. However, when each agent is constructed according to the DDPG optimization, the formulation of the DDPG variables remains similar to the conventional problem but has an index p to indicate

each agent and the hour t that progresses according to the time of the imported 24-HAMG. The DDPG variables of the proposed problem are formulated as follows:

$$S_{t,p} = S_p(t) = \begin{bmatrix} S_{t,p}^{Tr} \\ P_{t,p}^{MG} \\ P_{t,p}^{BESS} \\ SoC_{t,p}^{BESS} \\ TOU_t \\ FIT_t \\ CO2_t^{rate} \\ t \end{bmatrix} \tag{23}$$

$$a_{t,p} = \left[x_{t,p}^{BESS}\right];\ x_{t,p}^{BESS} \in [-1, 1] \tag{24}$$

$$P_{t,p}^{BESS} = a_{t,p} \cdot P_{rated}^{BESS} \tag{25}$$

$$R_{t,p} = -\left(C_{t,p}^{Exchanged} + C_{t,p}^{BESS} + C_{t,p}^{CO2} + \rho_{t,p}\right) \tag{26}$$

$$S_{t+1,p} = S_p(t+1) \tag{27}$$

where the current state, action, reward, and next state of the agent p are labeled as $S_{t,p}$, $a_{t,p}$, $R_{t,p}$, and $S_{t+1,p}$, respectively. The penalty factor $\rho_{t,p}$ is calculated using equation (21) and the parameters determined as the current state and the next state are normalized within the boundary $[-1, 1]$, similar to the conventional problem. However, the system parameters, defined as state variables, are estimated based on the 24-HAMG generated through a random sampling process. As a result, these parameters have varying values in each training iteration. These parameters have uncertainty levels higher than those estimated from the PPF process, resulting in an advantage for the direct learning of the agent p.

V. METHODOLOGY

To compare the training and testing processes of the conventional and proposed methods, this section outlines the procedures for constructing distribution models of random variables and conducting random sampling from these models. Additionally, the training and testing processes of both the conventional and proposed methods are described and compared.

A. DISTRIBUTION MODEL CONSTRUCTION

1) SOLAR PV GENERATION

To estimate the hourly solar PV's output power, hourly solar radiation (W/m^2) and hourly ambient temperature (°C) are sampled from the fitted PDFs to calculate the output power. The generation model for estimating the solar PV's output power can be represented as follows [13]:

$$P_t^{PV} = \eta_g P_{rated}^{PV} R_t^{Cell}\left[1 + \alpha_P\left(T_t^{Cell} - T_{STC}^{Cell}\right)\right] \tag{28}$$

$$R_t^{Cell} = \begin{cases} \chi_t^2 / (\chi_{cer}\chi_{std}) & \chi_t < \chi_{cer} \\ \chi_t / \chi_{std} & \chi_{cer} \le \chi_t \le \chi_{std} \\ 1 & \chi_t > \chi_{std} \end{cases} \tag{29}$$

$$T_t^{Cell} = T_t^{Amb} + R_t^{Cell}\left(T^{Nor} - 20\right) \tag{30}$$

The given detail indicates that there are two primary variables that vary each hour, including χ_t and T_t^{Amb}. To avoid fixed hourly values for χ_t and T_t^{Amb}, historical data of χ_t and T_t^{Amb} are fitted into a beta PDF [36], [37] and a normal PDF [36], respectively, to construct their PDFs for hourly random sampling. The formulation of the aforementioned PDFs can be shown as follows:

$$f(\chi) = \chi^{\alpha-1}(1-\chi)\beta\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)+\Gamma(\beta)} \tag{31}$$

$$f(T) = \frac{1}{\sigma_n\sqrt{2\pi}}\exp\left(-\frac{(T-\mu_n)^2}{2\sigma_n^2}\right) \tag{32}$$

where the radiation's (χ) beta PDF is represented as $f(\chi)$ with the random variable (α) and control deviation (β), while the gamma function is denoted as $\Gamma(\cdot)$. For the ambient temperature's (T) normal PDF, it is presented as $f(T)$ with the mean (μ_n) and standard deviation (σ_n), while the exponential function is represented as $\exp(\cdot)$.

2) APPLIANCE LOAD

To potentially cover the household's appliance load within the MG, historical data related to appliance load is fitted to a PDF for sampling and estimating the hourly household's baseload. Commonly, the normal PDF is utilized to fit the PDF of the household's appliance load [36], similar to the ambient temperature.

3) EV CHARGING DEMAND

One of this paper's contributions is to simulate the comprehensive charging behavior of the household's EV in the optimization process. Therefore, the EV usage is modeled using multi-variable construction consisting of the departure time, arrival time, traveling distance, EV type, SoC of the EV's battery, rated power charging, and TOU rate. Each EV charging scheduling in the MG can be shown in Fig. 3 and evaluated in the following steps:

Step 1: The departure $(t_{i,dep}^{EV})$ and arrival $(t_{i,arr}^{EV})$ times of the EV i are sampled from the fitted normal PDF, as referred to [38]. Therefore, the hour of the EV's stay at the house $(t_{i,stay}^{EV})$ is presented as follows:

$$t_{i,stay}^{EV} \in \left[t_{i,arr}^{EV}, t_{i,dep}^{EV}\right] \tag{33}$$

Step 2: The type (k_i^{EV}) and traveling distance (d_i^{EV}) of the EV i are sampled to label the EV type and EV travel. The EV types are randomly sampled using the survey data, as referred to [39], while the traveling distance is sampled from the fitted lognormal PDF, as referred to [38]. The lognormal PDF can be formulated as follows:

$$f(d_i^{EV}) = \frac{1}{d_i^{EV}\sigma_{Ln}\sqrt{2\pi}}\exp\left(-\frac{(\ln d_i^{EV} - \mu_{Ln})^2}{2\sigma_{Ln}^2}\right) \tag{34}$$

where the lognormal PDF function of the d_i^{EV} is represented as $f(d_i^{EV})$ with the mean (μ_{Ln}) and standard deviation (σ_{Ln}).

From the k_i^{EV} sampling, it makes to know the EV battery's capacity and consumption rate, while knowing d_i^{EV} can estimate the EV's SoC at the arrival time $SoC_{i,arr}^{EV}$. If the EV battery's capacity ($E_{i,cap}^{EV}$) and consumption rate (ε_i^{EV}) are determined, the $SoC_{i,arr}^{EV}$ will be evaluated as follows:

$$SoC_{i,arr}^{EV} = SoC_{i,dep}^{EV} - d_i^{EV} \cdot \varepsilon_i^{EV} \tag{35}$$

where the EV's SoC of the EV i at the departure time ($SoC_{i,dep}^{EV}$) is set as the maximum SoC ($SoC_{\max}^{EV}$).

Step 3: The starting time and stopping time for the charging of the EV i are labeled as the $t_{i,StartCh}^{EV}$ and $t_{i,StopCh}^{EV}$, respectively. The interval for EV charging is restricted to the time interval that intersects between the EV's stay at the house period ($t_{i,arr}^{EV}$ to $t_{i,dep}^{EV}$) and the off-peak TOU period. To this end, the interval time for EV charging will begin the $t_{i,StartCh}^{EV}$ to the $t_{i,StopCh}^{EV}$, but not beginning the $t_{i,arr}^{EV}$ to the $t_{i,dep}^{EV}$. This is the common behavior of the EV user who charges the EV within the off-peak TOU period. Thus, the hourly charging of the EV i denoted as $t_{i,Ch}^{EV}$ can be formulated as follows:

$$t_{i,Ch}^{EV} \in \left[t_{i,StartCh}^{EV}, t_{i,StopCh}^{EV}\right] = \left[t_{i,arr}^{EV}, t_{i,dep}^{EV}\right] \cap \forall_{t_{off-peak}^{TOU}} \tag{36}$$

Step 4: Set the charging time of the EV i ($t_{i,Ch}^{EV}$) equal to $t_{i,StartCh}^{EV}$.

Step 5: The EV's battery is charged with the rated power of the charger at $t_{i,Ch}^{EV}$.

Step 6: Update the SoC of the EV's battery at $t_{i,Ch}^{EV}$. The used SoC formulation is referred to [28].

Step 7: The $t_{i,Ch}^{EV}$ is regulated. If the $t_{i,Ch}^{EV}$ does not reach the $t_{i,StopCh}^{EV}$ or the $SoC_{i,ch}^{EV}$ is less than $SoC_{\max}^{EV}$, the $t_{i,Ch}^{EV}$ will be updated. Otherwise, the EV charging will be stopped.

From the above process, it can provide hourly power charging of the EV i in the MG. Therefore, each EV charging scheduling can be obtained by using the proposed process.

B. TRAINING AND TESTING PROCESS OF CONVENTIONAL METHOD

A single-agent DDPG with PPF, applied in [13] and [14] to optimally control the BESS installed the MG, is defined as the conventional method in this paper. The DDPG optimization concept involves four networks: actor, critic, target actor, and target critic networks. The actor performs continuous action prediction to control BESS, while the critic predicts the Q-value of each state, defined as the index in an action evaluation. The general Q-value formulation includes a discount factor to regulate the Q-value, preventing it from reaching infinity. If the action provides a higher reward, its Q-value is also higher. For the target networks, they are constructed to increase the convergence rate in the learning process of the actor and critic. Since DDPG learning relies on a trial-and-error process, learning rates are utilized to update the weights of the actor and critic networks, while soft updates are employed to update the weights of the target networks. Experiences from the environment interaction, consisting of state, action, reward, and next state, are grouped into a tuple and stored in the buffer memory. These experiences are sampled following the determined batch size for updating weights of all networks.

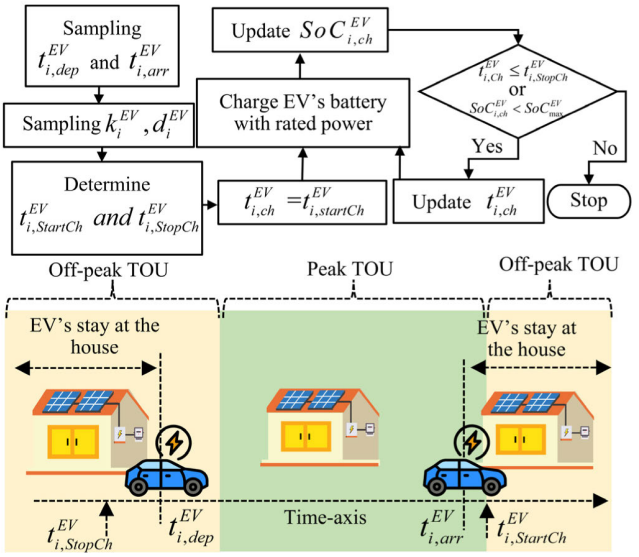

FIGURE 3. EV charging demand model.

As DDPG is a model-free data-driven method, the actor's weight update using gradient concepts are applied to determine the appropriate direction for updating. Moreover, the concepts of exploitation and exploration are utilized to enhance the performance of the agent's learning. The agent will choose actions with high Q-values at that state by following the best policy discovered up to that time, which represents the exploitation concept, while it must explore possible actions to potentially provide actions with higher Q-values by adding a random Gaussian noise to actions, this process is defined as the exploration concept. The exploration concept is formulated as follows [12]:

$$a_j = a_j^{bestPolicy} + G(0, \sigma_j^2) \tag{37}$$

$$\sigma_j = \exp\left(-j \times r_{decay}\right);\ j = [1, 2, \ldots, H \times iter] \tag{38}$$

where $a_j^{bestPolicy}$ is the action with high Q-values at that state by following the best policy discovered up to that time, while a_j is the action added by the noise. $G(0, \sigma_j^2)$ is a random Gaussian noise with a mean equal to zero and a standard deviation equal to σ_j. r_{decay} is a decay rate determined following the time slot H (24 hours) for BESS control and the training iteration $iter$. The value of σ_j is decreased as the training iteration increases according to the determined value of r_{decay}. Thus, the r_{decay} must be properly determined to decrease the impact of exploration before reaching the maximum training iteration, showing more detail according to [12].

From the DDPG optimization concept, the operation of a single-agent DDPG with PPF is set, and the training and testing processes are demonstrated as follows:

1) TRAINING PROCESS OF CONVENTIONAL METHOD

The training process of the conventional method can be depicted in Fig. 4 and operated in the following steps:

Step 1: The parameters of a single-agent DDPG are set, such as the number of hidden layers of all networks, activate function, learning rate, soft update factor, discount factor, decay rate, buffer memory size, batch size, and training iteration.

Step 2: The agent predicts the action at the initial hour to control BESS.

Step 3: When the action is taken into the BESS model, the BESS power (kW) is estimated to inject/absorb power into/from the MG.

Step 4: To evaluate the mean and standard deviation of system parameters or desired power flow parameters, the PPF is utilized to run. The PDFs of the solar radiation, ambient temperature, and appliance load, are imported to generate the PPF scenarios. Moreover, the hourly EV charging profiles for 10,000 vehicles are created following the Fig. 3. The profiles are fitted into a normal PDF based on a monte carlo simulation concept [39] and the above PDF is imported to generate the PPF scenarios along with the PDFs of other random variables.

By the PPF concept, the scenarios are generated based on the Nataf Transformation with Point Estimation Method (NTPEM), according to [39]. If the numbers of the input variables and point estimation are n and m, respectively, the number of PPF scenarios will be $(m-1)\,n+1$ per hour. In this paper, the number of input variables (random variables) is 4, and the number of points for estimation is 5. Therefore, the number of PPF scenarios per hour is 17. Consequently, the PPF will execute the DPF 17 times per hour, which is referred to as the PPF loop.

Step 5: While running the DPF each time, the desired parameters are stored and used to estimate their mean and standard deviation.

Step 6: Once the mean of the desired parameters are provided by the PPF, they are imported into the relevant objective function to compute the associated costs. In the conventional method, achieving a high confidence level to prevent violations is crucial for penalty calculation. This is achieved by utilizing the mean and standard deviation of the parameters in the penalty calculation. Commonly, a 95% confidence level is utilized in preventing violations of the power system [39]. Each parameter is considered both the mean ($y^{\mu}_{t,x}$) and standard deviation ($y^{\sigma}_{t,x}$). If the $y^{PPF}_{t,x}$ is the related constraint parameter x at the hour t, $y^{PPF}_{t,x}$ will be formulated $y^{\mu}_{t,x}+2y^{\sigma}_{t,x}$ for considering the maximum violation and $y^{\mu}_{t,x}-2y^{\sigma}_{t,x}$ for considering the minimum violation. Thus, a larger penalty value increases the likelihood of occurrence, making it more challenging for the agent to find an optimal solution. Subsequently, the calculated costs and penalties are mapped to the hourly reward.

Step 7: The hour is regulated. If the hour does not reach 24 hours, the desired-related parameters are mapped to the DDPG variables. Subsequently, the weights of all networks will be updated, providing further detail on the weight updating process as outlined in [12], along with the hour updating. Otherwise, the iteration (iter.) will be checked. If the iteration does not reach the maximum iteration (Max. iter.), it will be updated, and the hour will be reset to the initial hour. Otherwise, the weights of all networks will be saved, and the process will be stopped.

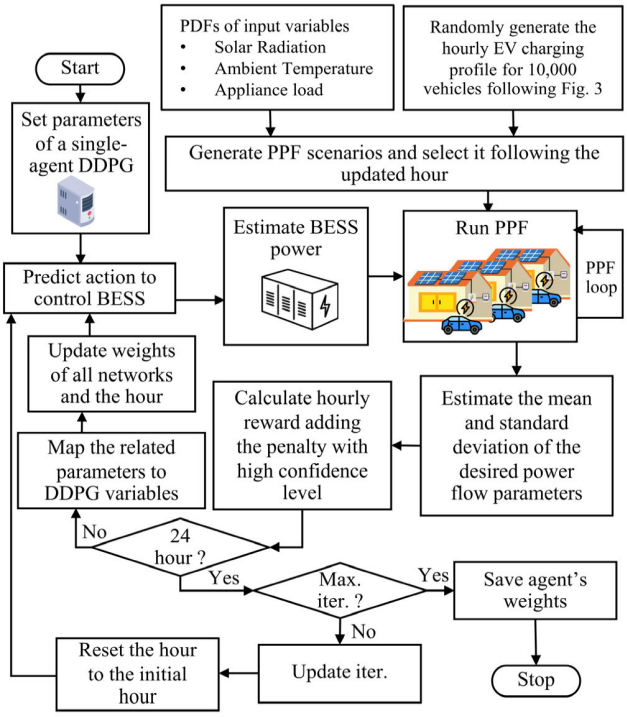

FIGURE 4. Training process of conventional method.

2) TESTING PROCESS OF CONVENTIONAL METHOD

From the training process, it provides a well-trained single agent. To obtain the hourly-optimal action for BESS controlling, the well-trained agent is tested which can be depicted in Fig. 5.

Firstly, load the well-trained agent and provide the optimal action to estimate BESS power at the initial hour. Next, the current-hourly PPF scenario is imported to run PPF. The estimation of the mean and standard deviation of the desired parameters is operated in the following step. Subsequently, the calculated-related parameters are mapped into the DDPG variables. Then, the hour will be checked. If the hour is not equal to 24, it will be updated, and the next action will be predicted for BESS control in the next hour. Otherwise, the hourly-optimal action is saved, and the process stops. This process enables the acquisition of hourly-optimal action, which is then utilized to verify the performance of the conventional method.

In the process of the performance verification, 1,000 DAPs and the verification process are generated to verify the hourly-optimal action obtained by the conventional method, which is shown in Fig. 6 and operated in the following steps:

Step 1: 1,000 DAPs are generated to use in the verification process. For a single DAP generation, the 24-hour-ahead profiles of solar radiation, ambient temperature, and appliance load are generated by using random sampling from their PDFs, while the 24-hour-ahead profile of EV charging demand is generated using the process in Fig. 3. These are generated 24 sets per a random variable, each having the initial hour beginning the 0^{th} hour until the 23^{th} hour. The 24-hour-ahead profile of all MG parameters that begins at the same initial hour is imported into the MG environment to simulate a single 24-HAMG. Therefore, 24 sets of 24-HAMG are generated, and these sets collectively form a single DAP. All DAPs are generated using the described process.

Step 2: Select the optimal action sequence following the hour sequence of the imported 24-HAMG.

Step 3: Take the action to the BESS model to estimate the BESS power. The action taken is a single action at the current hour within the selected action sequence.

Step 4: The PF-based unbalance is run to estimate the desired parameters.

Step 5: Related parameters are used to calculate the hourly reward.

Step 6: The hour will be checked. If the hour does not equal to 24 hours, the hour will be updated and the next action will be selected from the selected action sequence based on the updated hour. Otherwise, the reward at the initial hour of the imported 24-HAMG is stored, called the first reward.

Step 7: The number of 24-HAMG will be regulated. If the number of 24-HAMG does not reach 24, the next 24-HAMG will be imported. Then, the hour will be reset as the initial hour of the next 24-HAMG to control BESS for the next 24-HAMG. Otherwise, the reward at the initial hour of all 24-HAMGs saved in the sixth step is stored, called the first reward set.

Step 8: The number of DAPs will be checked. If it does not equal to 1,000, the DAP will be updated to the next DAP. Subsequently, the hour will be reset as the initial hour of the first 24-HAMG of the next DAP. Otherwise, the first reward set stored in the seventh step is saved to utilize as the performance index and then stops.

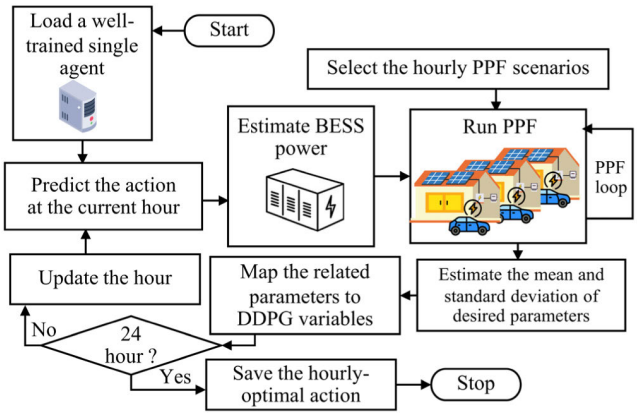

FIGURE 5. Hourly optimal action saving process of conventional method.

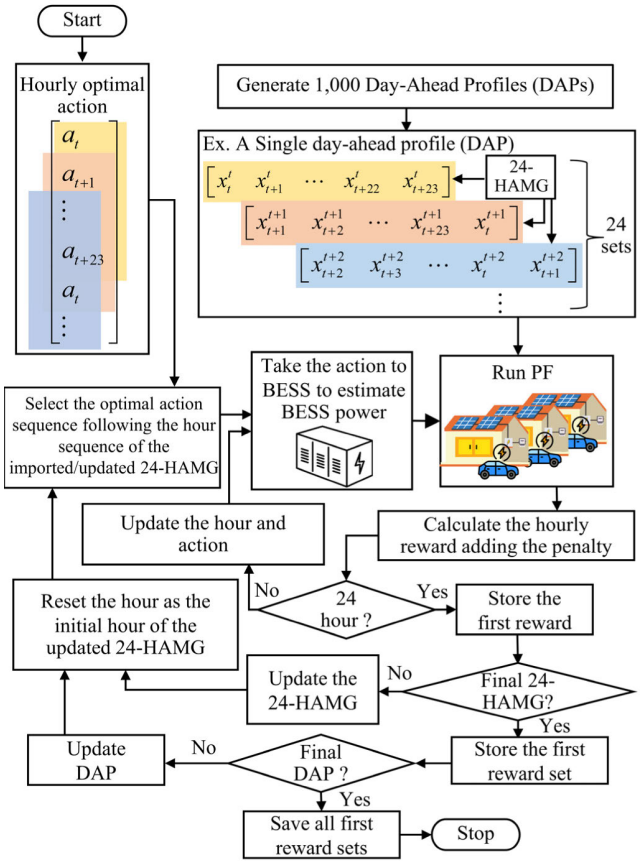

FIGURE 6. Hourly optimal action verification of conventional method.

C. TRAINING AND TESTING PROCESS OF THE PROPOSED METHOD

In the context of the proposed method, a multi-agent DDPG with random sampling is proposed. Through the random sampling process, uncertain scenarios are generated during the agent's training, causing the MG behavior to vary from hour to hour and across training iterations. This process effectively covers all uncertain scenarios, obviating the need for a forecasting model. Given the high uncertainty of MG behavior, a single agent may struggle to remember all uncertain scenarios. Therefore, a multi-agent concept is employed to address this challenge. Consequently, a multi-agent DDPG with random sampling can effectively control the BESS under all uncertain scenarios. Despite the 24-HAMG changing every hour, the proposed method can optimize BESS control to achieve the best solution. With this assumption, the well-trained multi-agent provided by the proposed method can be applied to the PEM task. The training and testing processes of the proposed method are represented in the following sub-subsections.

1) TRAINING PROCESS OF PROPOSED METHOD

The training process can be depicted in Fig. 7 and explained in the following steps:

Step 1: Construct 24 agents and set the DDPG parameters of each agent similar to setting the agent in the conventional method.

Step 2: All agents are labeled following the hours in a day, resulting in a total of 24 agents. Next, the agent labeled with the initial hour is selected for training first. Since each agent is responsible for controlling the BESS over a day (24 hours) to minimize the daily total cost, the training for this agent begins at the initial hour and continues until reaching the 24th hour.

Step 3: The selected agent will predict the action to control BESS. Then, the BESS model will receive it to estimate the BESS power.

Step 4: For the PF process, solar radiation, ambient temperature, and appliance load are sampled from their PDFs based on the current hour of the selected agent's training. This is to determine solar PV power, and appliance load at the current hour. Moreover, the hourly EV charging demand is generated by using the process in Fig. 3. Then, the hourly EV charging demand is only imported the demand at the current hour to the MG. From the above process, it can determine the hourly power of related elements in the MG to prepare the PF running.

Step 5: When the PF running is finished, the desired power flow parameters are estimated. Then, the reward adding the penalty is calculated.

Step 6: The number of the hour for the selected agent's training will be checked. If it does not reach 24, related parameters are mapped into the DDPG parameters. Subsequently, the weights of all networks of the agent will be updated along with the hour to prepare the state for the next action prediction of the agent. Otherwise, the iteration (iter.) for training will be regulated. If it does not reach the maximum iteration (Max. iter.), the iteration will be updated and the current hour will be reset to the initial hour to train the agent again. Otherwise, the agent's weights are saved for use in the testing process.

Step 7: The number of agents will be checked. If it does not equal its maximum number, the initial hour for training is updated to the labeled number of the updated agent. Otherwise, the training process will be stopped.

From the above process, the result is 24 well-trained agents. Thus, these well-trained agents exhibit good adaptability in dealing with uncertain situations because they are directly learned from occurring uncertainties in the MG, leading to obtaining the best actions for BESS control. The verification process for these agents is shown in the testing phase.

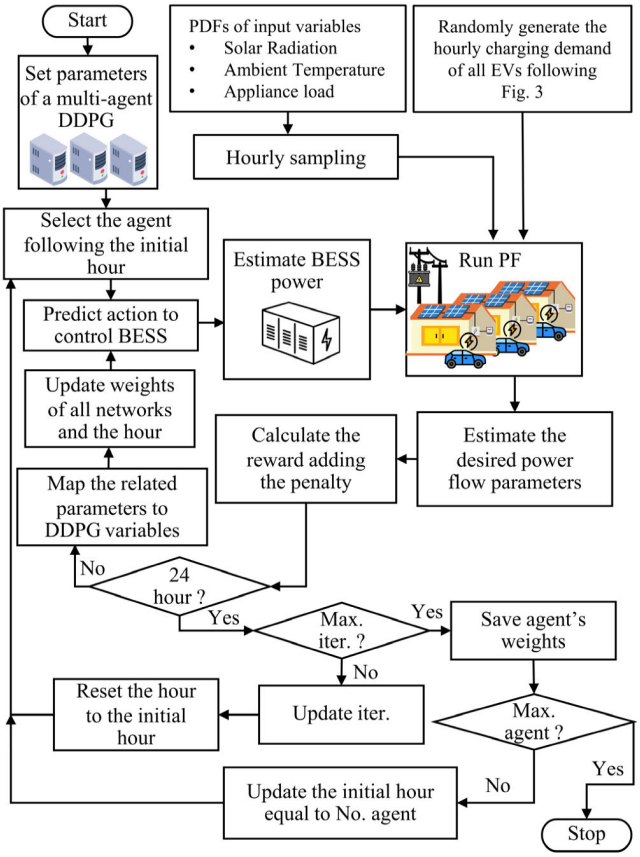

FIGURE 7. Training process of proposed method.

2) TESTING PROCESS OF PROPOSED METHOD

To verify the performance of 24 well-trained agents obtained by the training process, the testing process is designed to meet this aim. The process can be shown in Fig. 8 and explained in the following steps:

Step 1: Generate 1,000 DPAs to test the performance of all well-trained agents similar to the DAP generation of the conventional method.

Step 2: Load 24 well-trained agents to prepare for BESS controlling.

Step 3: Select the labeled agent following the initial hour of the imported 24-HAMG in the current DAP.

Step 4: The selected agent predicts the action to estimate the BESS power. Then, the BESS is injected/absorbed the power into/from the MG according to the BESS power value.

Step 5: The PF operation begins when the data sampled within the current hour from the 24-HAMG is imported into the MG. This imported data is then utilized to estimate the power consumption of all relevant elements within the MG system.

Step 6: The hourly reward adding the penalty is calculated. Then, the number of hours will be regulated. If it does not reach 24, related parameters will be mapped into the DDPG variables and the hour will be updated. Therefore, the agent will receive the state at the updated hour to predict the next action. Otherwise, the reward at the initial hour of the 24-HAMG, referred to as the first reward, is stored.

Step 7: The number of 24-HAMG will be checked. If it does not equal to 24, the 24-HAMG will be updated as the next 24-HAMG. Then, the hour will be reset to the initial hour of the next 24-HAMG, resulting in the next agent selection. Otherwise, all first rewards saved in the sixth step are stored as a set, called the first reward set.

Step 8: The number of DAPs will be regulated. If it does not equal to 1,000, the DAP will be updated to the next DAP. Subsequently, the hour will be reset as the initial of the first 24-HAMG of the updated DAP. Otherwise, the first reward sets stored in the seventh step are saved to utilize as the performance index and then stop.

VI. SIMULATION RESULTS

A. ASSUMPTION AND CASE STUDY

In this work, a novel EMS is proposed for a PEM task in a single MG, with careful consideration of high uncertainty. To construct the MG environment and the EMS based on DDPG, they are coded using the Python language in the Spyder program. The Pandapower library is applied to calculate PPF and unbalanced PF, while the TensorFlow library is employed to construct and train all networks of the agent. Moreover, the computer's specifications include an Intel(R) Core(TM) i7-8700 CPU clocked at 3.20GHz, coupled with 16.0GB of RAM. The simulated structures of the MG environment are explained in this subsection to provide clarity. Additionally, the DDPG parameters for training and testing, as well as defined case studies, are demonstrated in this subsection.

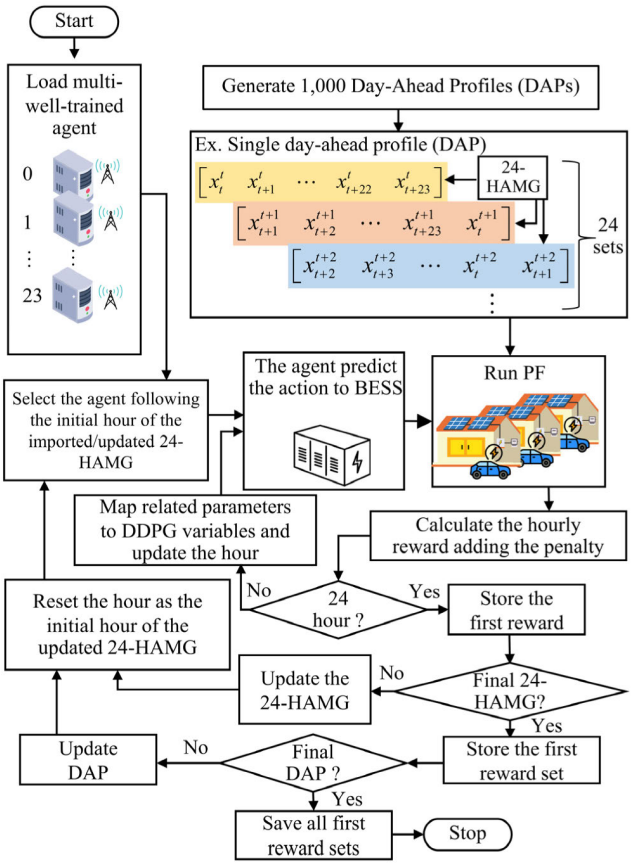

FIGURE 8. Testing process of proposed method.

1) MG ENVIRONMENT

To construct the MG environment tested, a low-voltage (230V) distribution system in Udon Thani, Thailand, of a Provincial Electricity Authority (PEA) is utilized as the MG, which is the MG2 in [13]. The MG is set as a grid-connected mode which is connected to a 22kV main grid through a distribution transformer. The transformer specification is set according to [5]. Furthermore, according to the IEEE standard, the maximum allowable HST for transformers in a distribution system is set at 110 °C [35]. The MG consists of 27 houses, each defined as a smart house. Each house features an electric appliance load, a 5 kW-rated solar PV rooftop, and a single EV with a 3.3 kW-rated charger. For the hourly evaluation of the appliance load, residential load profiles consuming less than 150 kWh from summer months over four years, spanning from March to June from 2017 to 2020 under the supervision of the PEA [40], are utilized as historical data. These profiles are used to fit hourly PDFs for estimating and sampling the hourly appliance load, allowing for the estimation of the hourly appliance load for each house. For the evaluation of the hourly solar PV generation, the hourly solar radiation and hourly ambient temperature from summer months over three years from 2015 to 2017 under the supervision of Department of Alternative Energy Development and Efficiency in Thailand [41] and the Thai Meteorological Department [42], respectively, are applied to fit the hourly PDFs. The constructed PDFs can be sampled to calculate the hourly solar PV power. The distributions of the hourly values of appliance load, solar radiation, and ambient temperature can be displayed and explained in [13].

In the context of EV vehicle specifications, the popular EV type, usage proportion of EV in each type, EV consumption rate, EV battery capacity, and efficiency for EV charging are used according to [39]. Furthermore, the parameters of PDFs related to the departure time, the arrival time, and the traveling distance are referred to [5].

The BESS is installed at the back of the transformer and controlled by the EMS. A vanadium redox flow battery, defined as a commercial BESS and a large BESS is selected as the BESS in the MG because it can be designed to meet the desired power and energy requirements easily. Additionally, it has low maintenance costs and low loss of life [43]. In this work, its specification is determined according to [14].

For the cost rate for evaluating related objective functions, the TOU rate at 22kV for buying energy [44] and the FIT rate for selling energy of the distribution system modified as the MG [45] are used to calculate the hourly exchanged energy cost. Moreover, the carbon emission rate for calculating the carbon emission cost is determined according to [13]. For

TABLE 1. DDPG parameters determination.

Networks	DDPG Parameters				
	Learning Rate	Number of Input Layer (dimension: number of states, number of action)	Number of Output Layer (dimension)	Number of Hidden Layers (neuros)	Activation Function (hidden layers, output layer)
Actor	0.005	1(8,0)	1(action)	2(400,400)	(ReLU, ReLU, linear)
Critic	0.03	1(8,1)	1(Q-value)		(ReLU, ReLU, None)

TABLE 2. Training parameters determination.

Methods	Training Parameters					
	Training Iteration	Decay Rate	Soft Update Factor	Buffer Memory Size	Batch Size	Discount Factor
Conventional	12,000	0.0001	0.05	50,000	256	0.99
Proposed	12,000 (500 per agent)	0.002 per agent				

TABLE 3. Training computational time.

Methods	Training Time (sec)
Conventional	812,162.16
Proposed	58,124.87
Decreased %	92.84 %

calculating the BESS degradation cost, the BESS capital cost is determined according to [14].

2) DDPG PARAMETERS FOR TRAINING AND TESTING

To compare the performance of the conventional and proposed methods, the DDPG parameters of both are set in a similar manner. The DDPG parameters of the agent can be shown in Table 1.

In the training parameters, with the inclusion of the multi-agent approach in the proposed method, certain training parameters are defined differently, as represented in Table 2. For example, in the conventional method with a single-agent, the training iteration is set at 12,000. In contrast, it is defined as 500 per agent in the proposed method, resulting in a total of 12,000 iterations equal to the training iteration of the conventional method. This leads to a comparison of training computational times between the proposed and conventional methods. Moreover, the decay rate for training is properly determined based on the training iteration to reduce exploration before saving the agent's weights for testing. Thus, the decay rates per agent are determined as 0.0001 in the conventional training and 0.002 in the proposed training, respectively. For other related training parameters, they are similarly set, as shown in Table 2.

3) CASE STUDY

Two case studies are constructed for comparison in this work, including:

- A single agent DDPG with PPF, defined the conventional method, is applied to control the BESS operation in the MG. By the PPF usage, the high confidence level is applied to provide the hourly-optimal solution.
- A multi-agent DDPG with random sampling, defined as the proposed method, is utilized to control the BESS operation in the MG. Without using the PPF, 24 well-trained agents are constructed to learn the MG behavior directly. These agents have a good ability to find the hourly-best solution for changed situations in the MG, but not the hourly-optimal solution.

B. TRAINING COMPUTATIONAL TIME

In the context of training computational time, it is represented in Table 3. This indicates that the conventional method spends more computational time than the proposed method. This situation results from the PPF being applied for estimating the mean and standard deviation of desired parameters, leading to a computational loop in the hourly training of the agent. In contrast, the hourly training process in the proposed method has a single PF loop in the hourly training of the agent, resulting in a lower computational burden. According to Table 3, the proposed method can decrease the training computational time by about 92% compared to the time obtained by the conventional method.

C. PEM PERFORMANCE

For the PEM performance, the actions provided by both conventional and proposed methods must properly control the BESS under changing DAPs with high uncertainties in solar PV generation, appliance load within the house, and EV charging. In the conventional method, the hourly-optimal action and the hourly-optimal SoC of the BESS obtained by a single well-trained agent using the process in Fig. 5 can be represented in Fig. 9 and Fig. 10, respectively.

TABLE 4. Summation of MRS provided by each method.

Initial hour of 24-HAMG	Summation of MRS ($)		Increased %
	Conventional	Proposed	
7	-53.907	-52.488	2.63
8	-53.681	-52.733	1.77
9	-54.073	-51.407	4.93
10	-54.023	-49.074	9.16
11	-54.197	-50.514	6.80
12	-54.446	-49.377	9.31
13	-54.062	-47.156	12.77
14	-54.104	-45.479	15.94
15	-53.857	-46.891	12.93
16	-53.680	-44.366	17.35
17	-53.607	-44.623	16.76
18	-53.452	-45.384	15.09
19	-53.814	-42.965	20.16
20	-53.685	-38.454	28.37
21	-53.747	-40.473	24.70
22	-54.018	-47.900	11.33
23	-52.080	-45.743	12.17
0	-51.616	-44.806	13.19
1	-51.258	-45.129	11.96
2	-51.083	-44.561	12.77
3	-51.880	-50.465	2.73
4	-50.305	-47.073	6.43
5	-51.019	-50.251	1.50
6	-50.812	-47.329	6.85

By applying the PPF in the testing process to provide the hourly-optimal action, the hourly mean of the desired power flow and BESS parameters mapped to the agent's hourly state variable have the same value in every testing iteration due to the hourly mean estimation of the PPF. This causes the actor of a single well-trained agent to predict the hourly-optimal action with the same values in every iteration. Consequently, the hourly-optimal action maintains the same values in every testing iteration when applying the PPF, as shown in Fig. 9.

From Fig. 9, The BESS is mostly charged from 7:00 to 21:00, as observed during this period with negative action values, leading to an increasing SoC in this period as represented in Fig. 10. Conversely, the BESS is discharged from 22:00 to 6:00. The reason for initiating discharging at 22:00 is that EV users will begin charging their batteries at this time, as 22:00 marks the start of the off-peak rate period of the TOU. Consequently, the BESS needs to be discharged to alleviate the burden on the transformer loading. The BESS discharge operations continue until reaching 6:00, resulting in a decreasing SoC during this period, which does not exceed the SoC limitation, as shown in red lines of Fig. 10.

Furthermore, the hourly-optimal action obtained by the conventional method involves long-period charging of the BESS. This results from a high confidence level consideration when applying the PPF. After 22:00, preventing undesired peak loads and considering a high confidence level are implemented due to heavy EV charging. Thus, the BESS needs to store energy for an extended period to use in heavy discharging after 22:00 due to EV charging, with a 95% confidence

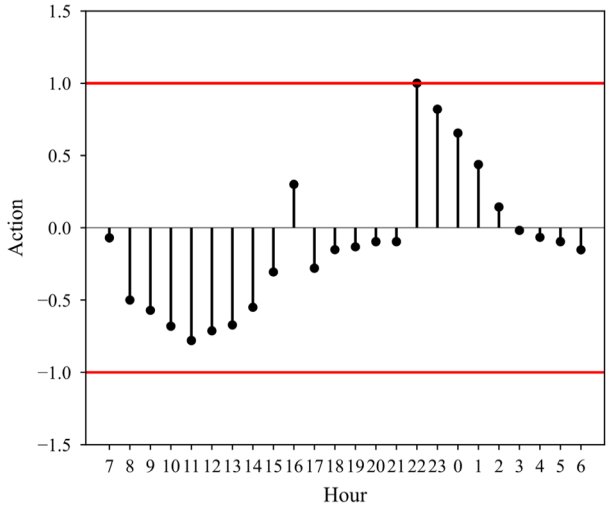

FIGURE 9. Hourly-optimal action of conventional method.

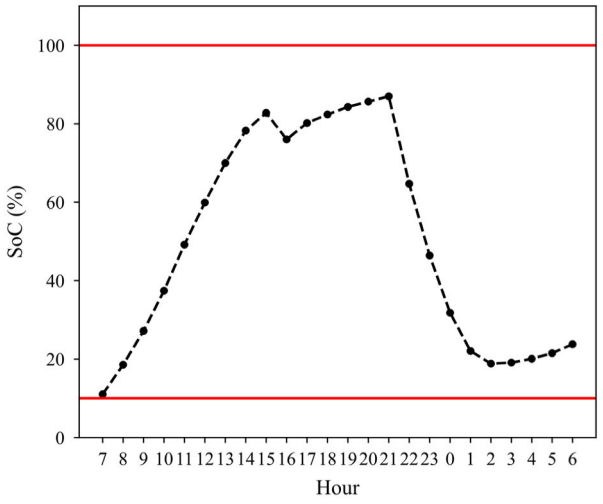

FIGURE 10. Hourly-optimal SoC of BESS of conventional method.

level consideration. Consequently, this leads to long-period charging for the BESS. While these solutions can prevent power system violations with a 95% confidence level, they result in higher costs, particularly from the drawn energy cost from the main grid to charge the BESS and the BESS degradation cost from heavy discharging.

From the training results of the proposed method, 24 well-trained agents are trained with random sampling instead of considering a high confidence level using PPF. Consequently, all hourly-state variables of the agent are changed in every testing iteration. The hourly action in each testing scenario does not have the same value. Therefore, the hourly-best action provided by the proposed method cannot be presented as a figure, but it can be shown through the reward performance, which is explained in the next result.

The reward sequence of each 24-HAMG estimated using 1,000 DPAs, derived from the testing process and defined

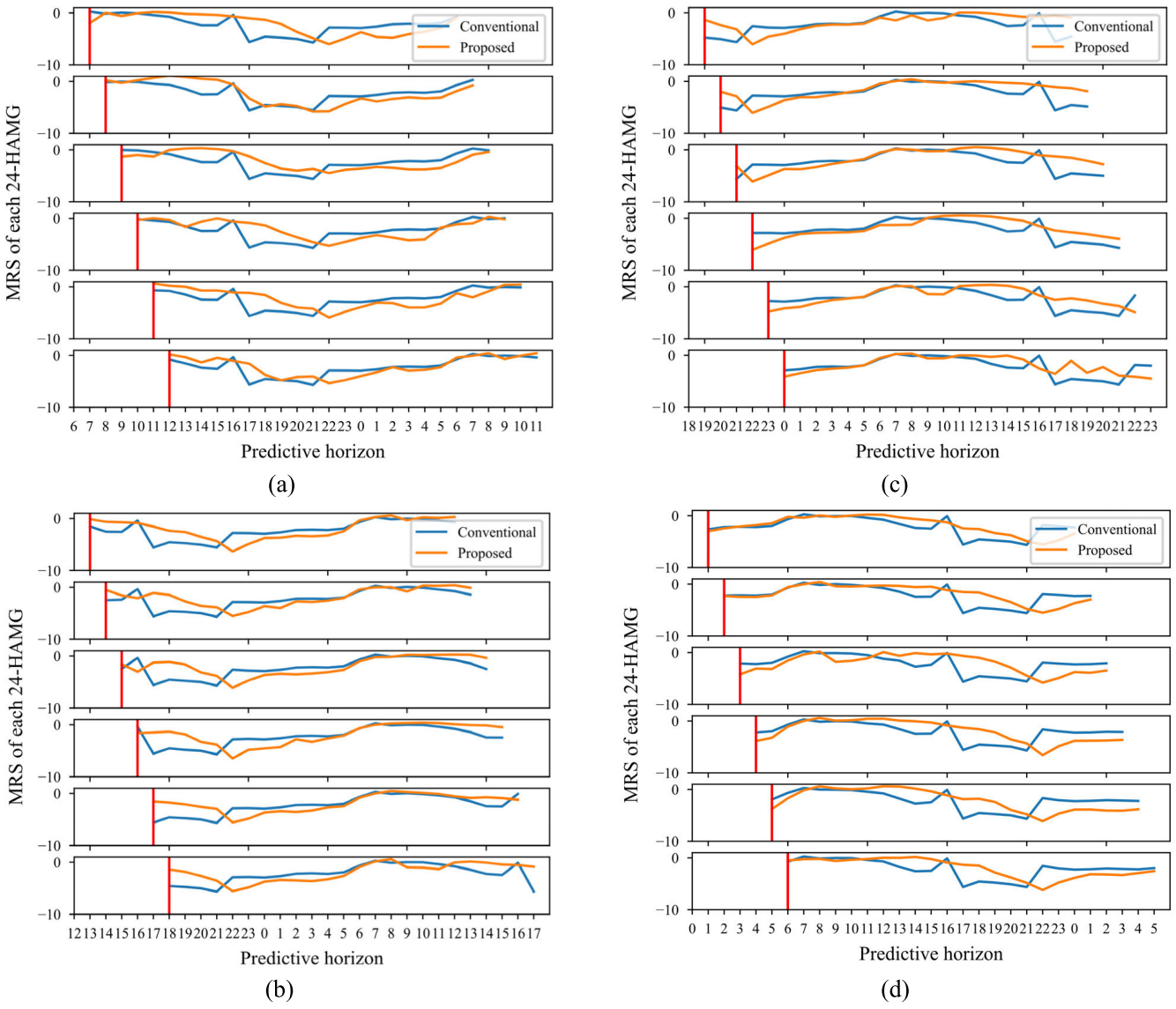

FIGURE 11. The MRS of each 24-HAMG estimated using 1000 DPAs; (a) initial hour from 7:00 to 12.00, (b) initial hour from 13:00 to 18.00, (c) initial hour from 19:00 to 0.00 and (d) initial hour from 1.00 to 6.00.

in Section V Subsections B and C, is served as the performance index in the PEM task. The hourly-optimal action obtained by the conventional method is used to provide the reward sequence of each 24-HAMG, and then it is stored to average as the MRS for each 24-HAMG, as shown in the blue line in Fig. 11. In contract, the orange line in Fig. 11 is represented as the MRS of each 24-HAMG when applying 24 well-trained agents constructed by the proposed method. Thus, all MRS can be represented in Fig. 11(a) to Fig. 11(d). From Fig. 11, the hourly-optimal action obtained by the conventional method and the 24 well-trained agents provided by the proposed method can properly control the BESS, as noticed from the MRS not having a large negative value. This indicates that the penalty added to the reward equals 0, which serves as evidence reflecting that the related constraints are not violated. Moreover, the MRS of the proposed method tends to have a higher value than that of the conventional method during the period from 7:00 to 21:00. In contrast, during the period from 22:00 to 6:00, the MRS of the conventional method tends to have higher values. To clarify, the summation of each MRS can be presented in Table 4. The table indicates that the summation of each MRS provided by the proposed method is higher than that obtained by the conventional method by 1.50% to 28.37%. Therefore, the proposed method demonstrates superior performance in the context of the day-ahead PEM task compared to the conventional method.

D. EXCHANGED ENERGY DISTRIBUTION

From the testing process, 24 sets of 24-HAMG in each DAP are applied to test the performance of optimization methods. The desired parameters at the initial hour of each

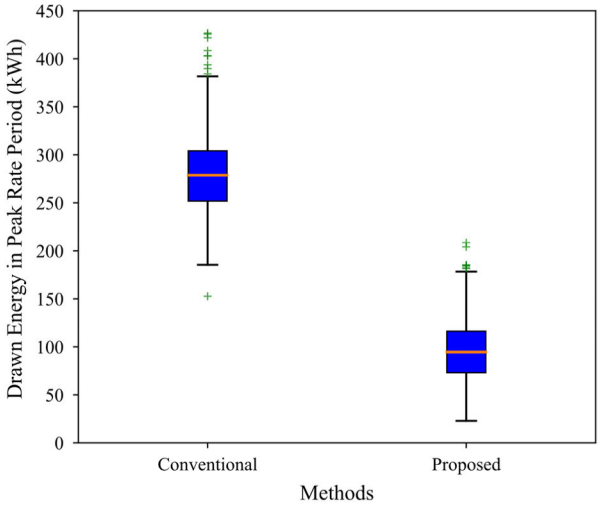

FIGURE 12. Drawn energy of MG in peak rate period of TOU.

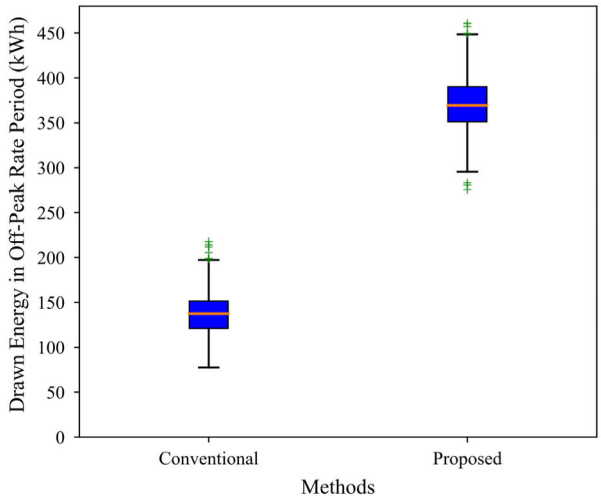

FIGURE 13. Drawn energy of MG in off-peak rate period of TOU.

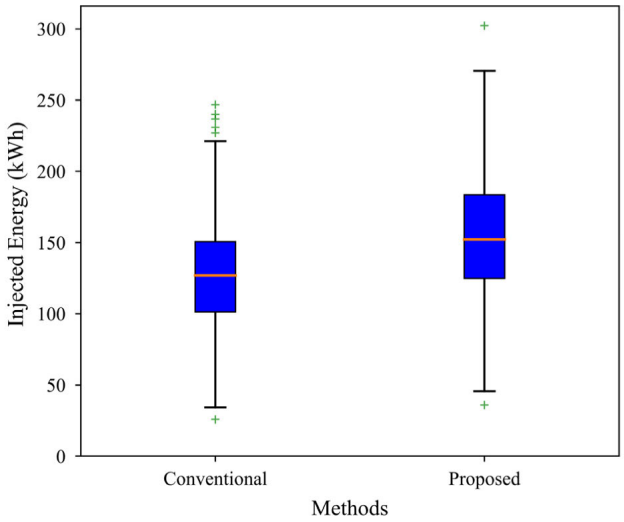

FIGURE 14. Injected energy of MG.

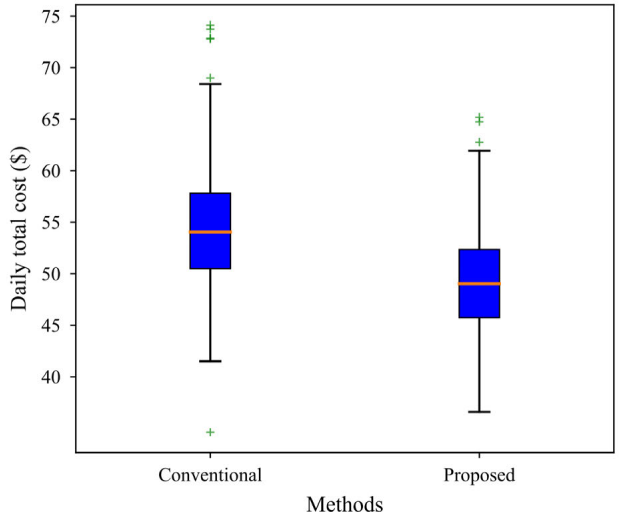

FIGURE 15. Daily total cost distribution.

24-HAMG are saved before the 24-HAMG is updated as the next 24-HAMG. For the PEM task in this work, the 24-HAMG is changed energy hour in a day, defined as the worst case of the PEM. To this end, the action at the initial hour of the 24-HAMG is the real action that is taken into the BESS to control before the 24-HAMG is updated. Therefore, the MG parameters at the initial hour of the 24-HAMG are the MG parameters that really occurred, not simulated. These parameters are saved to define as the performance indexes. By having 24 sets of 24-HAMG per DAP, the exchanged power is saved as 24 values per DAP, forming the exchanged power sequence. This results in having 1,000 exchanged power sequences according to the number of DAPs tested.

For 1,000 exchanged power sequences, each sequence is split into the drawn power sequence during both the peak rate period and the off-peak rate period of TOU, which are defined as the MG's power absorption from the main grid. Additionally, each sequence is split into the injected power sequence, which the MG exports power to the main grid. Consequently, there will be 1,000 sets of drawn power sequences, along with 1,000 sets of injected power sequences. When the summations of all drawn power sequences during both the peak rate period and the off-peak rate period of TOU are calculated, respectively, they yield the drawn energy for both the peak rate period and the off-peak rate period of TOU. This process results in having 1,000 values for both drawn energy in peak and off-peak periods. Moreover, the injected power sequence will be subjected to the same process. The distributions of these energies can be represented in Fig. 12, Fig. 13, and Fig. 14, respectively.

The drawn energy from the main grid, as shown in Fig. 12, exhibits a clearer tendency to be lower during peak rate periods when the proposed method is applied compared to when the conventional method is employed. Additionally, in the context of energy drawn during off-peak rate periods

TABLE 5. The mean and standard deviation of daily total cost.

Methods	Mean ($)	Standard Deviation
Conventional	54.245	5.376
Proposed	49.244	4.719
Decreased %	9.22 %	-

as according to Fig. 13, there is a tendency for it to be higher when implementing the proposed method compared to when using the conventional method. Therefore, the proposed method demonstrates greater intelligence in importing energy from the main grid than the conventional method. Consequently, the exchanged energy cost tends to be lower when the proposed method is implemented, thus guaranteeing its performance.

Moreover, in the context of the injected energy into the main grid, the proposed method can control the BESS to provide a tendency for the injected energy to be higher compared to the conventional method, as depicted in Fig. 14. Therefore, the revenue of the MG from injecting energy to the main grid has a higher tendency when the proposed method is applied.

E. DAILY TOTAL COST DISTRIBUTION

In the previous subsection, the reward at the initial hour of each 24-HAMG is saved similarly to the drawn/injected power, called the first reward. Thus, there are 1,000 sets of the first reward resulting from the actual BESS control. In Section VI Subsection C, the reward is verified without the penalty value. Therefore, each first reward can be shown as the hourly total cost. To this end, there are 1,000 sets of the hourly total cost. Then, when each hourly total cost is summed, it is defined as a daily total cost, which is determined as one of the performance indexes. Thus, there are 1,000 values of the daily total cost. The distribution of the daily total cost can be represented in Fig. 15.

From Fig. 15, the daily total cost has a lower tendency when the proposed method is implemented compared to the conventional method. This indicates that the proposed method can manage energy within the MG with a better performance than the conventional method. To clarify in the context of a figure, the mean and standard deviation of the daily total cost obtained by the conventional and proposed methods can be represented in Table 5.

In Table 5, the standard deviation of daily total costs obtained by the proposed method is lower than that provided by the conventional method. This indicates that the proposed method can provide a distribution of the daily total cost with less disorganization compared to the conventional method. Moreover, the mean of the daily total cost obtained by the proposed method is 9.22% lower than that obtained by the conventional method. Therefore, the proposed method demonstrates good performance in BESS controlling with lower related costs.

VII. DISCUSSIONS

To verify the performance of the conventional and proposed methods, the training and testing processes are operated to achieve this goal. When applying the PPF in the conventional method, the training process incurs a high computational burden due to the PPF loop for estimating the mean and standard deviation of related parameters. In contrast, the training process of the proposed method does not involve a PPF loop, thereby effectively mitigating this issue.

By unmitigated uncertainties of MG parameters through random sampling processes instead of relying on PPF, the DDPG agent can directly learn from the MG environment and provide the best solution in each scenario. Furthermore, transitioning from a single-agent DDPG to a multi-agent DDPG framework allows the multi-agent to allocate roles for predicting 24-hour-ahead actions in BESS control based on changing 24-HAMGs for every hour. As a result, during the testing process, the proposed method demonstrates better performance in BESS control, resulting in a higher summation of the MRS of each 24-HAMG, along with enhanced intelligence in importing/exporting energy to/from the main grid compared to that obtained by a single-agent DDPG with PPF. Additionally, the daily total costs achieved by the proposed method exhibit a lower mean value and narrower distribution range compared to those obtained by the conventional method. This confirms that the proposed method leads to lower cost attainment and higher accuracy in cost prediction for the MGO.

VIII. CONCLUSION

A novel EMS utilizing a multi-agent DDPG with random sampling is proposed for the PEM task within a single low-voltage MG installing a single BESS and considering uncertainties of solar PV generation, appliance load, and EV charging demand. Each agent is assigned the task of predicting the best action for BESS control 24 hours ahead, based on the values of random variables such as solar PV generation, appliance load, and EV charging demand within the same timeframe. Additionally, the 24-hour-ahead values of random variables change every hour throughout the day, generating from random sampling process and representing the worst-case scenario for the PEM task. Consequently, there are 24 agents (corresponding to the number of hours in a day) tasked with predicting 24-hour-ahead action for BESS control based on the updated 24-hour-ahead MG behavior values throughout the day. Through this concept, the PEM task can be applied in combination with multi-agent DDPG. There are three objective functions minimized in this work, including exchanged energy cost, BESS degradation cost, and carbon emission cost. Moreover, MG parameters, such as bus voltage, line current, hottest-spot temperature of the transformer, and power and SoC of the BESS are considered as the constraints in the PEM task.

Simulation results demonstrate that unmitigated uncertainties of random variables using random sampling processes

instead of utilizing PPF, and transitioning from a single-agent to a multi-agent framework, facilitate direct learning of MG uncertainty for the DDPG agent when performing the PEM task. Consequently, the proposed method leads to enhanced performance of the EMS with reduced training time, an increased trend of rewards, a decreased trend of daily costs, and improved intelligence in energy import/export operations from/to the main grid. In future work, the proposed method may be implemented in the PEM task of multi-MG. This can transition a conventional power system into a smart grid.

REFERENCES

[1] U.S. Energy Inf. Admin. (2019). *Electricity Explained*. Accessed: Apr. 04, 2023. [Online]. Available: https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php

[2] U.S. Environ. Protection Agency. (2022). *Sources of Greenhouse Gas Emissions*. Accessed: Apr. 04, 2023. [Online]. Available: https://www.epa.gov/ghgemissions/sources-greenhouse-gas-emissions

[3] L. Kumar and S. Jain, "Electric propulsion system for electric vehicular technology: A review," *Renew. Sustain. Energy Rev.*, vol. 29, pp. 924–940, Jan. 2014, doi: 10.1016/J.RSER.2013.09.014.

[4] A. Hirsch, Y. Parag, and J. Guerrero, "Microgrids: A review of technologies, key drivers, and outstanding issues," *Renew. Sustain. Energy Rev.*, vol. 90, pp. 402–411, Jul. 2018, doi: 10.1016/J.RSER.2018.03.040.

[5] C. Srithapon, P. Ghosh, A. Siritaratiwat, and R. Chatthaworn, "Optimization of electric vehicle charging scheduling in urban village networks considering energy arbitrage and distribution cost," *Energies*, vol. 13, no. 2, p. 349, Jan. 2020, doi: 10.3390/EN13020349.

[6] P. Xie, J. M. Guerrero, S. Tan, N. Bazmohammadi, J. C. Vasquez, M. Mehrzadi, and Y. Al-Turki, "Optimization-based power and energy management system in shipboard microgrid: A review," *IEEE Syst. J.*, vol. 16, no. 1, pp. 578–590, Mar. 2022, doi: 10.1109/JSYST.2020.3047673.

[7] S. Chakraborty, G. Modi, and B. Singh, "A cost optimized-reliable-resilient-realtime-rule based energy management scheme for a SPV-bes based microgrid for smart building applications," *IEEE Trans. Smart Grid*, vol. 14, no. 4, pp. 2572–2581, Jul. 2022, doi: 10.1109/TSG.2022.3232283.

[8] D. G. Kyriakou, F. D. Kanellos, and D. Ipsakis, "Multi-agent-based real-time operation of microgrids employing plug-in electric vehicles and building prosumers," *Sustain. Energy, Grids Netw.*, vol. 37, Mar. 2024, Art. no. 101229, doi: 10.1016/J.SEGAN.2023.101229.

[9] V. S. B. Kurukuru, A. Haque, S. Padmanaban, and M. A. Khan, "Rule-based inferential system for microgrid energy management system," *IEEE Syst. J.*, vol. 16, no. 1, pp. 1582–1591, Mar. 2022, doi: 10.1109/JSYST.2021.3094403.

[10] M. Jafari, Z. Malekjamshidi, D. D. Lu, and J. Zhu, "Development of a fuzzy-logic-based energy management system for a multiport multioperation mode residential smart microgrid," *IEEE Trans. Power Electron.*, vol. 34, no. 4, pp. 3283–3301, Apr. 2019, doi: 10.1109/TPEL.2018.2850852.

[11] T. T. Teo, T. Logenthiran, W. L. Woo, K. Abidi, T. John, N. S. Wade, D. M. Greenwood, C. Patsios, and P. C. Taylor, "Optimization of fuzzy energy-management system for grid-connected microgrid using NSGA-II," *IEEE Trans. Cybern.*, vol. 51, no. 11, pp. 5375–5386, Nov. 2021, doi: 10.1109/TCYB.2020.3031109.

[12] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning," *Int. J. Electr. Power Energy Syst.*, vol. 131, Oct. 2021, Art. no. 107048, doi: 10.1016/J.IJEPES.2021.107048.

[13] N. Kaewdornhan, C. Srithapon, and R. Chatthaworn, "Electric distribution network with multi-microgrids management using surrogate-assisted deep reinforcement learning optimization," *IEEE Access*, vol. 10, pp. 130373–130396, 2022, doi: 10.1109/ACCESS.2022.3229127.

[14] N. Kaewdornhan and R. Chatthaworn, "Model-free data-driven approach assisted deep reinforcement learning for optimal energy management in MicroGrid," *Energy Rep.*, vol. 9, pp. 850–858, Oct. 2023, doi: 10.1016/J.EGYR.2023.05.130.

[15] V. Herrera, A. Milo, H. Gaztañaga, I. Etxeberria-Otadui, I. Villarreal, and H. Camblong, "Adaptive energy management strategy and optimal sizing applied on a battery-supercapacitor based tramway," *Appl. Energy*, vol. 169, pp. 831–845, May 2016, doi: 10.1016/J.APENERGY.2016.02.079.

[16] T. Lv and Q. Ai, "Interactive energy management of networked microgrids-based active distribution system considering large-scale integration of renewable energy resources," *Appl. Energy*, vol. 163, pp. 408–422, Feb. 2016, doi: 10.1016/J.APENERGY.2015.10.179.

[17] N. Nikmehr, S. Najafi-Ravadanegh, and A. Khodaei, "Probabilistic optimal scheduling of networked microgrids considering time-based demand response programs under uncertainty," *Appl. Energy*, vol. 198, pp. 267–279, Jul. 2017, doi: 10.1016/J.APENERGY.2017.04.071.

[18] K. Borisoot, R. Liemthong, C. Srithapon, and R. Chatthaworn, "Optimal energy management for virtual power plant considering operation and degradation costs of energy storage system and generators," *Energies*, vol. 16, no. 6, p. 2862, Mar. 2023, doi: 10.3390/EN16062862.

[19] S. Hou, G. Desta Gebreyesus, and S. Fujimura, "Day-ahead multi-modal demand side management in microgrid via two-stage improved ring-topology particle swarm optimization," *Expert Syst. Appl.*, vol. 238, Mar. 2024, Art. no. 122135, doi: 10.1016/J.ESWA.2023.122135.

[20] N. Javaid, G. Hafeez, S. Iqbal, N. Alrajeh, M. S. Alabed, and M. Guizani, "Energy efficient integration of renewable energy sources in the smart grid for demand side management," *IEEE Access*, vol. 6, pp. 77077–77096, 2018, doi: 10.1109/ACCESS.2018.2866461.

[21] Z. A. Khan, A. A. Butt, T. A. Alghamdi, A. Fatima, M. Akbar, M. Ramzan, and N. Javaid, "Energy management in smart sectors using fog based environment and meta-heuristic algorithms," *IEEE Access*, vol. 7, pp. 157254–157267, 2019, doi: 10.1109/ACCESS.2019.2949863.

[22] S. K. Mishra, D. Puthal, J. J. P. C. Rodrigues, B. Sahoo, and E. Dutkiewicz, "Sustainable service allocation using a metaheuristic technique in a fog server for industrial applications," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4497–4506, Oct. 2018, doi: 10.1109/TII.2018.2791619.

[23] H. H. Goh, Y. Huang, C. S. Lim, D. Zhang, H. Liu, W. Dai, T. A. Kurniawan, and S. Rahman, "An assessment of multistage reward function design for deep reinforcement learning-based microgrid energy management," *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4300–4311, Nov. 2022, doi: 10.1109/TSG.2022.3179567.

[24] D. J. B. Harrold, J. Cao, and Z. Fan, "Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning," *Energy*, vol. 238, Jan. 2022, Art. no. 121958, doi: 10.1016/J.ENERGY.2021.121958.

[25] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020, doi: 10.1109/JIOT.2019.2957289.

[26] W. Kolodziejczyk, I. Zoltowska, and P. Cichosz, "Real-time energy purchase optimization for a storage-integrated photovoltaic system by deep reinforcement learning," *Control Eng. Pract.*, vol. 106, Jan. 2021, Art. no. 104598.

[27] E. Foruzan, L.-K. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5749–5758, Sep. 2018, doi: 10.1109/TPWRS.2018.2823641.

[28] N. Kaewdornhan, C. Srithapon, R. Liemthong, and R. Chatthaworn, "Real-time multi-home energy management with EV charging scheduling using multi-agent deep reinforcement learning optimization," *Energies*, vol. 16, no. 5, p. 2357, Mar. 2023, doi: 10.3390/EN16052357.

[29] A. Salari, S. E. Ahmadi, M. Marzband, and M. Zeinali, "Fuzzy Q-learning-based approach for real-time energy management of home microgrids using cooperative multi-agent system," *Sustain. Cities Soc.*, vol. 95, Aug. 2023, Art. no. 104528, doi: 10.1016/J.SCS.2023.104528.

[30] A. Salari, M. Zeinali, and M. Marzband, "Model-free reinforcement learning-based energy management for plug-in electric vehicles in a cooperative multi-agent home microgrid with consideration of travel behavior," *Energy*, vol. 288, Feb. 2024, Art. no. 129725, doi: 10.1016/J.ENERGY.2023.129725.

[31] F. Monfaredi, H. Shayeghi, and P. Siano, "Multi-agent deep reinforcement learning-based optimal energy management for grid-connected multiple energy carrier microgrids," *Int. J. Electr. Power Energy Syst.*, vol. 153, Nov. 2023, Art. no. 109292, doi: 10.1016/J.IJEPES.2023.109292.

[32] M. Huang, X. Lin, Z. Feng, D. Wu, and Z. Shi, "A multi-agent decision approach for optimal energy allocation in microgrid system," *Electric Power Syst. Res.*, vol. 221, Aug. 2023, Art. no. 109399, doi: 10.1016/J.EPSR.2023.109399.
[33] J.-O. Lee and Y.-S. Kim, "Novel battery degradation cost formulation for optimal scheduling of battery energy storage systems," *Int. J. Electr. Power Energy Syst.*, vol. 137, May 2022, Art. no. 107795, doi: 10.1016/j.ijepes.2021.107795.
[34] Provincial Electr. Authority (PEA). (2020). *Provincial Electricity Authority Power Quality Standards B.E.2563 (2020)*. [Online]. Available: https://www.pea.co.th/Portals/0/Document/WorkStandard/standardmix_2564_1.pdf
[35] *IEEE Guide for Loading Mineral-Oil-Immersed Transformers and Step-Voltage Regulators—Redline*, document C57.91-2011, 2012.
[36] H. R. Baghaee, M. Mirsalim, G. B. Gharehpetian, and H. A. Talebi, "Application of RBF neural networks and unscented transformation in probabilistic power-flow of microgrids including correlated wind/PV units and plug-in hybrid electric vehicles," *Simul. Model. Pract. Theory*, vol. 72, pp. 51–68, Mar. 2017, doi: 10.1016/J.SIMPAT.2016.12.006.
[37] B. R. Prusty and D. Jena, "Combined cumulant and Gaussian mixture approximation for correlated probabilistic load flow studies: A new approach," *CSEE J. Power Energy Syst.*, vol. 2, no. 2, pp. 71–78, Jun. 2016, doi: 10.17775/CSEEJPES.2016.00024.
[38] *National Household Travel Survey (NHTS)*, USA Dept. Transp., Washington, DC, USA, 2009.
[39] C. Srithapon, P. Fuangfoo, P. K. Ghosh, A. Siritaratiwat, and R. Chatthaworn, "Surrogate-assisted multi-objective probabilistic optimal power flow for distribution network with photovoltaic generation and electric vehicles," *IEEE Access*, vol. 9, pp. 34395–34414, 2021, doi: 10.1109/ACCESS.2021.3061471.
[40] Provincial Electr. Authority (PEA). (2023). *Load Research of PEA*. Accessed: Jun. 08, 2023. [Online]. Available: http://peaoc.pea.co.th/loadprofile/en/
[41] Dept. Alternative Energy Develop. Efficiency. (2023). *Hourly Solar Irradiation for Khon Kaen, Thailand*. Accessed: Jul. 10, 2023. [Online]. Available: https://weben.dede.go.th/webmax/
[42] Thai Meteorological Dept. (2023). *Hourly Temperature for Khon Kaen*. Accessed: Jul. 02, 2023. [Online]. Available: https://www.tmd.go.th/en/
[43] K. Lourenssen, J. Williams, F. Ahmadpour, R. Clemmer, and S. Tasnim, "Vanadium redox flow batteries: A comprehensive review," *J. Energy Storage*, vol. 25, Oct. 2019, Art. no. 100844, doi: 10.1016/j.est.2019.100844.
[44] Provincial Electr. Authority (PEA). (2018). *Electricity Tariffs for User*. Accessed: May 03, 2023. [Online]. Available: https://www.pea.co.th
[45] Energy Policy Planning Office (EPPO). (2022). *Review of the Purchase of Electricity From Incremental Renewable Energy Under Natural Gas Management*. Accessed: May 25, 2023. [Online]. Available: https://www.eppo.go.th/index.php/en/

NIPHON KAEWDORNHAN received the B.Eng. (Hons.) and M.Eng. degrees in electrical engineering from Khon Kaen University, Khon Kaen, Thailand, in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree with the Department of Electrical Engineering. His research interests include power system analysis, microgrids, energy management systems, renewable energy resources, and machine learning for energy management optimization and prediction.

RONGRIT CHATTHAWORN received the B.Eng. degree (Hons.) in electrical engineering from Khon Kaen University, Thailand, in 2009, and the M.Eng. degree in electrical engineering and the Ph.D. degree from Chulalongkorn University, Thailand, in 2011 and 2015, respectively. From 2015 to 2017, he was a Researcher with Energy Regulatory Commission (ERC), Thailand. Since 2022, he has been an Associate Professor with the Department of Electrical Engineering, Khon Kaen University. His research interests include power system planning, power system reliability, smart grids, energy management systems, and renewable energy resources.