Combined Heat and Power System Intelligent Economic Dispatch: A Deep Reinforcement Learning Approach

Suyang Zhou [1], Zijian Hu [1], Wei Gu [1*], Meng jiang [2], Qiteng Hong [3], Campbell Booth [3]

[1] School of Electrical Engineering, Southeast University, 2 Sipailou Xuanwu Qu, Nanjing, China
[2] Department of Computer Science and Engineering, University of Notre Dame
[3] Department of Electronic and Electrical Engineering, University of Strathclyde
[*] wgu@seu.edu.cn

This paper proposed a Deep Reinforcement learning (DRL) approach for Combined Heat and Power (CHP) system economic dispatch which is suitable for different operating scenarios and can significantly decrease the computational complexity without affecting accuracy. In the respect of problem description, a vast of Combined Heat and Power (CHP) economic dispatch problems are modeled as a high-dimensional and non-smooth objective function with a large number of non-linear constraints for which powerful optimization algorithms and considerable time are required to solve it. In order to reduce the solution time, most engineering applications choose to linearize the optimization target and devices model. To avoid complicated linearization process, this paper models CHP economic dispatch problems as Markov Decision Process (MDP) that making the model highly encapsulated to preserve the input and output characteristics of various devices. Furthermore, we improve an advanced deep reinforcement learning algorithm: distributed proximal policy optimization (DPPO), to make it applicable to CHP economic dispatch problem. Based on this algorithm, the agent will be trained to explore optimal dispatch strategies for different operation scenarios and respond to system emergencies efficiently. In the utility phase, the trained agent will generate optimal control strategy in real time based on current system state. Compared with existing optimization methods, the advantages of the proposed DRL methods are mainly reflected in the following three aspects: 1) Adaptability: under the premise of the same network topology, the trained agent can handle the economic scheduling problem in various operating scenarios without recalculation. 2) High encapsulation: The user only needs to input the operating state to get the control strategy, while the optimization algorithm needs to re-write the constraints and other formulas for different situations. 3) Time scale flexibility: It can be applied to both the day-ahead optimized scheduling and the real-time control. At the same time, we give a rigorous proof that the DRL method can converge to the optimal solution. To evaluate the performance of proposed economic dispatch algorithm, comprehensive numerical analysis is conducted. The result shows that the training time of our improved algorithm is 201 seconds and 318 seconds less than other two advanced DRL algorithm. And the difference on economic performance between this method and optimization methods is only 0.029%. If the wind power of the system is 0, the trained system can still find optimal dispatch strategy without re-training.

This work was supported in by the National Natural Science Foundation of China under Grant 51807024.

1. NOMENCLATURE

$\mathcal{S}$	State of CHP system
$\mathcal{A}$	Action for devices.
I	Indicator function
c	Equipment operating status vector
d	Power mismatch vector
r	Random variables.
R_π	Reward function
$V_\pi(s_t)$	Value function
A_π:	Advantage function.
$\eta(\tilde{\pi})$	Expected return
t	The t-th time slot.
P_{gt}	The electricity output of GT
P_{wind}	Wind power
P_{grid}	Trading electricity with grid
P_l,	Electricity load
Q_l,	Thermal load
Q_{gt}	The heat output of GT
Q_{gb}	The heat output of GB
α	Thermoelectric conversion efficiency of the GT
c_{gas}/c_{grid}	Natural gas cost /The grid interaction cost
ρ_{gas}/ρ_{grid}	unit price of natural gas/buy/sell electricity respectively
ϑ	Reasonable operating capacity
θ	Network parameter
$z_t(\theta)$	Probability ratio of updating parameter
ϵ	$Clip$ hyperparamert
γ	Discounting factor
P_{gt}^{max}	Maximum output of GT
P_{gt}^{min}	Minimum output of GT
P_{gb}^{max}	Maximum output of GB
P_{gt}^{min}	Minimum output of GT
q_{char}^{tst}	Charging rate of TST
q_{dis}^{tst}	Discharging rate of TST
$\pi_\theta(a\|s)$	Parameterized policies
r	Reward
$\mathbb{E}_{a_t,s_{t+1},\ldots}[\cdot]$	actions are sampled $a_t \sim \tilde{\pi}(\cdot\|s_t)$

2. Introduction

Co-generation units plays an increasingly important role in the latest power system for their high energy efficiency, excellent environmental-friendly performance and high flexibility. Considering the mutual conversion among various energies, there is an plenty headroom for us to optimize the current conventional CHP system, despite some widespread concerns over the way to improve the economy of the CHP operation [1][2].

The combined heat and power economic dispatch (CHPED) is a significant brunch in CHP researches, which aims at minimizing the total production cost or maximizing the operating income while keeping all constraints satisfied. CHPED problem is generally descripted as optimization problem with one or more optimizing objectives and a set of highly nonlinear and non-smooth constraints including energy supply-demand balance, capacity limits and other constraints.

The researches on CHPED mainly concentrate on two aspect: models and solutions. Several works have already been done in the CHP economic dispatch models domain. A thermal-electrolytic coupling method was proposed in [3], in which the CHP economic dispatch problem was decomposed into two heat and electricity sub-problems. Paper [4] and [5] proposed the CHP dispatch models which considered the detailed heat transfer process of the heat storage device and the cogeneration unit respectively. [6] established a two-stage dispatch model based on quantity adjustment and presented an iterative solution algorithm. An integrated response method for electro-thermal demand was proposed in [7] to improve the economy of CHP systems. An operational and structural Model based on efficiency matrices was proposed in [8][9], which was used for the dispatch of multi-energy system [10]. All economic dispatch problems are ultimately mathematically transformed into optimization problems, and the operation region of CHP systems can be modelled either convex or non-convex. Convex operation region is modelled by convex combination of electricity and heat extreme points [11][12] while non-convex operation region is usually modelled as mixed-integer model [13].

Some classical numerical methods have been successfully applied to CHPED including two-layer Lagrangian relaxation technique [3], Efficient Branch and Bound algorithm [14], dual and quadratic programming [15], etc. However, these methods have been criticized for their inability to cope with complex optimization problems which have highly nonlinear objective function and constraints. On the contrary, genetic algorithms, simulated annealing and evolutional algorithms could solve non-linear, non-smooth and non-convex optimization problem efficiently. Evolution programming-based algorithm was adopted in [16], in which the mutation search range could be controlled and the neighborhood of the best individual in a population could be searched. [17] presented multi-player harmony search algorithm for large-scale CHPED problem and obtained better convergence performance. Cuckoo optimization algorithm was powered by penalty function in [18]. This algorithm could yield better evolution and constraints handling methods. [19] improved basic genetic algorithm from avoiding excessive losses, excavating the information of parents and improving crossed offspring's quality three aspects.

These pioneering researches laid the foundation for the optimal dispatch of CHP system. However, it is worth noting that the solutions proposed by the existing research depend upon strict description of the CHP system. When the operating state changes, the strategy generated according to the original optimization problem is no longer the optimal strategy. In addition, conventional optimization methods do not achieve good encapsulation in engineering applications because the user needs to adjust the optimization target and constraint equation according to the operating state of the system.

We aim to address both of the two challenges by modeling CHP system as MDP problem and solving it by deep reinforcement learning (DRL) method. MDP is a discrete time stochastic control process and provides a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker [20]. At each time step, the process is in some CHP operating state, and the decision maker may choose any control action that is available for the current state. The process will return a corresponding reward to evaluate the quality of this question. By solving MDP, the decision maker could learn to choose optimal action for the current state to achieve maximum cumulative reward. By this model, the user only needs to consider the input of the system and the corresponding output, without having to consider the complex mathematical description of the system while retaining strict constraints.

Reinforcement learning based methods have so far attracted a number of researchers to apply them to power system optimization. [21] combined the artificial neural network and the Q-learning algorithms to achieve the optimal management of operation and maintenance of power grids. [22] applied the fuzzy reinforcement learning to energy trading process to improve the users' economy. [23] presented two variants of RL algorithms to solve economic problem and tested their performance on the IEEE 30 bus system. In this paper, a variant of Distributed Proximal Policy Optimization (DPPO) algorithm [24] for CHP economic dispatch problem has been introduced to our research. This algorithm is capable of handling different operation conditions without sacrificing stability or accuracy. When the system parameters change, the dispatch strategy can be directly given without long-term calculation by the chosen optimization methods. The Asynchronous Advantage Actor-Critic (A3C) [25] based agents and the Clipped Surrogate Objective [26] are adopted to improve the learning efficiency and stability. A comparison has been performed between the performance of this algorithm and two other common benchmark algorithms in CHP dispatch problem. Furthermore, the algorithm has been applied to day-ahead dispatch and real-time dispatch in our research, and the result has been compared with that from the mathematical optimization method. The contribution of this paper could be summarized as the following:

1) The CHP economic dispatch problem is modeled as Markov Decision Process (MDP). We have strict treatment of constraints and objective functions which ensures that the

results obtained by MDP are still the optimal solution in the feasible domain.

2) A variant DPPO algorithm for economic dispatch problem has been developed to improve algorithm exploring ability, which ensures that the result is the global optimal solution rather than a local optimal solution. The paper gives detailed proof of the most optimality and convergence of the algorithm.

3) The proposed method can be reused under a variety of operating scenarios without re-calculation, which improves the adaptability and provides more convenience for users. When the operating states change or emergency happen, users only need to input the current status to get an optimal control strategy instantly, instead of rewriting the constraint equation.

4) The proposed method has time scale flexibility. It can be applied to both day-ahead economic dispatch and real-time control.

This paper begins in Section II by describing the CHP system and the MDP model. Section III details the completion of the DPPO algorithm and the proof of its stability. Case studies are presented in Section IV and Section V gives the conclusions.

3. Problem Statements

In this section, the CHP operational environments and learning scene will be described.

A. System Structure

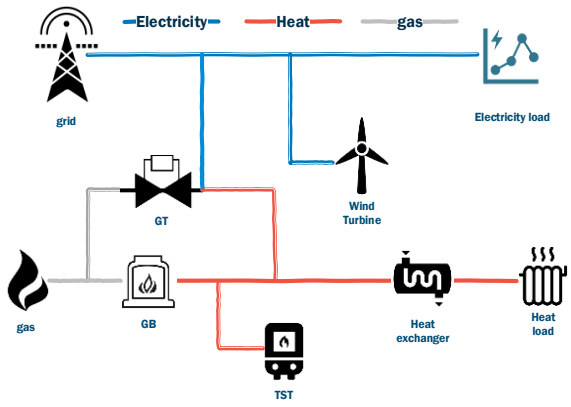

Fig. 1. Structure of the integrated energy system

A CHP system with four random variables (electricity load, heat load, wind and energy price) and four control variables (gas turbine output, gas boiler output, electricity traded with the grid and heat storage or heat release of thermal storage tank), which consequently increase the randomness and the complexity of calculation, is considered in this research. Fig. 1 illustrates the structure of the adopted system, consisting of a 5MWh gas turbine (GT), a 5MW gas boiler (GB), a wind turbine with the capacity of 1.5MW and a thermal storage tank (TST) with the capacity of 5MWh. Electricity load and heat load are connected to the end. This system could be divided into electric and thermal parts, and the GT realizes the coupling of the two parts. In the electric power system, the GT and the wind turbine are the sources of electricity supply. Meanwhile, the system is entitled to trade (buy or sell) electricity freely with the grid within constraints. In the thermal system, the GT and the GB convert gas into heat, and the TST is used to store any excess heat and to feed the stored heat back when there is a heat load after passing through the heat exchanger.

The proposed CHP system has four decision variables, including the output of the GT and that of the GB, the charge/discharge state of the heat storage tank (TST) and the trading volume with the grid. The variables are regulatable and are the main factors affecting the generation of electricity and heat. Apart from these controllable variables, the output of the wind turbine, the electricity load, the heat load and the time-of-use electricity price are randomly generated within the feasible range of real data in each iteration in order to guarantee the robustness of the algorithm in a stochastic environment. Operating parameters and limits for each device are listed in TABLE 6 in the Appendix.

B. Problem Modelling

The CHP economic dispatch problem is to determine the minimized unit cost of generating heat and power on the foundation that the heat and power loads along with other constraints are all met. To achieve the first-rate control strategy, optimal methods are publicly applied in CHP economic dispatch area, where the problem is described as a series of constraints and one or more objective functions [27]. Varieties of optimization algorithms [1] could be used to find optimal solution in feasible operation region.

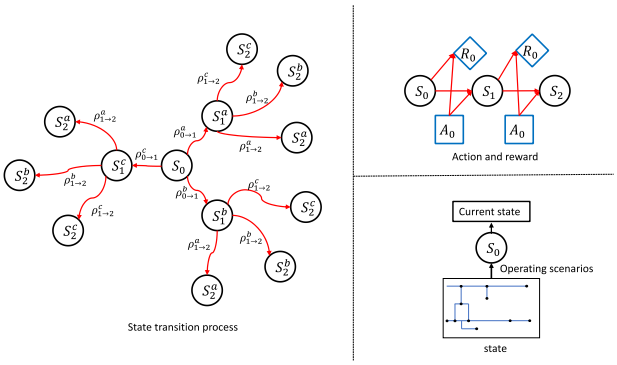

Fig. 2. Illustration of MDP process

MDP model is chosen in this research for its simplicity and computational efficiency, in which an agent interacts with environment over a number of time steps. At each time step t, the agent receives a state $\mathcal{S}$ and selects an action $\mathcal{A}$ according to its policy π. After performing the selected action $\mathcal{A}$, the agent, in return, receives the next state and receive a scalar reward r. The goal of the agent is to maximize the expected return from each state. So the economic dispatch of the CHP system is modeled as an infinite-horizon discounted Markov Decision Process (MDP), defined by the tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, r, \rho_0, \gamma)$, where $\mathcal{S}$ is an array of states, $\mathcal{A}$ is the array of actions, $P: \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the transition probability distribution, $r: S \to \mathbb{R}$ is the reward function, $\rho_0: \mathcal{S} \to \mathbb{R}$ is the distribution of the initial state s_0, and $\gamma \in (0,1)$ is the discount factor. The relationship can be described as the following:

$$\mathcal{S} = (I, c, d, r), \tag{1}$$

$$\mathcal{A} = \left(\Delta p_{gt}, \Delta p_{gb}, \Delta q_{tst}, \Delta p_{grid}\right), \tag{2}$$

I is an indicator function. In a training episode, I equals to 1 if power mismatch is lower than the limit ε for more than N consecutive time steps, otherwise I equals to 0. The stability of the strategy is improved by I.

$c = [p_{gt}, q_{gt}, p_{gb}, q_{tst}, p_{grid}, p_{wind}]$ is the equipment operating status vector.

$d = [(p_l - p_s), (q_l - q_s), p_l, q_l)]$ is the power mismatch vector and indicates the difference between the energy production and the load demand, where p_l is the electricity load, p_s is the electricity supplied, and q_l is the heat load and q_s is the heat supplied.

$r = [tst_i, rtp]$ denotes the value of random variables, where tst_i is the initial state of the TST and rtp is time-of-use price.

$\mathcal{A}$ suggests an action set for the decision variables which denotes the change amount of the decision variables in every single time step.

Let π denote a stochastic policy, and the following are standard definitions of the reward function R_π, the value function V_π, and the advantage function A_π:

$$R_\pi(s_t, a_t) = \mathbb{E}_{s_{t+1}, a_{t+1}, \cdots}[\sum_t^{t+l} \gamma^l r(s_t, a_t)], \quad (3)$$

$$V_\pi(s_t) = \mathbb{E}_{a_t, s_{t+1}, \cdots}[\sum_t^{t+l} \gamma^l r(s_t)], \quad (4)$$

$$A_\pi(s, a) = R_\pi(s, a) - V_\pi(s), \quad (5)$$

The following useful identity expresses the expected return of another policy $\tilde{\pi}$ in terms of the advantage over π, accumulated over time steps (see [28] for proof):

$$\eta(\tilde{\pi}) = \eta(\pi) + \mathbb{E}_{a_t, s_{t+1}, \cdots}[\sum_{t=0}^{\infty} A_\pi(s, a)], \quad (6)$$

where the notation $\mathbb{E}_{a_t, s_{t+1}, \cdots}[\cdot]$ indicates that actions are sampled $a_t \sim \tilde{\pi}(\cdot\,|s_t)$. This equation implies that any policy update $\pi \leftarrow \tilde{\pi}$ that has a nonnegative expected advantage at every state s ($\sum_t \tilde{\pi}(a|s) A_\pi(s, a) \geq 0$ is guaranteed to increase the policy performance η, or leave it constant in the case that the expected advantage is zero everywhere). This implies the classic result that the update performed by exact policy iteration, which uses the deterministic policy $\tilde{\pi}(s) = argmax A_\pi(s, a)$, improves the policy if there is at least one state-action that pairs with a positive advantage value and nonzero state visitation probability, otherwise the algorithm has converged to the optimal policy.

C. Constraints

Constraints are very important in the mathematical optimization problem. The premise of the optimal solutions is to set the allowable range for the constraints. To simulate the real operation of the CHP system, the strict constraints are set in state transition of MDP. In this section, we will demonstrate how we handle constrain in MDP model. For example, if the GT output has reached the maximum in current state S and the action choose by decision maker is still increasing the output of GT, the output of the GT in next state is still maximum to meet equipment operation limit.

Power demands: Electric and thermal power need to reach a supply and demand balance. The followings are the publicly used formulas:

$$p_{gt} + p_{wind} + p_{grid} = P_l, \quad (7)$$

$$q_{gt} + q_{gb} + q_{tst} = Q_l, \quad (8)$$

$$q_{gt} = \alpha p_{gt}, \quad (9)$$

where p_{gt}, p_{wind} and p_{grid} are the electric power of the GT, of the WT and traded with the grid, respectively. q_{gt}, q_{gb} and q_{tst} are the thermal power of the GT, the GB and the TST, respectively. α is the thermoelectric conversion efficiency of the GT. P_l and Q_l are the electricity and heat loads, respectively. This part of constraints is difficult to reflect in state transition P, and we convert it into reward function in the subsection D.

Equipment operation limit: Both the GT and the GB must meet their upper and lower limits of output. Grid interaction power is within the specified range shown in case study.

$$P_{gt}^{min} < P_{gt} < P_{gt}^{max} \quad (10)$$

$$P_{gb}^{min} < P_{gb} < P_{gb}^{max} \quad (11)$$

In MDP, if the P_{gt}or P_{gb} in next state is beyond restriction, the probability of moving from the current state to this state is 0 which means that the agent would not take action that will cause the device to exceed the limit.

Energy storage device constraint: Energy storage device operating constraints are detailed as follow:

$$Q_{(t+\Delta t)}^{tst} = \begin{cases} Q_t^{tst} + q_{char}^{tst} \cdot \Delta t & charge \\ Q_t^{Bt} - q_{dis}^{tst} \cdot \Delta t & discharge \end{cases} \quad (12)$$

s.t.

$$q_{dis}^{min} \leq q_{dis}^{tst} \leq q_{dis}^{max}$$
$$q_{char}^{min} \leq q_{char}^{tst} \leq q_{char}^{max}$$
$$Q_{min}^{tst} \leq Q_t^{tst} \leq Q_{max}^{tst}$$

Where Q_t^{tst} denotes heat storage of the heat storage tank at time t. q_{char}^{tst} and q_{dis}^{tst} is the charging rate and discharging rate of TST respectively. The constraints on charge and discharge rate are reflected in the action $\mathcal{A}[\Delta q_{tst}]$ in the MDP model: $q_{dis/char}^{min} < \mathcal{A}[\Delta q_{tst}] < q_{dis/char}^{max}$. The treat to heat storage capacity limits is same as Equipment operation limit.

D. Reward

In a Reinforcement learning problem, all objective function mentioned in optimization problems can be described as maximizing the expected cumulative reward signal [29]. Reasonable rewards must be set in order to guide the algorithm to continuously learn from the target. In this research, the rewards for all operational status were kept simple and consistent in different environments (i.e. the output of WT, the electricity load, the heat load and the time-of-use price). The reward consists of 3 sub-targets in "(14)": 1) total operating costs ($-\widetilde{c_{gas}} - \widetilde{c_{grid}}$): encouraging the agent to reduce the operating cost; 2) power mismatch ($-0.5 * \|d\|^2 + 5 * I[\|d\|^2 < \varepsilon]$): besides the penalty of power mismatch, additional rewards were added when the system reached a power balance, encouraging the agent to minimize the power mismatch; 3) storage tank status ($(s_{tst} - \vartheta)^2$): the penalty for heat storage was added in order to guarantee the stored heat is in a safer range, i.e. there should be enough storage to deal with unexpected situations but not too much storage.

$$c_{gas} = \rho_{gas}(\frac{p_{gt}^{t_d}}{\eta_{gt}} + \frac{Q_{gb}^{t_d}}{\eta_{gb}})\Delta t_d, \quad (13)$$

$$c_{grid} = \rho_{grid} p_{gt}^{t_d} \Delta t_d, \quad (14)$$

$$d = (\sum P_{output} - P_l, \sum Q_{output} - Q_l) \quad (15)$$

$$r = -\widetilde{c_{gas}} - \widetilde{c_{grid}} - 0.5 * \|d\|^2 + 5I[\|d\|^2 < \varepsilon] - 0.1 * (s_{tst} - \vartheta)^2 \quad (16)$$

Where c_{gas} and c_{grid} represent the natural gas cost and the grid interaction cost respectively (superscript ~ means that this parameter has been normalized), ρ_{gas} and ρ_{grid} are unit price of natural gas and buy/sell electricity respectively, t_d

denotes one timestep in economic dispatch, d is the power mismatch vector, $I[\cdot]$ is the indicator function, ε is the maximum cumulative power mismatch, s_{tst} is the current heat storage capacity of the heat storage tank and ϑ is the reasonable operating capacity.

E. Random environment setting

To simulate all possible CHP system operating status, the algorithm is trained on different types of courses. We collected the real operational data of CHP system from https://data.open-power-system-data.org/ and determine the upper and lower limits of four variables. In every episode a new course is generated within the upper and lower limits. Different load, time-of-use price, WT output and initial state of TST are considered, which could include the following typical types: a) morning: high heat load with very low electricity load; b) midday: higher electricity load and lower heat load; c) evening: high electricity load and high heat load. There are examples in Section IV which consist of a sequence of random instantiations of the above environment types within user-specified parameter ranges. Both the time-of-use price and the WT output are generated randomly within the pre-defined range.

4. DPPO For Economic Dispatch Problem

Purpose of this algorithm is for economic dispatch problem in rich simulated CHP environments with continuous/sequential state and action spaces. It is required that the algorithms are robust across a wide range of state variation and are effective for CHP systems with high uncertainty. Finally, the strategy learnt by the algorithm should also satisfy all the constraints and ultimately achieve the optimal function. It is described in this section how to derive a practical algorithm for the CHP system. A DPPO setup has been considered to learn the parameterized policies $\pi_\theta(a|s)$ with the parameter vector θ, and a baseline function V_π. The architecture consists of a set of agents, the repeatedly generating trajectories of experience, and one chief learner that uses the experiences sent from agents to learn π off-policy. Besides, we will give a proof of convergence and optimality of the algorithm. FIGURE 3. Illustrates how to train the intelligent agent and how to use it in different scenarios. In training process, the generated random variables were passed into action network, and the action network will generate actions accordingly. Then the value network will evaluate the action strategy. Besides, the simulation CHP environment will proceed to the next state according to the current action and return the reward value. In order to simulate as many situations as possible, each state will be executed 300 times. Finally, update the action network and value network parameters with the goal of maximizing the product of the reward and the evaluation value. In other words, the action network can be understood as an experience pool and accumulate experience about CHP economic dispatch during the training process. When the training is completed, the user can use the action network in real-time scheduling or day-ahead optimization scheduling.

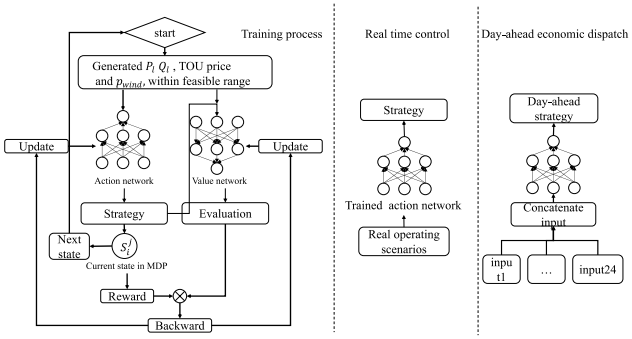

Fig. 3. Training process and how to use

A. Proximal Policy Optimization with Clipped Surrogate Objective

The accurate way to update the algorithm is mentioned in the previous section. Unfortunately, this proposed update is not possible in continuous dispatch problem since the computation of $\tilde{\pi}(s) = argmax A_\pi(s,a)$ is excessively time consuming. Hence, the accurate update way is inaccessible in the approximate setting for the estimation and approximation error. As a result, there will be some states s for which the expected advantages are negative, i.e. $\sum_a \tilde{\pi}(a|s)A_\pi(s,a) < 0$. The complex dependency of $\rho_{\tilde{\pi}}(s)$ on $\tilde{\pi}$ makes it difficult to converge to the optimal policy.

Instead, PPO algorithms is introduced to make some changes on target value function $\eta(\tilde{\pi})$. It is implied by equation (6) that our approach is guaranteed to improve the true objective η by performing the following maximization:

$$J = \max_\theta \mathbb{E}_{a_t, s_{t+1}, \ldots}[\sum_{t=0}^{\infty} A_\pi(s,a)]. \quad (17)$$

In practice, Clipped Surrogate Objective function has been chosen to replace this maximization for the higher robustness than updating policies directly as J. (Clipped Surrogate Objective and Kullback-Leibler divergence (KL) penalty are two widely accepted methods for policy update in PPO algorithms. John Schuman found that KL penalty performed worse than clipped surrogate objective in [26]. Besides, several modifications were added to the core algorithm in both [24] and [26], which includes the normalization of inputs and the accumulation of rewards through timestep with a window of length n as well as bootstrap from the value function after n-steps. Similar augmentations were adopted in this paper.)

Given a parameterized policy π_θ, where $\pi_\theta(a_t|s_t)$ is a differentiable function of parameter vector θ, our research supposed that trajectory $(s_t, a_t, r_t)_{t=k}^{t=k+n}$ was generated by the agents with the policy π_θ. Let $z_t(\theta)$ denote the probability ratio $\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{old}}(a_t|s_t)}$, so $z_t(\theta_{old}) = 1$ means the strategy π_θ has not changed. Hence, the n-steps clipped target value function J could be re-written as:

$$J(\theta) = \mathbb{E}\big[\min(z_t(\theta)\,\hat{A}_t, clip(z_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\big] \quad (18)$$

where ϵ is a hyperparameters, the second $clip$ term modifies the surrogate objective function $J(\theta)$ by clipping the probability ratio, by which the range of action changes were clipped in a reasonable scope. Finally, the minimum objective is taken to ensure the objective function is a lower bound on the unclipped objective. In that case, the change in probability ratio was ignored only when it improves the objective, otherwise it was included when it deteriorates the

objective. The key idea of this target value function is that the probability ratio ($z_t(\theta)$ was clipped at $1-\epsilon$ or $1+\epsilon$ depending on whether the advantages is positive or negative. This assures that the policy change would not be too intense when the advantage is positive, and the update direction is correct when the advantage is negative. As aforementioned, given $A_\pi(s,a)$ was estimated in continuous problems, $\hat{A}_t$ represents an advantage estimator for n timesteps (where n is much less than the episode length) as:

$$\hat{A}_t = \sum_{t=k}^{k+n-1} \gamma^{t-k}\left(r_t + \gamma V_\theta(s_{t+1}) - V_\theta(s_t)\right), \quad (19)$$

Algorithm I DPPO-chief

for iteration=1, 2…M **do**
  for actor=1, 2… N **do**
    Run policy π_θ for K timesteps, collecting $\{s_t, a_t, r_t\}$
    Estimate $R_\pi(s_t,a_t), V_\pi(s_t)$ and A_π
  end for
  push data to main PPO
  $\pi_{old} \leftarrow \pi_\theta$
  Optimize surrogate loss and update global action π and critic network parameters
end for

B. *Distributed Settings*

To achieve good performance in various randomly generated scenes, agents must be guaranteed to explore in as many different environments as possible. Therefore, distributed setup has been introduced to the PPO algorithm. Data was collected in different environments by multiple threads simultaneously and all parallel threads share a global learner. The chief learner learns and develops through the experience collected by different threads. The chief learner setting is similar to A3C in [25]. The difference exists where in our setting that each thread does not compute nor push the gradient of its own policy update to the global PPO net, which promotes the efficiency of the multi-threaded data collection.

A Distributed Proximal Policy Optimization algorithm that uses clipped surrogate objective and distributed architecture is shown in Algorithm I. In each episode, each of the N (parallel) workers (agents) runs policy π_θ for K timesteps, collecting data $\{s_t, a_t, r_t\}$ and estimating the reward function $R_\pi(s_t,a_t)$, the value function $V_\pi(s_t)$ and the advantage function A_π. Besides, workers are required to push data to the chief net. Then the surrogate loss is constructed on NK timesteps of data, and optimized with Adam optimization [30][31]. Pseudocode are provided in Algorithm II. U is the number of sub-iterations with policy update when a batch of data was collected.

Algorithm II agents

for iteration=1, 2... **do**
  for actor=1, 2… N **do**
    Run policy π_θ for K timesteps, collecting $\{s_t, a_t, r_t\}$
    Estimate $R_\pi(s_t,a_t), V_\pi(s_t)$ and A_π
  end for
  push data to main PPO
  $\pi_{old} \leftarrow \pi_\theta$
  for $m \in \{1, 2, \ldots U\}$ **do**
    $J_{CLIP}(\theta) = \mathbb{E}_{\rho\theta_{old}}(\tau)\left[\min(z_t(\theta)\hat{A}_t, clip(z_t(\theta), 1-\epsilon, 1+\epsilon)\hat{A}_t)\right]$
    Send collect data to chief
    Wait until all agents end this episode
    Chief compute main $\nabla_\theta J$
    Update chief-policy parameters θ
  end for
end for

C. *Observations and Network Structure for DPPO Algorithm*

When applied in economic dispatch in CHP system, the agent receives two sets of observation: 1) A set of states information, containing the operating status of the GT, the GB, the TST and the Grid. The agents collect this data set in every timestep and then push it to the main PPO net. 2) A set of uncertain information, including the output of the wind turbine, the energy price and the load which are all difficult to give accurate values. Hence, these data sets are generated stochastically in each iteration, due to the high randomness of the wind power, the energy price and the load. Then the action network and the value network compute the action set and $V_\pi(s_t)$, respectively, with the input of observations. The actor network architecture of actor network is illustrated in FIGURE 4.

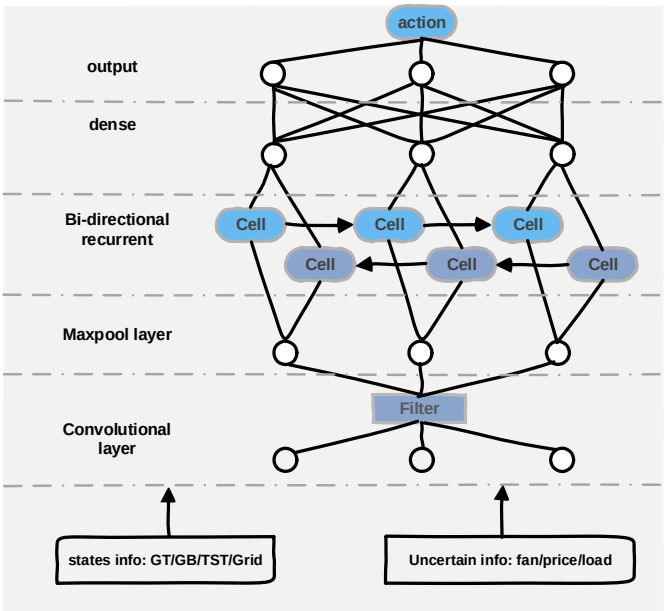

Fig. 4. Schematic drawing of network architecture

Since the learner updates on entire batches of trajectories, it is able to parallelize most of its computations. We use an architecture similar to [31], consisting convolutional layer, bi-directional recurrent layer and dense layer. This optimization architecture increases the effective batch size to thousands and LSTM-based agents also obtain significant speedups on the learner by exploiting the network structure dependencies and operation fusion [32]

5. Case Study

MDP and DPPO algorithm is applied to the optimized scheduling of the CHP system in FIGURE. 1. This experiment aims to prove that the proposed algorithm is applicable to the optimization problems with stochastic environments. The performance of the DPPO was compared with that of the classical optimization methods. The experiment also aims to find out whether this algorithm is capable of coping with emergencies such as a WT

disconnection. More parameters settings are in the supplementary material. For the modelling of the MDP and DPPO algorithm, Python is selected as the programming language and TENSERFLOW is used as the framework. For the benchmark optimization methods, MATLAB is used to model the system and Yalmip toolkit and GUROBI are chosen as optimization solver.

A. Case I: Comparison with other DRL algorithms

The performance of the proposed DPPO algorithm was compared with those of the Distributed Deterministic Policy Gradient (DDPG) [33] and of the A3C [25] which had already been recognized as effective solutions to continuous problem. The experiment CHP system was set in random mode with diverse load, wind power and energy price. FIGURE. 5 shows that all three algorithms applied in learning a stable regulation strategy, and DDPO took substantially less training time (402 seconds) under the same number of iterations than DPPG (720 seconds) and A3C (603 seconds). The DPPO and DPPG algorithms converge, while the a3c algorithm does not converge. (The hyperparameters of the two algorithms was shown in TABLE 5 in appendix)

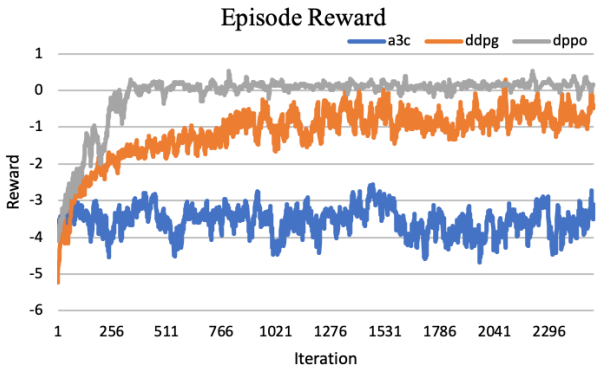

Fig. 5. Comparison of the three algorithms on CHP environment, trained for 2500 iterations. The training process took 402 seconds for DPPO, 720 seconds for DDPG and 603 seconds for A3C, respectively.

Then, the trained DDPG and DPPO algorithms was applied to the three aforementioned typical CHP operating conditions (detailed parameters were shown in TABLE 7 in appendix). Fig. 4 and TABLE I demonstrate the results. The agents of both the DPPO and DDPG algorithms learnt to adjust the output of each equipment reliably in different operating conditions, with only small variations. They learnt that the GT, which generating both electricity and heat, was more efficient than the GB. Both algorithms succeeded in achieving their economic optimums by learning how to adjust all devices to cope with changes in the environment. The final marginally higher reward of DPPO than that of DDPG, as shown in FIGURE. 5, is attributed to the difference in the two algorithms, i.e. the DPPO can explore more scenarios and ensure that the solution is the optimal solution.

In economic dispatch realm, economy is the first criterion on the basis of meeting operation restrictions. Therefore, the DPPO presented with the drastically reduced time cost in this paper is the superior candidate for multi-objective optimization problems with large scale variables, e.g. the economic dispatch problem

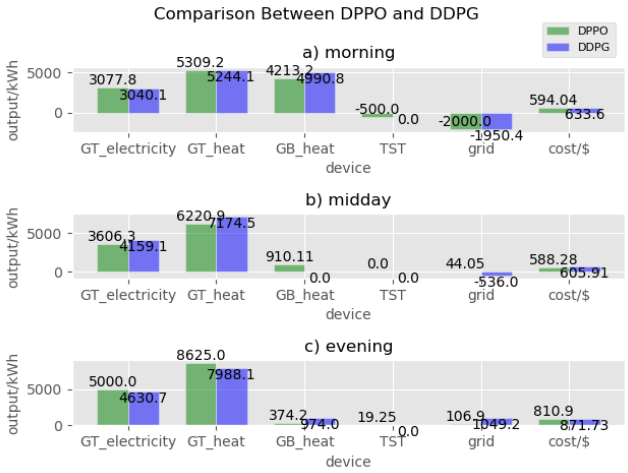

Fig. 6. Performance comparison DDPG and DPPO algorithms on typical conditions. (Negative output means that TST releases heat to the CHP system, and conversely stores heat energy)

Table 1 Detailed Results

	Morning		Midday		Evening	
Algorithm	DPPO	DDPG	DPPO	DDPG	DPPO	DDPG
Cost/($)	594.04	633.6	588.28	605.91	810.9	871.73
Heat error	0.04	0.004	0.036	0.026	0.007	0.0008
Electric error	0.009	0.001	0.012	0.0098	0.064	0.037

B. Case II: Details in One Episode

In this case, we will show how agent work in one episode. The specific parameters are set as follows and FIGURE 7 demonstrate the detailed adjustment process. To meet the device operating constraints, action range is $[-0.02,\ 0.02]$. By comparing TABLE 2 and TABLE 7 in the appendix, it can be found that the current situation has a lower electrical load level and a higher thermal load level, wind power and time-of-use electricity price are relatively low. In theory, user should increase the output of GB to meet the heat load without excessive electrical load. At the same time, due to the lower energy price, the heat storage tank should reserve some heat.

Table 2 Case II

electricity load (kW)	wind (kW)	heat load (kW)	TOU price ($/kWh)
6000	700	9000	0.0627

Figure 7 shows the actual adjustment process of DRL agent. It increases the output of GT and GB to meet the user load within the feasible domain and sells excess power to the grid to reduce operating costs. Furthermore, it finds that the energy price is lower at this time, which is suitable for charging TST.

The strategy generated by DRL agent is in line with theoretical analysis and take the economy into account. In actual operation, the user only needs to input the detailed information of the current load, electricity price, etc., to get the control strategy which increases the flexibility and ease of use.

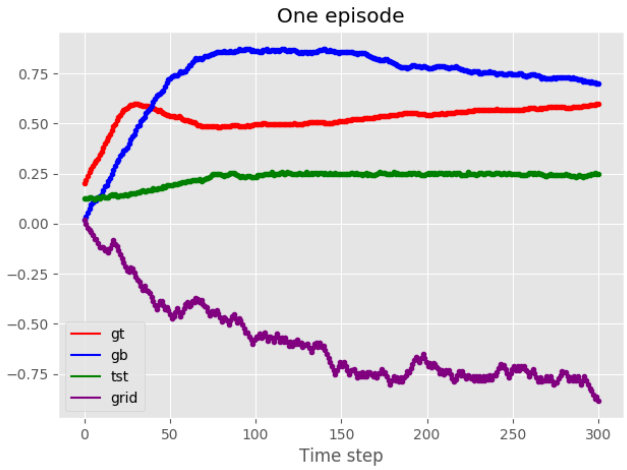

Fig. 7. Operating states

C. Case III: Day-ahead Economic Dispatch Problem

The performance of the DPPO algorithm was evaluated, and the DPPO was then applied to day-ahead generation dispatch problem, whose result was subsequently compared with that of the optimization method. It was assumed that the electricity load had to be matched by the electric power output of the CHP system while the heat load had not to be precisely matched as long as the users' acceptable temperature range was guaranteed [27]. Operating parameter settings are provided in TABLE 7 in appendix. FIGURE 8 shows the comparison between the economic dispatch strategies generated by the DPPO algorithm and the optimization method. Yalmip toolkit and GUROBI solver were adopted, as shown in FIGURE 7, b) and d) respectively, to model and to solve our CHP system. The result demonstrates the following characteristics: 1) The GB output was time period dependent. For example, the output of GB was relatively higher when the heat load is higher from 0 am to 5 am and from 19 pm to 24 pm. 2) The GT undertook most of the electricity and heat loads. 3) The TST was used less frequently, for only 7 time periods, than other devices in the system. load demand reliably, with only acceptable variations in heat load across the time. The strategy learnt through DRL are similar to that through the optimization method, despite a slight numerical gap at every time step.

FIGURE 8. a) and c) imply that the DPPO succeeded in discovering the economical approaches on handling the load changes by choosing the GT as the main load bearer for its more economical performance and adjusting other decision variables based on the environment. However, comparing the real heat load with the CHP output, the DPPO results are different from the results of the optimization algorithm. This is due to the difference in the objective function. In DRL method, we aim to minimize the error between load and generation. However, in practical applications, we should not consider the difference between the heat load and the user's comfort. Therefore, in the optimization method, we aim at the user's indoor temperature not exceeding the limit which results in a large difference between the CHP output thermal power and the actual load. In the DRL method, we have not completed the conversion between this goal and the reward.

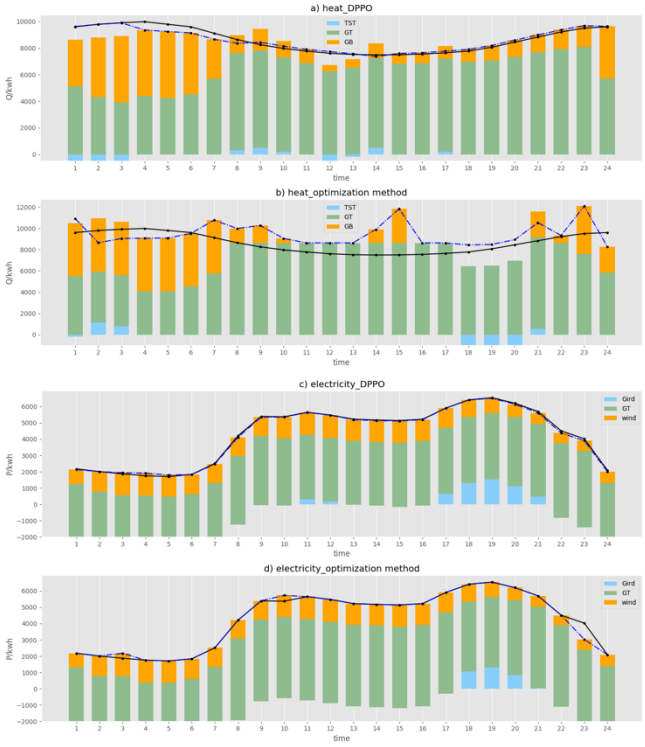

Fig. 8. Comparisons of dispatch strategies. In the heat subplots, the black curves marked with point means ideal heat load, the blue curves marked with point shows the heat output of CHP. In the electricity subplots, the black curves marked with point means ideal electricity load, the blue curves marked with point shows the electricity output of CHP. Beside the illustration, if the bottom of the histogram is less than 0, that part is used to charge the TST. In the electricity subplots, the two curves have the same meaning with the heat subplots, and if the bottom of the histogram is lower than 0, that part means selling the electricity to the grid.

In addition to the qualitative analysis, the heat load error and the cost are listed in TABLE 3. The heat load error in DPPO, as shown in the second column in TABLE 3, was successfully kept at a very low level, indicating the user's comfort zone was well preserved, which approves the accuracy of the DPPO algorithm. The economic performance, i.e. the cost, of the two methods was also compared. The DPPO operated at lower costs for the majority time periods as the lower costs are highlighted in green in TABLE 3. Attributed to the assumption that the heat load does not need to be strictly met in real time, the DPPO attempted to maintain the temperature merely within the range of the user's comfort zone for economical reason, i.e. declining the output of the equipment over a certain period of time and increasing the output when the temperature is about to drop out of the user's comfort zone. In contrast, the optimization method simply kept the temperature to the optimal value. Judging from the total cost of the day, the DPPO has the tiny advantage by making a 0.03% saving as costing of the optimization method.

Table 3 Heat Load Mismatch and Cost

	heat load error	Cost/$	
	DPPO	optimization methods	DPPO
0:00	0.0049	674.77	580.74
1:00	0.00035	592.65	564.15
2:00	0.002	595.18	557.27
3:00	0.025	528.56	554.05
4:00	0.008	529.33	544.08
5:00	0.009	567.91	546.31
6:00	0.009	650.84	526.36
7:00	0.033	702.68	616.83
8:00	0.03	800.41	737.72
9:00	0.03	747.36	694.55
10:00	0.025	709.5	708.3
11:00	0.026	706.25	651.15
12:00	0.032	690.84	646.26
13:00	0.034	762.13	657.53
14:00	0.02	868.03	726
15:00	0.02	691.24	657.45
16:00	0.03	753.85	731.47
17:00	0.02	770.38	807.27
18:00	0.03	799.15	847.59
19:00	0.0292	794.57	840.43
20:00	0.032	921.02	810.2
21:00	0.031	712.18	718.76
22:00	0.031	815.96	728.76
23:00	0.031	538.56	613.26
total	0.007	16924.029	16874.28

D. Case IV: WT Failure

It is essential to investigate whether the DPPO could deal with any emergency. To evaluate the DPPO on different tasks, the trained network was subjected to an extreme CHP environment, i.e. without the WT, and the result was compared with the normal case. The comparison of the dispatch strategies in the two operating status is shown in FIGURE 9. When there was no WT output, the DPPO algorithm acquired a robust dispatch strategy compared with the normal strategy: 1) In the morning setting with the low electricity load and the high heat load, the DPPO managed to increase the output of the GT appropriately in order to slightly reduce the heat output of the GB, to use the stored heat in the TST and to sell the same amount of power to the grid. The strategy can be rationalized by the fact that the GT has the best economic efficiency in the system. By turning up the output of the GT, the gap in the electricity supply caused by the absence of the WT was accurately met and, simultaneously, excessive heat was generated to relieve the burden of the GB.

2) In the midday and the evening settings with the high enough electricity load and the declined heat load, the DPPO decided to reduce the output of the GT and to increase the power purchase from the grid, for optimal economic efficiency since it is cheaper to buy electricity from the grid rather than to generate. Meanwhile, the DPPO adjusted the output of the GB accordingly to meet the rest of the heat load and stored the excessive heat in the TST for future use, which further promotes the economic performance. In TABLE 4, the resulted changes are demonstrated.

Compared with the optimization method, the DPPO is also more advantageous in the solving speed. This is resulted from one of its characteristics that once the training of the DPPO is completed, there is no need to retrain for each new situation. In other words, for any new operation status, the calculation time for the DPPO is always next to zero.

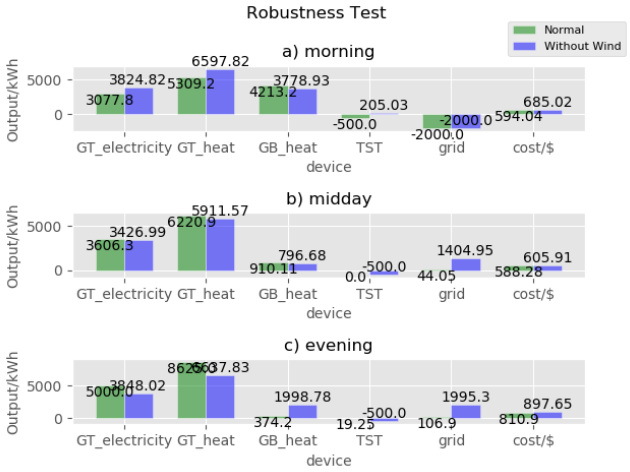

Fig. 9. Robustness Test

Table 4 ROBUSTNESS TEST RESULT

	Morning		Midday		Evening	
Condition	Normal	No wind	Normal	No wind	Normal	No wind
Cost/($)	594.04	685.02	588.28	691.50	810.9	897.65
Heat error	0.04	0.03	0.0036	0.04	0.007	0.0027
Electric error	0.009	0.03	0.012	0.03	0.064	0.037

6. CONCLUSION

We proposed and analyzed the DPPO algorithms for optimizing the stochastic CHP economic dispatch problem. We modeled the CHP economic dispatch problem as infinite-horizon discounted Markov decision process and set constrains to simulate the real environment. A form of reward signal was designed to lead the algorithm to the goal. We introduced proximal policy optimization methods that use multiple epochs of stochastic gradient ascent to perform each policy update and proved the convergence of the algorithm. Besides, we also used asynchronous advantage actor-critic to improve the convergence rate of the distributed framework, which subsequently improved the data collection speed, making it applicable to CHP settings where samples are expensive.

In the domain of the CHP economic dispatch, we successfully taught the agents to schedule the devices in the CHP system when chasing the economic optimum while satisfying load demand. Due to the fact that the dispatch activities involve five continuous variables, it is essential to optimize the high-dimensional and sequential policies. Thus, we utilized the LSTM-convolutional neural network policies that used two types of observations as inputs. Our analysis

shows the DPPO algorithm could optimize the certain objective to a constraint.

In case study, the result shows that the training time of our improved algorithm is 201 seconds and 318 seconds less than other two advanced DRL algorithm. And the difference on economic performance between this method and optimization methods is only 0.029%. The proposed method can cope with more situations, have better time scale flexibility, and is easier to use on the basis of the same economic performance as the optimization method.

However, there are still shortcomings in solving economic dispatch problems with DRL methods. For examples, all optimization goals are reflected in the reward formula, which is not conducive to achieving multi-objective optimization, and optimization goals closer to the user's needs (As we analyzed in Case III).

Since the method we proposed is economical when compared with optimization algorithms and other DRL algorithm, we hope it can sever as a choice for future work on optimization problem. The result promises high robustness and high efficiency in learning economic dispatch policy. Although the results may not be optimal yet, it would definitely benefit from improvements such as combining the proposed DPPO with other recognized optimization theory, improving its applicability to real-world settings.

7. References

[1] M. Nazari-Heris, B. Mohammadi-Ivatloo, and G. B. Gharehpetian, "A comprehensive review of heuristic optimization algorithms for optimal combined heat and power dispatch from economic and environmental perspectives," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 2128–2143, Jan. 2018.

[2] W. Gu, Z. Wang, Z. Wu, Z. Luo, Y. Tang, and J. Wang, "An Online Optimal Dispatch Schedule for CCHP Microgrids Based on Model Predictive Control," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2332–2342, Sep. 2017.

[3] Tao Guo, M. I. Henwood, and M. van Ooijen, "An algorithm for combined heat and power economic dispatch," *IEEE Trans. Power Syst.*, vol. 11, no. 4, pp. 1778–1784, Nov. 1996.

[4] Y. Dai *et al.*, "A General Model for Thermal Energy Storage in Combined Heat and Power Dispatch Considering Heat Transfer Constraints," *IEEE Trans. Sustain. Energy*, vol. 9, no. 4, pp. 1518–1528, Oct. 2018.

[5] Y. Dai *et al.*, "Dispatch Model of Combined Heat and Power Plant Considering Heat Transfer Process," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1225–1236, Jul. 2017.

[6] S. Lu, W. Gu, J. Zhou, X. Zhang, and C. Wu, "Coordinated dispatch of multi-energy system with district heating network: Modeling and solution strategy," *Energy*, vol. 152, pp. 358–370, Jun. 2018.

[7] S. Bahrami and A. Sheikhi, "From Demand Response in Smart Grid Toward Integrated Demand Response in Smart Energy Hub," *IEEE Trans. Smart Grid*, pp. 1–1, 2015.

[8] M. Geidl and G. Andersson, "A modeling and optimization approach for multiple energy carrier power flow," in *2005 IEEE Russia Power Tech*, St. Petersburg, Russia, 2005, pp. 1–7.

[9] G. Chicco and P. Mancarella, "Matrix modelling of small-scale trigeneration systems and application to operational optimization," *Energy*, vol. 34, no. 3, pp. 261–273, Mar. 2009.

[10] P. Mancarella, "MES (multi-energy systems): An overview of concepts and evaluation models," *Energy*, vol. 65, pp. 1–17, Feb. 2014.

[11] C. Sondergren and H. F. Ravn, "A method to perform probabilistic production simulation involving combined heat and power units," *IEEE Trans. Power Syst.*, vol. 11, no. 2, pp. 1031–1036, May 1996.

[12] R. Lahdelma and H. Hakonen, "An efficient linear programming algorithm for combined heat and power production," *Eur. J. Oper. Res.*, vol. 148, no. 1, pp. 141–151, Jul. 2003.

[13] S. Makkonen and R. Lahdelma, "Non-convex power plant modelling in energy optimisation," *Eur. J. Oper. Res.*, vol. 171, no. 3, pp. 1113–1126, Jun. 2006.

[14] A. Rong and R. Lahdelma, "An efficient envelope-based Branch and Bound algorithm for non-convex combined heat and power production planning," *Eur. J. Oper. Res.*, vol. 183, no. 1, pp. 412–431, Nov. 2007.

[15] F. J. Rooijers and R. A. M. van Amerongen, "Static economic dispatch for co-generation systems," *IEEE Trans. Power Syst.*, vol. 9, no. 3, pp. 1392–1398, Aug. 1994.

[16] K. P. Wong and C. Algie, "Evolutionary programming approach for combined heat and power dispatch," *Electr. Power Syst. Res.*, vol. 61, no. 3, pp. 227–232, Apr. 2002.

[17] M. Nazari-Heris, B. Mohammadi-Ivatloo, S. Asadi, and Z. W. Geem, "Large-scale combined heat and power economic dispatch using a novel multi-player harmony search method," *Appl. Therm. Eng.*, vol. 154, pp. 493–504, May 2019.

[18] M. A. Mellal and E. J. Williams, "Cuckoo optimization algorithm with penalty function for combined heat and power economic dispatch problem," *Energy*, vol. 93, pp. 1711–1718, Dec. 2015.

[19] D. Zou, S. Li, X. Kong, H. Ouyang, and Z. Li, "Solving the combined heat and power economic dispatch problems by an improved genetic algorithm and a new constraint handling strategy," *Appl. Energy*, vol. 237, pp. 646–670, Mar. 2019.

[20] D. P. Bertsekas, "Dynamic Programming and Optimal Control 3rd Edition, Volume II," p. 233.

[21] R. Rocchetta, L. Bellani, M. Compare, E. Zio, and E. Patelli, "A reinforcement learning framework for optimal operation and maintenance of power grids," *Appl. Energy*, vol. 241, pp. 291–301, May 2019.

[22] S. Zhou, Z. Hu, and W. Gu, "Artificial intelligence based smart energy community management: A reinforcement learning approach," *CSEE J. Power Energy Syst.*, 2019.

[23] E. A. Jasmin, T. P. Imthias Ahamed, and V. P. Jagathy Raj, "Reinforcement Learning approaches to Economic Dispatch problem," *Int. J. Electr. Power Energy Syst.*, vol. 33, no. 4, pp. 836–845, May 2011.

[24] N. Heess *et al.*, "Emergence of Locomotion Behaviours in Rich Environments," *ArXiv170702286 Cs*, Jul. 2017.

[25] V. Mnih *et al.*, "Asynchronous Methods for Deep Reinforcement Learning," *ArXiv160201783 Cs*, Feb. 2016.

[26] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *ArXiv170706347 Cs*, Jul. 2017.

[27] W. Gu *et al.*, "Residential CCHP microgrid with load aggregator: Operation mode, pricing strategy, and optimal dispatch," *Appl. Energy*, vol. 205, pp. 173–186, Nov. 2017.

[28] J. Langford, "Approximately Optimal Approximate Reinforcement Learning," 2002.

[29] D. Silver *et al.*, "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018.

[30] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ArXiv14126980 Cs*, Dec. 2014.

[31] B. Shi, X. Bai, and C. Yao, "An End-to-End Trainable Neural Network for Image-Based Sequence Recognition and Its Application to Scene Text Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298–2304, Nov. 2017.

[32] J. Appleyard, T. Kocisky, and P. Blunsom, "Optimizing Performance of Recurrent Neural Networks on GPUs," *ArXiv160401946 Cs*, Apr. 2016.

[33] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," *ArXiv150902971 Cs Stat*, Sep. 2015.

8. APPENDIX

A. Hyperparameters of DPPO algorithm

Table 5 DPPO HYPERPARAMETERS

Hyperparameter	Value
Discount γ	0.9
Adam update rate for actor	0.0001
Adam update rate for critic	0.0005
Update step	10
Minibatch size	24
Clipping parameters ϵ	0.2

B. DEVICE OPERATING PARAMETERS

Table 6 Device Operating Parameters

Hyperparameter	Value
Discount γ	0.9
Adam update rate for actor	0.0001
Adam update rate for critic	0.0005
Update step	10
Minibatch size	24
Clipping parameters ϵ	0.2

C. Environment Variables

Table 7 Environment Variables

time interval	electricity load (kW)	wind (kW)	heat load (kW)	TOU price ($/kWh)
00:00—01:00	2,178	875	9600	0.065
01:00—02:00	2,009	1,234.00	9792	0.065
02:00—03:00	1,873	1,390.00	9907.2	0.065
03:00—04:00	1,755	1,392.00	9984	0.065
04:00—05:00	1,704	1,336.00	9792	0.065
05:00—06:00	1,839	1,223.00	9600	0.065
06:00—07:00	2,517	1,173.00	9120	0.08
07:00—08:00	4,211	1,136.00	8640	0.08
08:00—09:00	5,397	1,158.00	8256	0.095
09:00—10:00	5,735	1,312.00	7968	0.095
10:00—11:00	5,651	1,369.00	7776	0.095
11:00—12:00	5,481	1,376.00	7603.2	0.08
12:00—13:00	5,227	1,315.00	7516.8	0.08
13:00—14:00	5,176	1,301.00	7488	0.08
14:00—15:00	5,143	1,343.00	7497.6	0.08
15:00—16:00	5,227	1,310.00	7545.6	0.08
16:00—17:00	5,909	1,208.00	7641.6	0.08
17:00—18:00	6,417	1,055.00	7776	0.095
18:00—19:00	6,545	896	8064	0.095
19:00—20:00	6,206	773	8448	0.095
20:00—21:00	5,698	672	8832	0.095
21:00—22:00	4,510	626	9216	0.095
22:00—23:00	3,025	624	9504	0.065
23:00—24:00	2,093	703	9600	0.065