Multi-energy Management of Interconnected Multi-microgrid System Using Multi-agent Deep Reinforcement Learning

Sichen Li, Di Cao, Weihao Hu, Qi Huang, Zhe Chen, and Frede Blaabjerg

***Abstract*—The multi-directional flow of energy in a multi-microgrid (MMG) system and different dispatching needs of multiple energy sources in time and location hinder the optimal operation coordination between microgrids. We propose an approach to centrally train all the agents to achieve coordinated control through an individual attention mechanism with a deep dense neural network for reinforcement learning. The attention mechanism and novel deep dense neural network allow each agent to attend to the specific information that is most relevant to its reward. When training is complete, the proposed approach can construct decisions to manage multiple energy sources within the MMG system in a fully decentralized manner. Using only local information, the proposed approach can coordinate multiple internal energy allocations within individual microgrids and external multilateral multi-energy interactions among interconnected microgrids to enhance the operational economy and voltage stability. Comparative results demonstrate that the cost achieved by the proposed approach is at most 71.1% lower than that obtained by other multi-agent deep reinforcement learning approaches.**

***Index Terms*—Interconnected multi-microgrid system, energy management, combined heat and power, demand response, deep reinforcement learning.**

I. Introduction

MULTI-ENERGY microgrid (MG) is a new paradigm for the generation, transmission, and consumption from heterogeneous energy carriers such as electrical and thermal energy sources at the distribution-network (DN) level [1], [2]. Typically, the components of a multi-energy MG at this level include distributed energy resources, energy-coupling equipment, local active loads, and energy storage systems (ESSs) [3]. Recently, energy coupling innovations such as micro-turbines and electric heat pumps have been integrated with multiple energy carriers to enhance the economics and environmental sustainability of energy systems [4], [5]. Multi-energy MGs have thus evolved into a cost-effective and reliable strategy for providing both multi-energy supply via enhanced utilization of renewable energy sources (RESs) and multi-energy coordination. However, owing to the local nature of their power supply, multi-energy MGs and RESs have limited energy supply capabilities [6], [7]. To overcome this deficiency, several neighboring MGs can share energy in certain areas to address the capacity limitations of individual MGs [8], [9]. Hence, in addition to directly connecting each individual MG to the DN, multiple MGs can be interconnected into a multi-microgrid (MMG) system over an area and to the DN to improve the economic benefits and power supply reliability for both the MMG system and DN [10], [11]. In a multi-energy MG, consumers typically demand large amounts of electrical and thermal energy simultaneously [12]. Thus, energy management is critical for reliable and efficient multi-energy MG operation and control [13], particularly for multi-energy MMG systems.

An entirely centralized control system generally involves a specialized controller that performs a variety of functions, including gathering data and calculating, optimizing, and determining the control actions that will be applied to the controlled units. In addition, the central controller and controlled units must interact via an extensive communication network to execute all these functions from a single site. For example, particle swarm optimization is applied in [14] to determine the optimal scheduling of interconnected MMGs in a centralized manner for minimizing the operating costs. However, given the real-time requirements for MG operation and complex energy management of interconnected MMGs (for energy management of a single MG and interactions between MGs and between MGs and the DN), the approach in [14] cannot achieve the desired results. In view of this, an imperialist competitive-based algorithm with faster, more accurate, and stronger global convergence than the approach in [14] is introduced in [15] for energy management of interconnected MMGs. Furthermore, an improved linear control and dispatch model is devised in [16] for integrated energy systems to reduce the complexity of calculation and control.

The abovementioned studies are aimed at adjusting central-

Manuscript received: July 30, 2022; revised: December 8, 2022; accepted: December 28, 2022. Date of CrossCheck: December 28, 2022. Date of online publication: March 27, 2023.

This work was supported by Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows (No. BX202210).

S. Li, D. Cao (corresponding author), W. Hu, and Q. Huang are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu, China, and Q. Huang is also with Southwest University of Science and Technology, Mianyang, China (e-mail: sichenli@std. uestc. edu. cn; caodi@uestc. edu. cn; whu@uestc. edu. cn; hwong@uestc.edu.cn).

Z. Chen and F. Blaabjerg are with the Department of Energy Technology, Aalborg University, DK-9220 Aalborg, Denmark (e-mail: zch@et. aau. dk; fbl@et. aau.dk).

DOI: 10.35833/MPCE.2022.000473

ized approaches to MG energy management. However, given the requirement of concurrently processing large amounts of data at a single location, centralized control is not a plug-and-play approach, which is required in an MG setup [17]. Generally, centralized control is useful in standalone power systems that must maintain a critical supply and demand balance over the long term in a slow-changing infrastructure [17]. In addition, MGs in a DN may be owned by different parties, and energy management within each MG may be determined by specific policies and economic rules [18]. This may limit information exchange between the distribution system operator and MGs and between MMGs owing to privacy and security concerns.

Overall, centralized approaches are unsuitable for energy scheduling in interconnected MMG systems. Instead, distributed and decentralized management is a major trend for such systems [1]. A distributed stochastic optimal scheduling scheme with minimal information exchange overhead is proposed in [19] for iterative energy scheduling of interconnected MMG systems, which are decomposed into MGs with local and reduced complexity. In [20], to measure the benefits of proactive resource trading within an interconnected MMG system, a distributed alternating direction method of multipliers is proposed to optimize synergistic operations of MMGs, thus determining the optimal solution in few iterations. The primal dual-multiplier method (PDMM) introduced in [21] outperforms the approach in [20] in terms of processing time and accuracy. Accordingly, the PDMM is applied in [22] to an interconnected MMG energy management system, showing the desired results. Other common algorithms such as Lagrangian relaxation algorithm [8] and consensus algorithm [23] have been successfully applied to distributed decision-making. A decentralized approach does not require each controller to establish communication channels with other controllers. Thus, some features of the distributed approach are preserved, and control is executed using only local information. A decentralized bilevel algorithm is used in [24] for energy management to coordinate the operation of interconnected MMGs within a distribution system. In [25], a decentralized two-stage approach is proposed for local energy trading in MMGs with an integrated pricing mechanism.

Although the abovementioned decentralized approaches have achieved promising results, energy management of interconnected systems still faces four main shortcomings.

1) Decentralized approaches are based on models, but deriving an accurate and efficient physical model of an interconnected MMG energy management system is difficult because the power flow and energy coupling relationships are influenced by numerous factors [26].

2) Model-based approaches depend on specific interconnected MMG environments. Hence, their generalization is limited when applied to a variety of MMG environments [1], [27].

3) By adding MGs and devices to the DN, the amount of data to be transmitted, processed, and stored in power systems increases rapidly. Therefore, the computational cost of energy management drastically increases when using conventional model-based approaches [1].

4) Model-based approaches fail to respond to continuously changing conditions and require continuous problem-solving, likely delaying real-time decisions. Hence, massive amounts of data should be used fully, accurately, and efficiently for proper energy management.

Among the available techniques to overcome the abovementioned drawbacks are data-driven such as model-free deep reinforcement learning. Such techniques can fully exploit information in interactions within the interconnected MMG environment to learn an optimal management policy without requiring accurate physical models of the interconnected MMGs. After learning is complete, the obtained policy can be used for end-to-end complex decision-making, such as instantly (e. g., within milliseconds) generating optimal actions in response to a real-time system state without any prior knowledge. In [1], deep reinforcement learning is applied to manage the energy sources of an MMG system, in which each MG is controlled by an agent through decentralized control. However, agents do not have a coordinated mechanism. Thus, each agent is independently optimized by its own reward function and lacks a coordinated relationship with other agents. However, such deployment mode does not fit an MMG system, in which energy interactions between MGs are intended to meet the energy needs of a single MG and realize energy complementarity within the system by integrating multiple MGs. Therefore, a high-quality coordinated optimization approach is essential for the efficient operation of MMG systems [6]. In [10], an approach based on multi-agent deep reinforcement learning (MADRL) is proposed to coordinately manage the energy resources of an MMG system. A peer-to-peer energy trading system for energy management of small-scale distributed energy resources is introduced in [28] based on MADRL. An MADRL-enabled demand response system is proposed in [29] to minimize electricity costs and improve grid reliability. These approaches establish a coordinated mechanism between controllable units (e. g., MGs). Nevertheless, they do not provide each unit a clear distinction between the impacts of other units' control strategies on its own electrical characteristics. In this situation, it may ignore the characteristics of energy conversion and transfer that are owned by different MGs and the voltage fluctuations caused by the frequent interactions of energy between MGs in the MMG system during the process of energy management, thereby causing certain hidden dangers to the stable operation of MG, ultimately resulting in the bad performance of the energy management of MMGs.

To improve energy management, we propose a novel MADRL-based approach called deep dense individual attention (D2IA) architecture in multi-agent soft actor critic (MASAC) algorithm to manage the energy of multi-energy interconnected MMG systems, aiming to minimize the operation cost while satisfying voltage limitations.

The main contributions of this study are summarized as follows.

1) A general energy network model is established. Along with the balance between energy supply and demand, a comprehensive electricity-thermal energy MG model is derived

to describe the internal structure and operation mechanism of each energy network.

2) A novel MADRL-based approach is proposed for a multi-energy interconnected MMG system, in which each MG controller is modeled as an intelligent agent. Instead of concatenating all the system information into a neural network, as in [10], [28], and [29], the proposed approach deploys individual attention mechanisms in each MG controller. In addition, it applies the deep dense architecture in reinforcement learning (D2RL) to enhance the nonlinear expression of the attention mechanism. This provides each MG with the ability to determine the degree of impact of other MG control strategies on its operation state.

3) Unlike model-based approaches, the proposed approach does not require optimization of the complex energy management problem of interconnected MMG systems in real time. Using the proposed approach, the MG controllers can build decision-making functions offline and deploy them online to select the optimal decision based on the latest system state data in a fully decentralized manner.

4) We evaluate the performance of the proposed approach against various benchmark approaches using real-world historical data.

The remainder of this paper is organized as follows. Section II introduces the mathematical model of the interconnected MMG system and formulates the optimization problem as a Markov game. In Section III, the proposed approach to solve MG energy management is detailed. To highlight the effectiveness of the proposed approach, Section IV reports extensive simulation and comparative results. Finally, Section V concludes this paper.

II. Mathematical Model of Interconnected MMG System

A. Architecture of Interconnected MMG System

The typical structure of an interconnected MMG system with combined heat and power (CHP) is shown in Fig. 1. Each MG establishes heat and electricity networks, where the heat network consists of a micro-turbine, a power-to-heat (PtH) unit (e. g., heat pump), and thermal load, while the electricity network consists of a diesel generator (DG), RES (e.g., solar, wind power), ESS, and electrical load. The power line (black lines in Fig. 1) and heat pipe (red lines) integrate the individual MGs into an interconnected MMG system, such that electrical and thermal energy can be supplied between MGs.

The individual MGs shown in Fig. 1 are autonomous systems that are geographically dispersed and equipped with RES, energy storage, and DG units to sustain local loads. Energy management at the individual MG level is intended to handle time-varying issues associated with RESs and loads using the complementary characteristics of the ESS, DG, and RESs. Individual MGs are often not intended to manage geographically scattered RESs and loads of a region but rather to manage a local area. Consequently, overall energy management may not be the globally optimal for that region [6], [30]. As shown in Fig. 1, MGs can be linked to form an interconnected MMG system using power lines and heat pipes. In an interconnected MMG system, energy management is aimed to handle spatial variations in RESs and loads within a region or wider area by coordinating the exchange of power between MGs and trading energy between the DN and individual MGs [6].

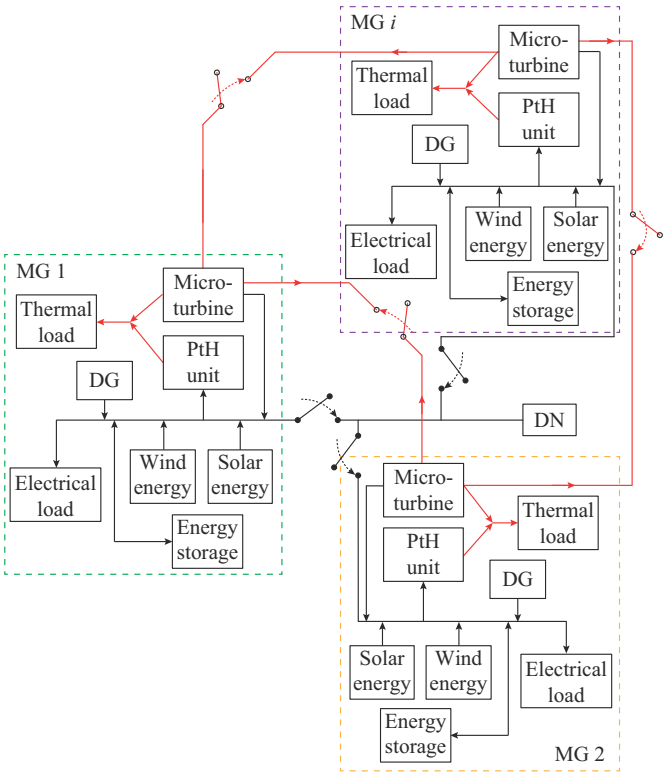

Fig. 1. Typical architecture of interconnected MMG system with CHP.

This paper considers a decentralized energy management system that controls the operation of an interconnected MMG system, in which each MG has its own controller. Each MG controller gathers two types of information from the local area to aid in decision-making. The forecasted values of wind power, solar power, electricity prices, and thermal and electrical loads of a local area, etc., are of the first type. The second type comes from real-time monitoring of system conditions such as time information, state of charge (SOC) of the ESS, and voltage at each node in the local area. The energy management system analyzes the second type of information and controls the power flow to maintain the security of the MMG system while maximizing the economic benefits.

B. Optimization Model

We formulate the energy management problem of the interconnected MMG system as a Markov game that can be solved using MADRL. To model the Markov game, various components should be described, as detailed below.

1) State. State variables reflect the current operating conditions of the individual MG and are gathered by each MG controller in the energy management system for decision-making. The state of the g^{th} MG at time t, $s_t^g \in S^g$, is defined as:

$$s_t^g=\{t,\widetilde{EP}_t,SOC_{i,t}^g,\widetilde{TL}_t^g,\widetilde{EL}_t^g,P_{i,t-1}^{g,DG},\widetilde{P}_{i,t}^{g,WT},\widetilde{P}_{i,t}^{g,PV},\widetilde{NL}_t^g,NV_t^g\} \quad (1)$$

where $\widetilde{EP}_t$ is the forecasted electricity price at time t; $SOC_{i,t}^g$ is the SOC of the ESS in the g^{th} MG at node i; $\widetilde{TL}_t^g$ is the forecasted cumulative thermal load of each node; $\widetilde{EL}_t^g$ is the forecasted cumulative electrical load of each node; $P_{i,t-1}^{g,DG}$ is the operation power of the DG at time $t-1$; $\widetilde{P}_{i,t}^{g,WT}$ and $\widetilde{P}_{i,t}^{g,PV}$ are the forecasted power values of the wind turbine (WT) and photovoltaic (PV) system, respectively; $\widetilde{NL}_t^g$ is the forecasted electrical load of each node in the g^{th} MG; and NV_t^g is the voltage of each node.

2) Action. In the Markov game, the action of the g^{th} MG is defined as:

$$a_t^g=\{P_{i,t}^{g,ESS},P_{i,t}^{g,DG},P_{i,t}^{g,PtH},Q_{i,t}^{g,PV},Q_{i,t}^{g,ESS},Q_{i,t}^{g,DG},Q_{i,t}^{g,WT}\} \quad (2)$$

where $P_{i,t}^{g,ESS}$, $P_{i,t}^{g,DG}$, and $P_{i,t}^{g,PtH}$ are the active power values of the ESS, DG, and PtH units, respectively; and $Q_{i,t}^{g,PV}$, $Q_{i,t}^{g,ESS}$, $Q_{i,t}^{g,DG}$, and $Q_{i,t}^{g,WT}$ are the reactive power values of the PV system, ESS, DG, and WT, respectively.

3) Reward function. The reward function in the Markov game is the objective function for optimization. It can be defined as:

$$r_t^g=-\sum_{g=1}^{G}(C_t^{EP}+C_t^{g,DG}+C_t^{g,Gas}+C_t^{g,Voltage}) \quad (3)$$

where G is the number of MGs; C_t^{EP} is the cost of buying electricity from the DN; $C_t^{g,DG}$ is the cost of the DG; $C_t^{g,Gas}$ is the cost of buying gas; and $C_t^{g,Voltage}$ is the cost of unstable voltage [31], [32]. These costs are given by:

$$C_t^{EP}=\sigma P_t^{DN}\cdot EP_t\cdot\Delta t \quad (4)$$

$$C_t^{g,DG}=C_t^{g,Fuel}+C_t^{g,Operation} \quad (5)$$

$$C_t^{g,Fuel}=\sum_{i=1}^{I}s_{i,t}^{DG}(a(P_{i,t}^{g,DG})^2+bP_{i,t}^{g,DG}+c)\Delta t \quad (6)$$

$$C_t^{g,Operation}=\sum_{i=1}^{I}C_{i,t}^{g,start/stop}\left|s_{i,t}^{g,DG}-s_{i,t-\Delta t}^{g,DG}\right| \quad (7)$$

$$C_t^{g,Gas}=\sum_{i=1}^{I}\frac{H_{i,t}^{g,CHP}\Delta t}{\eta_{CHP}L_{cvng}}\cdot GP_t \quad (8)$$

$$C_t^{g,Voltage}=\sum_{i=1}^{I}C_{i,t}^{g,Voltage} \quad (9)$$

$$C_{i,t}^{g,Voltage}=\begin{cases}2.5-50\left|1-v_{i,t}^g\right| & v_{i,t}^g\in(0.95,1.05)\\ -50\left|1-v_{i,t}^g\right| & v_{i,t}^g\in[0.8,0.95)\cup v_{i,t}^g\in(1.05,1.25]\\ -500 & \text{otherwise}\end{cases} \quad (10)$$

where EP_t is the electricity price at time t; σ is the motivating factor, which is similar to the one defined in [33] and used to balance the importance of the economic reward in the total reward; P_t^{DN} is the power exchange between the MMG system and DN at time t; $C_t^{g,Fuel}$ and $C_t^{g,Operation}$ are the fuel and operation costs of the DG in the g^{th} MG, respectively; a, b, and c are the coefficients of the DG fuel cost function; $C_{i,t}^{g,start/stop}$ is the start/stop cost of the DG in node i of the g^{th} MG at time t; $s_{i,t}^{g,DG}$ is the operation state of the DG; $H_{i,t}^{g,CHP}$ denotes the thermal output of CHP; η_{CHP} is the energy generation efficiency; L_{cvng} is the low calorific value of natural gas; GP_t is the gas price; $v_{i,t}^g$ is the node voltage; and I is the number of nodes in the interconnected MMG system.

4) Transition function. The transition function $s_{t+1}^g=\mathcal{T}(s_t^g,a_t^g,\mathcal{S})$ maps current state s_t^g to next state s_{t+1}^g according to action a_t^g and the randomness from the environment $\mathcal{S}$. Action a_t^g mainly alters the deterministic elements in s_t^g. For example, $SOC_{t+1}^g=SOC_t^g+a_t^g\Delta t/E_{\max}^g$, where $E_{\max}^g$ is the ESS capacity of the g^{th} MG. For $\widetilde{EP}_t$, $\widetilde{TL}_t^g$, $\widetilde{EL}_t^g$, $\widetilde{P}_{i,t}^{g,WT}$, $\widetilde{P}_{i,t}^{g,PV}$, and $\widetilde{NL}_t^g$ in s_t^g, the state transition is influenced by $\mathcal{S}$.

C. Constraints

The constraints are divided into those related to heat and power. Most heat networks comprise one source and several intermediate/load nodes interconnected via supply and return pipelines. Thus, the primary and secondary networks are the generation and consumption networks, respectively. Transporting thermal energy from the source node to intermediate/load nodes comprises two steps: ① the source node generates thermal energy; ② thermal energy is transported through the water supply pipe of the primary network to the thermal exchanger of the secondary network at the beginning. After entering the thermal exchanger of the secondary network, the thermal energy is transferred to all the intermediate/load nodes, and the water temperature changes drastically through this process. The water returns to the system through return pipelines to complete thermal circulation [34].

1) Loop pressure equation

The head loss is the change in pressure caused by pipe friction. In accordance with the loop pressure equation, the sum of all the head losses around a closed loop must equal zero. The loop pressure for the entire hydraulic network is given by [35]:

$$\boldsymbol{B}\boldsymbol{h}_f=\boldsymbol{0} \quad (11)$$

where $\boldsymbol{B}$ is the loop incidence matrix that relates loops to branches; and $\boldsymbol{h}_f$ is the vector of head losses.

2) Nodal flow balance

Thermal flow follows Kirchhoff's law. At each node, the amount of inflowing water is equal to that of outflowing water.

$$\sum_{p\in P_{s,i}^{s/r}}m_{p,t}^{s/r}=\sum_{p\in P_{e,i}^{s/r}}m_{p,t}^{s/r} \quad (12)$$

where $P_{s,i}^{s/r}$ and $P_{e,i}^{s/r}$ are the sets of pipes starting and ending at node i in supply/return network, respectively; and $m_{p,t}^{s/r}$ is the mass flow rate of supply/return pipe p at time t.

3) Thermal energy balance

The generation and consumption of thermal energy must be balanced. In a thermal system, considering that thermal energy is transferred between the heat source and load nodes through hot and cold water via a thermal pipe, the required thermal energy equals the difference in energy between the beginning of the supply pipe and the end of the return pipe [34].

$$H_{p,t}=C_{WH}(m_{p,t}^{s/r}T_{s,p,t}^{s}-m_{p,t}^{s/r}T_{e,p,t}^{r})\quad p\in N_{SN} \quad (13)$$

where $H_{p,t}$ is the thermal demand of pipe p at time t; C_{WH} is

water heat capacity; N_{SN} is the set of source nodes; and $T_{s,p,t}^{s/r}$ and $T_{e,p,t}^{s/r}$ are the temperatures at the start and end of supply/return pipe p at time t, respectively.

The following constraints related to the power component must be satisfied during optimization of energy management.

$$\sum_k I_{i,k} P_{k,t} - P_{i,t}^{Load} = V_{i,t} \sum_{j=1}^{J} V_{j,t} \left(G_{ij,t} \cos\theta_{ij,t} + B_{ij,t} \sin\theta_{ij,t}\right) \quad (14)$$

$$\sum_s I_{i,s} Q_{s,t} - Q_{i,t}^{Load} = V_{i,t} \sum_{j=1}^{J} V_{j,t} \left(G_{ij,t} \sin\theta_{ij,t} - B_{ij,t} \cos\theta_{ij,t}\right) \quad (15)$$

$$V_{i,\min} \le V_{i,t} \le V_{i,\max} \quad (16)$$

$$0 \le P_{i,t}^{PV} \le P_{i,\max}^{PV} \quad (17)$$

$$0 \le P_{i,t}^{WT} \le P_{i,\max}^{WT} \quad (18)$$

$$0 \le P_{i,t}^{PtH} \le P_{i,\max}^{PtH} \quad (19)$$

$$P_{i,t}^{DG} - P_{i,t-\Delta t}^{DG} \le R_{up}^{DG} \Delta t s_{t-\Delta t}^{DG} + P_{i,\min}^{DG} (s_t^{DG} - s_{t-\Delta t}^{DG}) + P_{i,\max}^{DG} (1 - s_t^{DG}) \quad (20)$$

$$P_{i,t-\Delta t}^{DG} - P_{i,t}^{DG} \le R_{dn}^{DG} \Delta t s_t^{DG} + P_{i,\min}^{DG} (s_{t-\Delta t}^{DG} - s_t^{DG}) + P_{i,\max}^{DG} (1 - s_{t-\Delta t}^{DG}) \quad (21)$$

$$-P_{i,\max}^{ESS} \le P_{i,t}^{ESS} \le P_{i,\max}^{ESS} \quad (22)$$

$$SOC_{i,\min}^{ESS} \le SOC_{i,t}^{ESS} \le SOC_{i,\max}^{ESS} \quad (23)$$

$$(P_{i,t}^{PV})^2 + (Q_{i,t}^{PV})^2 \le (S_i^{PV})^2 \quad (24)$$

$$(P_{i,t}^{WT})^2 + (Q_{i,t}^{WT})^2 \le (S_i^{WT})^2 \quad (25)$$

$$(P_{i,t}^{DG})^2 + (Q_{i,t}^{DG})^2 \le (S_i^{DG})^2 \quad (26)$$

$$(P_{i,t}^{ESS})^2 + (Q_{i,t}^{ESS})^2 \le (S_i^{ESS})^2 \quad (27)$$

where $k \in \{PV, WT, PtH, DG, ESS\}$; $I_{i,k}$ is an element in the generator incidence matrix, and when generator k is connected to node i, $I_{i,k}$=1; $P_{k,t}$ is the active power of generator k at time t; $P_{i,t}^{Load}$ is the load demand of node i; J is the number of nodes in the MG; $V_{i,t}$ is the voltage amplitude of node i; $G_{ij,t}$ and $B_{ij,t}$ are the real and imaginary parts of the admittance element between nodes i and j, respectively; $\theta_{ij,t}$ is the voltage phase difference between nodes i and j; $s \in \{PV, WT, DG, ESS\}$; $Q_{s,t}$ is the reactive power of node s at time t; $Q_{i,t}^{Load}$ is the reactive power of node i; $V_{i,\min}$ and $V_{i,\max}$ are the lower and upper voltage limits of node i, respectively; $P_{i,t}^{PV}$, $P_{i,t}^{WT}$, $P_{i,t}^{PtH}$, $P_{i,t}^{DG}$, and $P_{i,t}^{ESS}$ are the active power values of the PV system, WT, PtH unit, DG, and ESS of node i, respectively; R_{up}^{DG} and R_{dn}^{DG} are the ramp-up and ramp-down limits of the DG, respectively; s_t^{DG} is a binary variable that represents the operation state of the DG at time t (1, on; 0, off); $P_{i,\max}^{PV}$, $P_{i,\max}^{WT}$, $P_{i,\max}^{PtH}$, $P_{i,\max}^{DG}$, and $P_{i,\max}^{ESS}$ are the upper limits of the active power for the PV system, WT, PtH unit, DG, and ESS of node i, respectively; $SOC_{i,t}^{ESS}$ is the SOC of the ESS of node i; $SOC_{i,\min}^{ESS}$ and $SOC_{i,\max}^{ESS}$ are the lower and upper limits of the SOC of ESS, respectively; $Q_{i,t}^{PV}$, $Q_{i,t}^{WT}$, $Q_{i,t}^{DG}$, and $Q_{i,t}^{ESS}$ are the reactive power values of the PV system, WT, DG, and ESS, respectively; and S_i^{PV}, S_i^{WT}, S_i^{DG}, and S_i^{ESS} are the apparent power values of the PV system, WT, DG, and ESS of node i, respectively.

Equations (14) and (15) establish AC power flow constraints, while (16) is the voltage amplitude constraint of each node. Formulas (17) - (22) are operation power constraints, while (23) is the SOC constraint for the ESS, and (24)-(27) ensure that the reactive power generated at the inverter of the PV system, WT, DG, and ESS does not exceed the available capacity.

III. Proposed Approach to Solve MG Energy Management

Conventional reinforcement learning algorithms can not suitably deal with Markov games owing to the curse of dimensionality and lack of coordination between agents while considering privacy and information security. To solve this problem, we propose an MADRL-based approach for energy management of complex interconnected MMGs. Our MADRL-based approach called D2IA-MASAC features centralized training and decentralized execution to ensure that each MG has autonomous energy management within an interconnected MMG system. The proposed approach consists of two neural network stages. The actor stage is responsible for making decisions, and the critic stage is responsible for guiding the actor stage to approximate the optimal policy [36].

1) MASAC Overview

The MASAC algorithm is a multi-agent variant [37] of the soft actor-critic (SAC) algorithm [38]. Each MG controller is modeled as an SAC agent within the centralized training framework to manage the energy of the interconnected MMG system.

Considering N agents with parameterized critic network $Q_{\phi_j}^j, j \in [1, N]$, the update of the individual critic network for the j^{th} agent can be defined as:

$$Loss = \mathbb{E}[(Q_{\phi_j}^j(\boldsymbol{X}_t, \boldsymbol{A}_t) - y)^2] \quad (28)$$

$$y = r_t^j + \gamma \mathbb{E}[Q_{\phi_j'}^j(\boldsymbol{X}_{t+1}, \boldsymbol{A}_{t+1}) - \alpha \ln(\mu_{\theta_j'}^j(s_{t+1}^j))] \quad (29)$$

where *Loss* denotes the optimization objective; $\boldsymbol{X}_t = [s_t^1, s_t^2, ..., s_t^N]$ is the state concatenation at time t; $\boldsymbol{A}_t = [a_t^1, a_t^2, ..., a_t^N]$ is the action concatenation at time t; $Q_{\phi_j}^j(\boldsymbol{X}_t, \boldsymbol{A}_t)$ and $Q_{\phi_j'}^j(\boldsymbol{X}_{t+1}, \boldsymbol{A}_{t+1})$ are the centralized action-value functions with parameter ϕ_j and ϕ_j', respectively; γ is the reward discount factor; α is the temperature ratio used to balance exploration and exploitation; and $\mu_{\theta_j'}^j$ denotes the target actor network with parameter θ_j'.

N agents are parameterized by actor network $\mu_{\theta_j}^j, j \in [1, N]$, and the individual policies are updated by ascent according to the following gradient:

$$\nabla J(\mu_{\theta_j}^j) = \mathbb{E}[\nabla_{\theta_j} \ln(\mu_{\theta_j}^j(s_t^j))(Q_{\phi_j}^j(\boldsymbol{X}_t, \boldsymbol{A}_t) - \alpha \ln(\mu_{\theta_j}^j(s_t^j)))] \quad (30)$$

Parameters ϕ_j' and θ_j' can be updated as:

$$\begin{cases} \phi_j' \leftarrow \tau\phi_j + (1-\tau)\phi_j' \\ \theta_j' \leftarrow \tau\theta_j + (1-\tau)\theta_j' \end{cases} \quad (31)$$

where the soft replacement parameter $\tau \ll 1$.

2) Architectures of D2IA-MASAC Networks

The architecture of the critic network used in the proposed approach is shown in Fig. 2. Each agent has its critic, which is composed of D2RL [39] and attention mechanism

[40]. Considering the first agent in Fig. 2 as an example, the input of the critic is the agent state and actions. Each agent information (i.e., states and actions) is encoded by the D2RL.

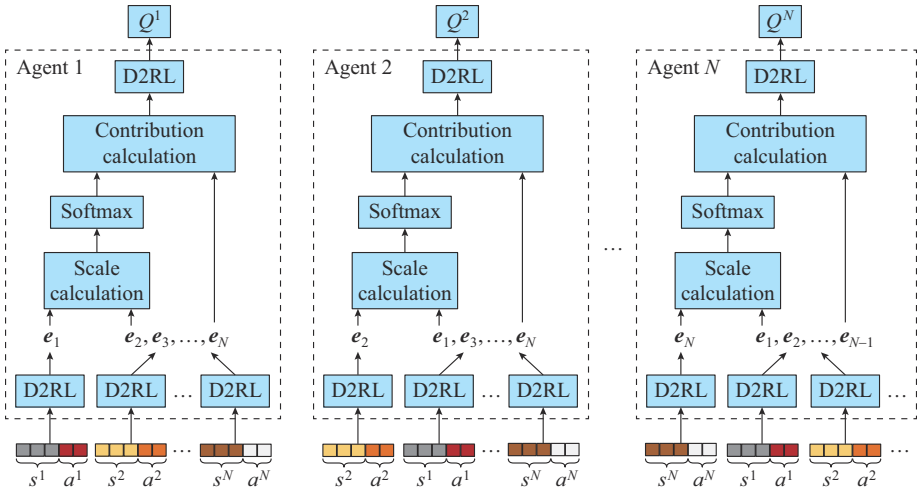

Fig. 2. Architecture of critic network used in proposed approach.

Deep neural network is a powerful approach for extracting features from inputs and mapping them to the desired outputs. Hence, it can enhance the nonlinear capability of the critic network. Merely increasing the depth of a dense neural network may be ineffective because the original input information gradually vanishes as the network deepens [41]. The D2RL can prevent this problem. Its architecture is shown in Fig. 3(a). Input data i are first processed by dense neural network h_1 to obtain output m_1. The input of h_2 is the concatenation of i and m_1 instead of m_1 alone, as shown in Fig. 3(b). A loop is established until obtaining output O. This strategy leverages the nonlinear mapping ability of deep dense neural networks by incorporating original input i into each hidden layer of the network, such that the information of input i is retained. The calculation from the input of the critic network belonging to the first agent of $\boldsymbol{e}_1$ is given by:

$$\begin{cases} \boldsymbol{I}_1 = concat(s^1, s^2, ..., s^N, a^1, a^2, ..., a^N) \\ \boldsymbol{I}_{u+1} = ReLu(concat(\boldsymbol{I}_1, \boldsymbol{I}_u)\cdot \boldsymbol{W}_u^1 + \boldsymbol{B}_u^1) \quad u = 2, 3, \ldots, U-1 \\ \boldsymbol{\zeta} = ReLu(\boldsymbol{I}_U \boldsymbol{W}_U^1 + \boldsymbol{B}_U^1) \\ \boldsymbol{e}_1 = \boldsymbol{\zeta} \boldsymbol{W}_O^1 + \boldsymbol{B}_O^1 \end{cases} \tag{32}$$

where $concat(\cdot)$ denotes the concatenate operation function; $ReLu(\cdot)$ denotes the rectified linear unit activation function; I_u is the input of the u^{th} hidden layer; $\boldsymbol{W}_u^1$ and $\boldsymbol{B}_u^1$ are the weight and bias matrices of the u^{th} hidden layer of the first agent, respectively; U is the number of hidden layers; $\boldsymbol{\zeta}$ denotes the latent feature extracted by the U hidden layers from input $\boldsymbol{I}_1$; and $\boldsymbol{W}_O^1$ and $\boldsymbol{B}_O^1$ are the weight and bias matrices of the output layer of the first agent, respectively.

The calculations for $\boldsymbol{e}_i, i \in [1, N]$ are processed by the attention mechanism to emphasize representative information of other agents while discarding irrelevant details for decision-making. The detailed calculation process is given as (33)-(35).

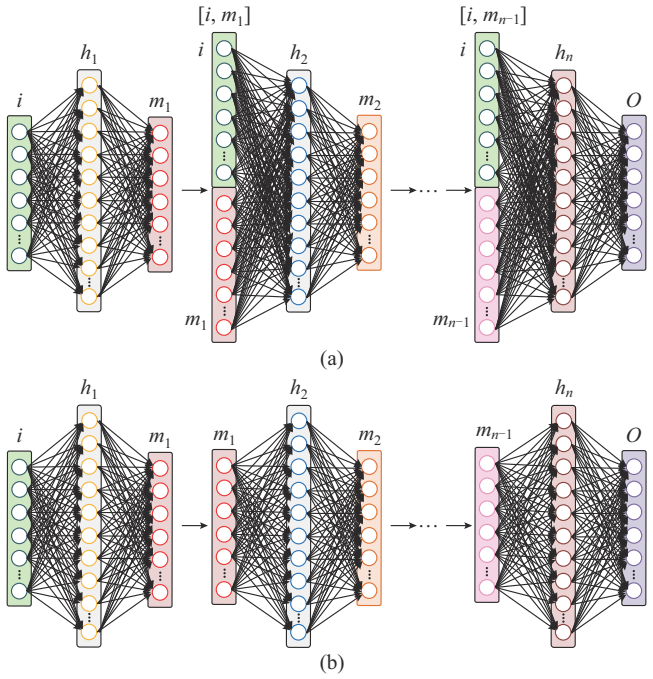

Fig. 3. Architectures of D2RL and conventional dense neural network. (a) D2RL. (b) Conventional dense neural.

$$\omega_i = \frac{\exp(\chi_{1(i)})}{\sum_{k=2}^{N} \exp(\chi_{1(k)})} \quad i \in [2, 3, ..., N] \tag{33}$$

$$[\chi_{1(2)}, \chi_{1(3)}, \ldots, \chi_{1(N)}] = \frac{\boldsymbol{\varpi}(\boldsymbol{e}_1, [\boldsymbol{e}_2^{\mathrm{T}}, \boldsymbol{e}_3^{\mathrm{T}}, ..., \boldsymbol{e}_N^{\mathrm{T}}])}{\sqrt{d_\beta}} \tag{34}$$

$$\boldsymbol{\varepsilon}_1 = \sum_{k=2}^{N} \omega_k \boldsymbol{e}_k \tag{35}$$

where $\boldsymbol{\varpi}(\boldsymbol{e}_1, [\boldsymbol{e}_2^{\mathrm{T}}, \boldsymbol{e}_3^{\mathrm{T}}, ..., \boldsymbol{e}_N^{\mathrm{T}}]) = [\boldsymbol{e}_1 \boldsymbol{e}_2^{\mathrm{T}}, \boldsymbol{e}_1 \boldsymbol{e}_3^{\mathrm{T}}, ..., \boldsymbol{e}_1 \boldsymbol{e}_N^{\mathrm{T}}]$; d_β is the network dimension; and $\boldsymbol{\varepsilon}_1$ denotes influence of other agents for the first agent.

The action-value function is expressed as:

$$Q^1(\boldsymbol{X},\boldsymbol{A})=f_{D2RL}(\boldsymbol{e}_1,\boldsymbol{\varepsilon}_1) \tag{36}$$

where f_{D2RL} denotes the deep-blue D2RL in the first agent. Training process of the proposed approach is described in Algorithm 1.

Algorithm 1: training process of D2IA-MASAC approach for energy management in interconnected MMG system

Initialize parameters and replay buffer

for episode from 1 to P **do**
  Randomly choose electricity price and PV and WT power data of the κ^{th} day
  Randomly generate thermal and electrical loads in range
  for t = 1 to T **do**
    for MG j = 1 to N **do**
      Sample action a_t^j from $\mu_{\theta_j}^j(s_t^j)$
      Input s_t^j and a_t^j to environment to obtain r_t^j and s_{t+1}^j
      Store transition $(s_t^j, a_t^j, r_t^j, s_{t+1}^j)$ in memory of the j^{th} MG controller
      Randomly sample batch-sized transitions from the j^{th} replay buffer
      Update ϕ_j, θ_j, ϕ_j', and θ_j' using (28)-(31)
    end for
  end for
end for

IV. Numerical Results

A. Experimental Setup

We evaluated the performance of the proposed D2IA-MASAC approach through numerical experiments on an interconnected MMG system, whose architecture is illustrated in Fig. 4. Therein, there are three MGs that each contain a heat network. Each colored region shown in Fig. 4 denotes a heat network.

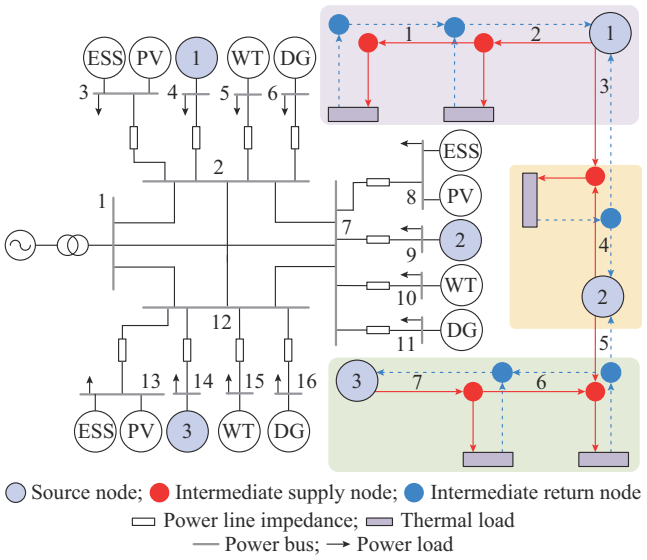

Fig. 4. Architecture of interconnected MMG system.

The parameters of lines and heat pipes in MMG system shown in Fig. 4 are listed in Tables I and II, respectively. The interconnected MMG system includes three MGs, each of which consists of a heat network and an electricity network. Each MG has the ESS, DG, WT, PV system, and micro-turbine (i.e., source node in Fig. 4). The parameters related to this case study are listed in Tables III and IV, and those of the proposed approach are listed in Table V. The electricity price offered by the DN follows the hourly pricing from PJM [42], and the natural gas prices follow the monthly Natural Gas Industrial Price from the US Energy Information Administration [43]. The training data contain the first 200 days of data, and the test data are from days 201 to 300 in 2017. The forecasted power of electrical and thermal loads, PV system, and WT of the three MGs in the interconnected MMG system on a typical day are shown in Fig. 5. The regions of MG1 largely include electrical loads and have a high RES capacity. There are fewer electrical loads in MG2 than in MG1, and the RES capacity of MG2 is lower. MG3 has the highest proportion of RESs among the three MGs. In addition, the electrical loads on MG3 are relatively low. The thermal loads of MG1 and MG2 are similar, and that of MG3 is the highest.

TABLE I
Parameters of Lines in MMG System

Line	From node	To node	Resistance (Ω)	Reactance (Ω)
1	1	2	0.0125	0.005
2	2	3	0.0375	0.015
3	2	4	0.0300	0.012
4	2	5	0.0225	0.009
5	2	6	0.0150	0.006
6	2	7	0.7500	0.300
7	2	12	0.7500	0.300
8	1	7	0.0125	0.005
9	7	8	0.0250	0.010
10	7	9	0.0200	0.008
11	7	10	0.0225	0.009
12	7	11	0.0275	0.011
13	7	12	0.7500	0.300
14	1	12	0.0125	0.005
15	12	13	0.0175	0.007
16	12	14	0.0200	0.008
17	12	15	0.0250	0.010
18	12	16	0.0250	0.010

TABLE II
Parameters of Heat Pipes in MMG System

Pipe	Length (m)	Diameter (m)
1	205	0.08
2	200	0.05
3	200	0.05
4	200	0.05
5	200	0.05
6	195	0.06
7	208	0.07

TABLE III
PARAMETERS OF SYSTEM MODEL

Symbol	Value	Symbol	Value
a (\$/((kW)^2·h))	1040	σ	6
b (\$/MWh)	30.4	η_{CHP}	0.36
c (\$)	1.3	L_{cvng} (MWh/m^3)	9.7×10^{-3}
$C^{g,start/stop}$ (\$)	5		

TABLE IV
PARAMETERS OF INTERCONNECTED MMG SYSTEM

MG	$P^{PV}_{i,\max}$ (MW)	$P^{WT}_{i,\max}$ (MW)	$P^{PtH}_{i,\max}$ (MW)	$P^{DG}_{i,\max}$ (MW)	$P^{DG}_{i,\min}$ (MW)	$P^{ES}_{i,\max}$ (MW)
MG1	0.20	0.25	0.005	0.05	0.02	0.050
MG2	0.15	0.20	0.005	0.05	0.02	0.045
MG3	0.10	0.15	0.005	0.05	0.02	0.040

MG	$E_{i,\max}$ (MWh)	$SOC^{ES}_{i,\max}$	$SOC^{ES}_{i,\min}$	R^{DG}_{up}	R^{DG}_{dn}
MG1	0.20	1	0.1	0.025	0.025
MG2	0.18	1	0.1	0.025	0.025
MG3	0.16	1	0.1	0.025	0.025

TABLE V
PARAMETERS OF PROPOSED APPROACH

Parameter	Value
Temperature ratio	10
Reward discount factor	0.95
Memory capacity	1×10^6
Learning rate of actor	3×10^{-4}
Learning rate of critic	3×10^{-4}
Soft replacement	1×10^{-3}
Batch size for updating	256

B. Comparisons with Benchmark Approaches

To assess the effectiveness of our proposal, we compared it with various benchmark approaches.

1) Evaluated Benchmark Approaches

1) TD3 [44]: MG controllers are modeled as TD3 agents, each of which is trained individually to maximize its own reward function without coordinating with other MGs.

2) MATD3 [10]: MATD3 is an approach to optimize the energy management of interconnected MMG systems. This is a multi-agent variant of TD3.

3) MAAC [45]: MAAC uses a single common attention network to collect information from all MGs in a centralized manner and constructs a common gradient space for all MG controllers to update the parameters of the neural network. The most notable difference between MAAC and MATD3 is that the former introduces a common attention network.

2) Performance on Training Set

Figure 6 shows the normalized cumulative rewards (i. e., values of the objective function) obtained using different reinforcement-learning-based approaches during the training process. The agents have no initial knowledge for decision-making to obtain a high reward.

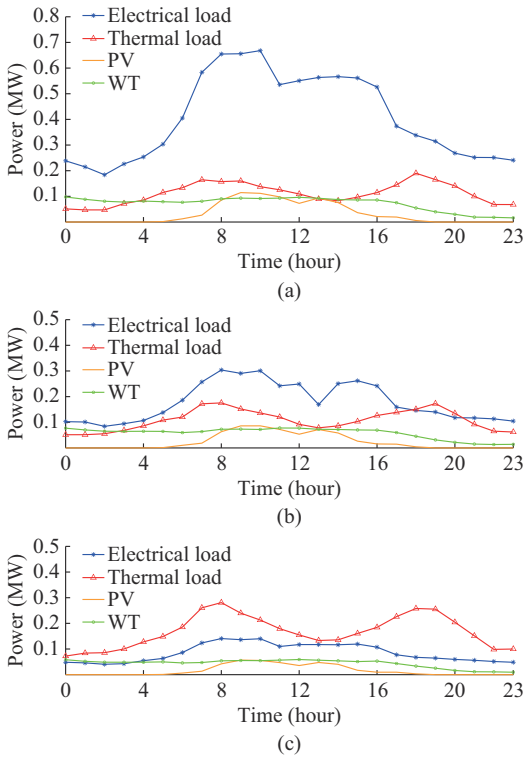

Fig. 5. Forecasted power of electrical and thermal loads, PV system, and WT of three MGs. (a) Data of MG1. (b) Data of MG2. (c) Data of MG3.

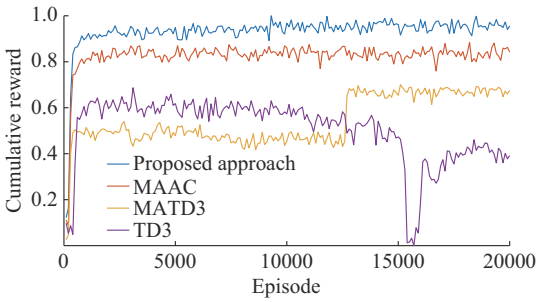

Fig. 6. Normalized cumulative rewards obtained using different reinforcement-learning-based approaches during training process.

Consequently, they explore the action space to gain experience. As training proceeds, the agents gradually learn the energy management strategy of the interconnected MMG system to increase the reward. The training curves increase until complete convergence, representing the end of training.

The reward curve of TD3 shows the lowest value and oscillates considerably during training compared with other approaches, demonstrating that a lack of coordination impedes a suitable energy management of an interconnected MMG system. MATD3 is a multi-agent variant of TD3. Compared with TD3, MATD3 uses the multi-agent framework inspired by the approach in [37] to construct a coordination relationship between each MG for energy management in an interconnected MMG system. Evidently, MATD3 outperforms TD3 in terms of training stability and performance owing to the coordination mechanism. The MAAC [45] further enhances the coordination between MGs by using a common attention network in the critic part, outperforming MATD3.

Using the common attention network in MAAC, the network parameters of each agent include part of the network parameters shared among agents, and each agent has an attention attribute to accelerate consensus among agents and thus enhance their coordination. However, the attention network in MAAC does not seem to fit the optimization of the energy management of interconnected MMG systems because such systems have complex coupling characteristics between MGs in the energy network with combined heat and power, energy exchange among MMGs, and energy trading between the MMG and DN. In an interconnected MMG system, each MG has its system parameters (e.g., thermal and electrical loads, rated power of devices, heat and power flow), and each MG may have negative effects on others through a common network owing to the complex and individual operation. Hence, the proposed approach uses the network architecture described in Section III, and its training outperforms that of MAAC, indicating the effectiveness of our proposal.

3) Performance on Test Set

To further demonstrate the effectiveness of the proposed approach, we select stochastic optimization approaches for comparison. We performed the comparisons using the NLopt library [46].

For optimization of deterministic information scenario with NLopt, uncertain information about electricity prices, electrical and thermal loads, solar and wind power, and other factors is assumed to be known beforehand. NLopt is then used to solve a deterministic optimization problem based on global information. Note that this approach is infeasible in a realistic scenario owing to the randomness of electricity prices, electrical load, etc.

Table VI lists the average cost of the proposed and benchmark approaches over 100 test days.

TABLE VI
AVERAGE COST OF PROPOSED AND BENCHMARK APPROACHES

Approach	Average cost ($/day)	Percentage (%)
TD3	1870.14	565.00
MATD3	1145.15	345.97
MAAC	420.58	127.06
Proposed	331.00	100.00
NLopt	86.11	26.02

The percentage represents the average cost of the benchmark approaches divided by that of the proposed approach. The interconnected MMG system spends 465.00%, 245.97%, and 27.06% more under control of TD3, MATD3, and MAAC compared with the proposed approach, respectively. TD3 has no mechanism for coordination between agents and provides the worst results. Hence, coordination among agents are necessary in an MMG system. The results of MATD3, which are better than those of TD3, further demonstrate this point. The proposed approach is superior to MAAC, which has a common attention layer. Therefore, using an individual attention mechanism to manage the energy of an interconnected MMG system is more suitable than using a common attention mechanism. Compared with other approaches, our proposal achieves the closest performance to that of NLopt, confirming its effectiveness.

C. Evaluation on Test Set Over a Day

To show the effectiveness of the proposed approach in guaranteeing secure operation, the voltage of each bus for a day in the interconnected MMG system is shown in Fig. 7. The red, blue, and green lines indicate the voltage profiles of MG1, MG2, and MG3, respectively, while marks *, Δ, ∇, and + indicate the voltage profiles at nodes 1, 2, 3, and 4, respectively. The shaded part indicates the area where the voltage exceeds the limit. The voltage is constrained within a safe range at each bus and is more stable using our proposal than the rule-based approach, in which the rule-based approach follows: ESS, DG, and PtH units do not operate, and each controllable reactive power unit has the reactive power associated with its power factor and active power. These results suggest that the proposed approach can ensure the safe operation of interconnected MMG systems by providing relevant solutions.

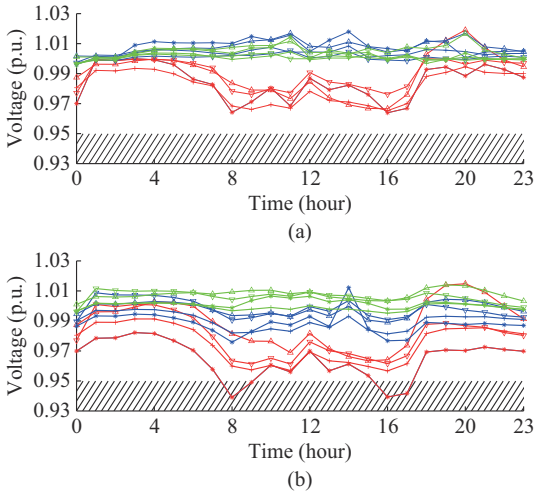

Fig. 7. Voltage profile of each bus for a day in interconnected MMG system. (a) Voltage profile with proposed approach. (b) Voltage profile with rule-based approach.

The energy management result of the proposed approach on a test day is depicted in Fig. 8 to further illustrate the effectiveness of the proposed approach.

MG1 has the lowest RES proportion and the largest fluctuation range in its voltage profile compared with the other two MGs, which has the most severe undervoltage, as shown in Fig. 7(b). Therefore, MG1 has a different charging mode from MG2 and MG3, as shown in Fig. 8(b). Specifically, MG1 has a low voltage early in the day. To increase the voltage, the ESSs of MG1 discharge their power to avoid further voltage drops. On the contrary, MG3 has the highest RES penetration, which can be used to supply the electrical load to stabilize the voltage. Hence, its ESS can be used for energy arbitrage, that is, charging when the electricity price is low and discharging when the electricity price is high.

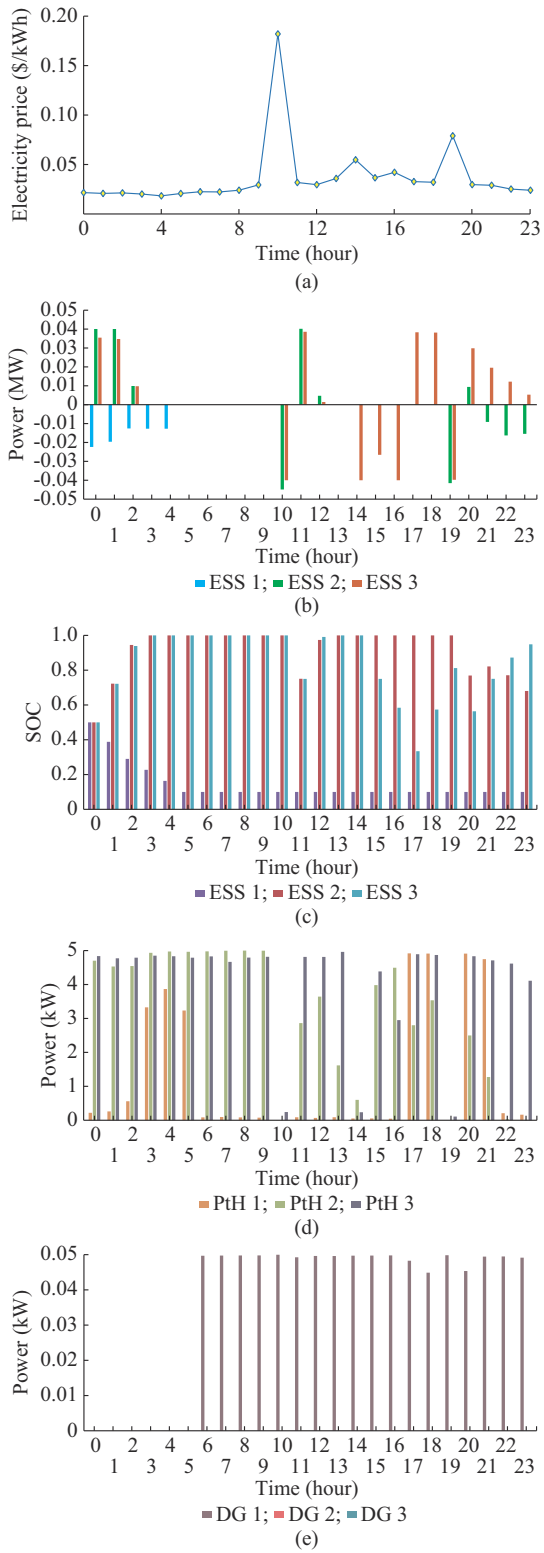

Fig. 8. Energy management results of proposed approach on a test day. (a) Real-time electricity price. (b) Charging/discharging power of ESSs in interconnected MMG system. (c) Energy of ESSs in interconnected MMG system. (d) Power of PtH units in interconnected MMG system. (e) Power of DGs in interconnected MMG system.

Similarly, the ESS of MG2 performs energy arbitrage, but it behaves more conservatively than that of MG3. Specifically, as shown in Fig. 8(c), the discharging power of MG3 is larger than that of MG2, resulting in the lowest SOC of MG3 at the 17th hour being lower than that of MG2 at the 23rd hour. As MG2 has a lower RES penetration than MG3, if the ESS has a low SOC, the ESS may fail to supply enough energy to the electrical load, thus leading to undervoltage in MG2. Therefore, sufficient energy is required for the ESS of MG2 to stabilize the voltage profile. In other words, voltage stability is more important in MG2 than that in MG3 for the charging/discharging management of the ESS. The period from the 21st hour to the 23rd hour is the time of the day when electricity prices are low, as shown in Fig. 8(a). During this period, the charging power of MG3 is more reasonable when analyzed from the energy arbitrage perspective, as shown in Fig. 8(b). However, because MG2 has a lower RES penetration than MG3, MG2 needs an additional energy supply from the ESS from the 21st hour to the 23rd hour to stabilize the voltage.

Figure 8(d) shows the operation of the PtH units, which convert electrical energy into thermal energy. In the heat network, the thermal demand is satisfied by the micro-turbine and PtH units. Because the PtH units consume power, like the ESS, their power profile follows a rule similar to that of the ESS.

As shown in Fig. 8(e), the DG of MG1 does not operate from the 0th hour to the 5th hour and operates at almost maximum power from the 6th hour to the 23rd hour. This is because from the 0th hour to the 5th hour, the electrical load is low, and the voltage of MG1 is relatively stable. To reduce the DG operation and related fuel costs, the DG should not work when the voltage is stable. After the 5th hour, as people start their activities, the electricity consumption gradually increases, and the voltage fluctuates, as shown in Fig. 7(b), thus requiring DG operation to compensate for the fluctuations. The DG of MG1 operates at almost maximum power from the 6th hour to the 23rd hour given the severe undervoltage during this period, as shown in Fig. 7(b). As the voltages of MG2 and MG3 are more stable after regulation, their DGs do not operate to prevent the related costs.

As shown in Fig. 9, because MG1 has the lowest RES penetration and largest voltage fluctuations compared with the other two MGs, MG2 and MG3 transfer their energy to MG1 through the power line to stabilize its voltage profile.

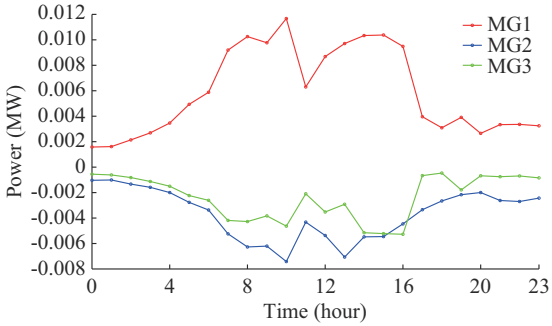

Fig. 9. Power transference between MGs.

V. Conclusion

Energy management at the MMG system level focuses on scheduling power exchange between MGs and power trades

with the DN to minimize the total cost. In this paper, a decentralized multi-energy management approach is proposed to achieve optimal synergies between multiple MGs. The approach encourages active resource transactions between interconnected MGs by optimizing the reward function to leverage diverse multi-energy generation/loads to synergize multiple MGs. The coordinated optimization of the interconnected MMG system is modeled as a Markov game and solved using the proposed approach. The proposed approach is decentralized and does not require information sharing between MGs to protect privacy and guarantee resource autonomy.

Simulations and comparative results show that: ① compared with benchmark approaches, the proposed approach is the most stable during training and achieves the best training results; ② on the test set, the proposed approach can achieve the best energy management for the interconnected MMG system, demonstrating its generalization ability; and ③ the interpretability of the operation of different MG assets and the power transfer between MGs further confirm the effectiveness of the proposed approach. At the MMG system level, suitably managing its internal energy can not only lower the cost of the system but also help keep the voltage of each node stable through an effective management of how MGs exchange energy with each other. In addition, we can determine from these data that the proposed approach allows MG controllers to distinguish the degree to which the control strategies of other MG controllers influence their working states, thus providing the MG controller with the capability to effectively manage multiple energies, making the proposed approach more economical and safer than other benchmark approaches.

REFERENCES

[1] C. Guo, X. Wang, Y. Zheng *et al*., "Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning," *International Journal of Electrical Power & Energy Systems*, vol. 131, p. 107048, Oct. 2021.

[2] D. E. Olivares, A. Mehrizi-Sani, A. H. Etemadi *et al*., "Trends in microgrid control," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1905-1919, Jul. 2014.

[3] S. Fang, T. Zhao, Y. Xu *et al*., "Coordinated chance-constrained optimization of multi-energy microgrid system for balancing operation efficiency and quality-of-service," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 5, pp. 853-862, Sept. 2020.

[4] M. A. Mohamed, A. Hajjiah, K. A. Alnowibet *et al*., "A secured advanced management architecture in peer-to-peer energy trading for multi-microgrid in the stochastic environment," *IEEE Access*, vol. 9, pp. 92083-92100, Jun. 2021.

[5] H. Zou, J. Tao, S. K. Elsayed *et al*., "Stochastic multi-carrier energy management in the smart islands using reinforcement learning and unscented transform," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106988, Sept. 2021.

[6] B. Zhao, X. Wang, D. Lin *et al*., "Energy management of multiple microgrids based on a system of systems architecture," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6410-6421, Nov. 2018.

[7] B. Zhou, J. Zou, C. Y. Chung *et al*., "Multi-microgrid energy management systems: architecture, communication, and scheduling strategies," *Journal of Modern Power Systems and Clean Energy*, vol. 9, no. 3, pp. 463-476, May 2021.

[8] N. Liu, J. Wang, and L. Wang, "Hybrid energy sharing for multiple microgrids in an integrated heat-electricity energy system," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1139-1151, Jul. 2019.

[9] M. Fathi and H. Bevrani, "Statistical cooperative power dispatching in interconnected microgrids," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 3, pp. 586-593, Jul. 2013.

[10] T. Chen, S. Bu, X. Liu, *et al*., "Peer-to-peer energy trading and energy conversion in interconnected multi-energy microgrids using multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 715-727, Jan. 2022.

[11] H. Gao, J. Yang, S. He *et al*., "Decision-making method of sharing mode for multi-microgrid system considering risk and coordination cost," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 6, pp. 1690-1703, Nov. 2022.

[12] J. Wang, S. You, Y. Zong *et al*., "Optimal dispatch of combined heat and power plant in the integrated energy system: a state of the art review and case study of Copenhagen," *Energy Procedia*, vol. 158, pp. 2794-2799, Feb. 2019.

[13] Y. Guo and C. Zhao, "Islanding-aware robust energy management for microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1301-1309, Mar. 2018.

[14] N. Nikmehr and S. Najafi-Ravadanegh, "Probabilistic optimal power dispatch in multi-microgrids using heuristic algorithms," in *Proceedings of 2014 Smart Grid Conference (SGC)*, Tehran, Iran, Dec. 2014, pp. 1-6.

[15] N. Nikmehr and S. Najafi-Ravadanegh, "Reliability evaluation of multi-microgrids considering optimal operation of small scale energy zones under load-generation uncertainties," *International Journal of Electrical Power & Energy Systems*, vol. 78, pp. 80-87, Jun. 2016.

[16] X. Chen, C. Kang, M. O'Malley *et al.*, "Increasing the flexibility of combined heat and power for wind power integration in China: modeling and implications," *IEEE Transactions on Power Systems*, vol. 30, no. 4, pp. 1848-1857, Jul. 2015.

[17] Y. S. F. Eddy, H. B. Gooi, and S. X. Chen, "Multi-agent system for distributed management of microgrids," *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 24-34, Jan. 2015.

[18] M. N. Alam, S. Chakrabarti, and A. Ghosh, "Networked microgrids: state-of-the-art and future perspectives," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1238-1250, Mar. 2019.

[19] D. Xu, B. Zhou, K. Chan *et al.*, "Distributed multienergy coordination of multimicrogrids with biogas-solar-wind renewables," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3254-3266, Jun. 2019.

[20] D. Xu, B. Zhou, N. Liu *et al.*, "Peer-to-peer multienergy and communication resource trading for interconnected microgrids," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 4, pp. 2522-2533, Apr. 2021.

[21] T. W. Sherson, R. Heusdens, and W. B. Kleijn, "Derivation and analysis of the primal-dual method of multipliers based on monotone operator theory," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 2, pp. 334-347, Jun. 2019.

[22] M. A. Mohamed, T. Jin, and W. Su, "Multi-agent energy management of smart islands using primal-dual method of multipliers," *Energy*, vol. 208, Oct. 2020.

[23] Z. Li, C. Zang, P. Zeng *et al*., "Fully distributed hierarchical control of parallel grid-supporting inverters in islanded AC microgrids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 679-690, Feb. 2018.

[24] Z. Wang, B. Chen, J. Wang *et al.*, "Decentralized energy management system for networked microgrids in grid-connected and islanded modes," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1097-1105, Mar. 2016.

[25] A. Paudel, M. Khorasany, and H. B. Gooi, "Decentralized local energy trading in microgrids with voltage management," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 1111-1121, Feb. 2021.

[26] D. Cao, W. Hu, J. Zhao *et al.*, "Reinforcement learning and its applications in modern power and energy systems: a review," *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 6, pp. 1029-1042, Nov. 2020.

[27] Q. Zhang, K. Dehghanpour, Z. Wang *et al.*, "Multi-agent safe policy learning for power management of networked microgrids," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 1048-1062, Mar. 2021.

[28] D. Qiu, Y. Ye, D. Papadaskalopoulos *et al.*, "Scalable coordinated management of peer-to-peer energy trading: a multi-cluster deep reinforcement learning approach," *Applied Energy*, vol. 292, pp. 1-16, Jun. 2021.

[29] R. Lu, Y. Li, Y. Li *et al.*, "Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management," *Applied Energy*, vol. 276, pp. 1-10, Jul. 2020.

[30] D. Xu, Q. Wu, B. Zhou *et al.*, "Distributed multi-energy operation of coupled electricity, heating, and natural gas networks," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2457-2469, Oct. 2020.

[31] W. Liu, P. Zhuang, H. Liang *et al*., "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 6, pp. 2192-2203, Jun. 2018.
[32] D. Chen, K. Chen, Z. Li *et al*., "PowerNet: multi-agent deep reinforcement learning for scalable power grid control," *IEEE Transactions on Power Systems*, vol. 37, no. 2, pp. 1007-1017, Mar. 2022.
[33] M. Shin, D.-H. Choi, and J. Kim, "Cooperative management for PV/ESS-enabled electric vehicle charging stations: a multiagent deep reinforcement learning approach," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3493-3503, May 2020.
[34] Z. Li, L. Wu, Y. Xu *et al.*, "Multi-stage real-time operation of a multi-energy microgrid with electrical and thermal energy storage assets: a data-driven MPC-ADP approach," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 213-226, Jan. 2022.
[35] X. Liu, J. Wu, N. Jenkins *et al*., "Combined analysis of electricity and heat networks," *Applied Energy*, vol. 162, pp. 1238-1250, Jan. 2016.
[36] S. Li, W. Hu, D. Cao *et al.*, "Electric vehicle charging management based on deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 3, pp. 719-730, May 2022.
[37] R. Lowe, Y. Wu, A. Tamar *et al*., "Multi-agent actor-critic for mixed cooperative-competitive environments," in *Proceedings of Advances in Neural Information Processing Systems (NeurIPS)*, California, USA, Dec. 2017, pp. 6379-6390.
[38] T. Haarnoja, A. Zhou, P. Abbeel *et al.*, "Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor," in *Proceedings of International Conference on Machine Learning (ICML)*, Stockholm, Sweden, Jul. 2018, pp. 1-9.
[39] S. Sinha, H. Bharadhwaj, A. Srinivas *et al.* (2020, Sept.). D2RL: deep dense architectures in reinforcement learning. [Online]. Available: https://openreview.net/forum?id=mYNfmvt8oSv
[40] S. Li, W. Hu, D. Cao *et al.*, "EV charging strategy considering transformer lifetime via evolutionary curriculum learning-based multi-agent deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 4, pp. 2774-2787, Jul. 2022.
[41] S. Li, W. Hu, D. Cao *et al.*, "A multi-agent deep reinforcement learning-based approach for the optimization of transformer life using coordinated electric vehicles," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 11, pp. 7639-7652, Nov. 2022.
[42] PJM. (2017, May). Historical hourly electricity price from PJM. [Online]. Available: https://www.pjm.com/
[43] U. S. Energy Information Administration. (2018, Jul.). Historical gas price from U. S. Energy Information Administration. [Online]. Available: https://www.eia.gov/
[44] S. Yao, J. Gu, H. Zhang *et al*., "Resilient load restoration in microgrids considering mobile energy storage fleets: a deep reinforcement learning approach," in *Proceedings of 2020 IEEE Power & Energy Society General Meeting (PESGM)*, Montreal, Canada, Aug. 2020, pp. 1-5.
[45] L. Yu, Y. Sun, Z. Xu *et al.*, "Multi-agent deep reinforcement learning for HVAC control in commercial buildings," *IEEE Transactions Smart Grid*, vol. 12, no. 1, pp. 407-419, Jan. 2021.
[46] NLopt. (2022, Aug.). Library for nonlinear optimization. [Online]. Available: https://nlopt.readthedocs.io/en/latest/

Sichen Li is currently working toward the Ph. D. degree in control science and engineering from the University of Electronics Science and Technology of China, Chengdu, China. His research interests include demand response and applications of machine learning in power system operation and control.

Di Cao received the Ph. D. degree from University of Electronic Science and Technology of China, Chengdu, China, in 2021. He is currently a Postdoctoral Researcher at University of Electronic Science and Technology of China. His research interests include optimization of distribution network and applications of machine learning in power systems.

Weihao Hu received the B.Eng. and M.Sc. degrees in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, and the Ph. D. degree from Aalborg University, Aalborg, Denmark, in 2012. He is currently a Full Professor and the Director of the Institute of Smart Power and Energy Systems, University of Electronics Science and Technology of China, Chengdu, China. His research interests include artificial intelligence in modern power systems and renewable power generation.

Qi Huang received the B. S. degree in electrical engineering from Fuzhou University, Fuzhou, China, in 1996, the M.S. degree from Tsinghua University, Beijing, China, in 1999, and the Ph.D. degree from Arizona State University, Phoenix, USA, in 2003. His current research and academic interests include power system instrumentation, power system monitoring and control, and power system high performance computing.

Zhe Chen received the B.Eng. and M.Sc. degrees from the Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph. D. degree from the University of Durham, Durham, UK. He is a Full Professor with the Department of Energy Technology, Aalborg University, Aalborg, Denmark. His research interests include power electronics and electric machines, wind energy, and modern power systems.

Frede Blaabjerg received the Ph. D. degree in electrical engineering from Aalborg University, Aalborg, Denmark, in 1995. He was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. His current research interests include power electronics and its applications such as in wind turbines, photovoltaic (PV) systems, reliability, harmonics and adjustable speed drives.