This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

Coordinating Multi-Energy Microgrids for Integrated Energy System Resilience: a Multi-Task Learning Approach

Yi Wang, *Member, IEEE,* Dawei Qiu, *Member, IEEE,* Xiaotian Sun, *Student Member, IEEE,* Zhaohong Bie, *Fellow, IEEE,* and Goran Strbac, *Member, IEEE*

***Abstract*—High-impact and low-probability events have occurred more frequently than before, which can seriously damage energy supply infrastructures. As localized small energy systems, multi-energy microgrids (MEMGs) can provide a viable solution for the system-wise load restoration of integrated energy systems (IESs), due to their enhanced flexibility and controllability. However, existing literature tends to realize MEMGs as corrective response rather than load restoration resource after extreme events, which cannot fully exploit the benefits of multi-MEMGs on IES resilience. This paper introduces a decentralized operating paradigm for the real-time coordination of local multi-MEMGs towards system-wise IES load restoration, while a novel topology-aware multi-task reinforcement learning method with soft modularization is proposed to solve it. The multi-task learning framework enables MEMGs to simultaneously learn scheduling decisions across different network topologies and better adapt to unanticipated contingencies. Additionally, to avoid insecure MEMG operations, a physics-informed safety layer is embedded on top of the multi-task learning framework for action corrections. Case studies have been conducted on two IESs (33-bus power, 20-node gas, and 20-node heat network as well as 69-bus power, 40-node gas, and 62-node heat network) to evaluate the effectiveness of the proposed method in enabling effective coordination among multi-MEMGs towards system-wise IES load restoration.**

***Index Terms*—Multi-energy microgrids, Resilience, Integrated energy systems, Multi-task reinforcement learning, Safety layer.**

NOMENCLATURE

A. Abbreviation

MEMG	Multi-energy microgrid
IES	Integrated energy system
PDN	Power distribution network
GDN	Gas distribution network
DHN	District heating network
GG	Gas-fired generator
ES	Energy storage
CHP	Combined heat and power
EHP	Electrical heat pump
GW	Gas well
GB	Gas boiler
GS	Gas storage
TS	Thermal storage
MARL	Multi-agent reinforcement learning
MGCC	Microgrid central controller
OEF	Optimal energy flow

B. Indices and Sets

$i \in \mathcal{I}$	Index and set of MEMGs
$t \in T$	Index and set of time steps
$g \in \mathcal{GG}$	Index and set of GGs
$g \in \mathcal{GW}$	Index and set of GWs
$g \in \mathcal{CHP}$	Index and set of CHPs
$e \in \mathcal{EHP}$	Index and set of EHPs
$g \in \mathcal{GB}$	Index and set of GBs
$k \in \mathcal{ES}$	Index and set of ESs
$b \in \mathcal{B}_i^{mg}$	Index and set of buses/nodes in MEMG i
$l \in \mathcal{L}_i^{mg}$	Index and set of lines in MEMG i
$d \in \mathcal{PD}$	Index and set of loads in the PDN
$d \in \mathcal{GD}$	Index and set of loads in the GDN
$d \in \mathcal{HD}$	Index and set of loads in the DHN

C. Parameters

Δt	Time resolution (1 hour)
λ_d^p	Shedding cost of load d in PDN (£/kWh)
λ_d^g	Shedding cost of load d in GDN (£/Sm^3)
λ_d^h	Shedding cost of load d in DHN (£/kWh)
$\overline{P}_{d,t}^{pd}$	Baseline load d in PDN at time t (kW)
$\overline{G}_{d,t}^{gd}$	Baseline load d in GDN at time t (Sm$^3/h$)
$\overline{H}_{d,t}^{hd}$	Baseline load d in DHN at time t (kW)
$\overline{P}_{i,g}^{gg}$	Maximum limit of active power of GG g in MEMG i (kW)
$\underline{P}_{i,g}^{gg}$	Minimum limit of active power of GG g in MEMG i (kW)
$\overline{Q}_{i,g}^{gg}$	Maximum limit of reactive power of GG g in MEMG i (kVAR)
$\underline{Q}_{i,g}^{dg}$	Minimum limit of reactive power of GG g in MEMG i (kVAR)
$\eta_{i,g}^{gg}$	Coefficient for gas consumption of GG g in MEMG i (Sm$^3/kWh$)
$\overline{G}_{i,g}^{gw}$	Maximum gas output of GW g in MEMG i

This work was supported by two UK EPSRC projects: 'Integrated Development of Low-Carbon Energy Systems (IDLES): A Whole-System Paradigm for Creating a National Strategy' (project code: EP/R045518/1) and 'Technology Transformation to Support Flexible and Resilient Local Energy Systems' (project code: EP/T021780/1), and one Horizon Europe project: 'Reliability, Resilience and Defense technology for the griD' (Grant agreement ID: 101075714). *(Corresponding author: Dawei Qiu.)*

Yi Wang, Dawei Qiu, and Goran Strbac are with the Department of Electrical and Electronic Engineering, Imperial College London, London, SW7 2AZ, U.K. (e-mail: {yi.wang18, d.qiu15, g.strbac}@imperial.ac.uk).

Xiaotian Sun and Zhaohong Bie are are with the School of Electrical Engineering, Xi'an Jiaotong University, Xi'an, 710049, China (e-mail: xtsun@stu.xjtu.edu.cn, zhbie@mail.xjtu.edu.cn).

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

(Sm^3/h)

$\underline{G}_{i,g}^{gw}$ Minimum gas output of GW g in MEMG i (Sm^3/h)

$\overline{P}_{i,g}^{chp}$ Power capacity of CHP g in MEMG i (kW)

$\overline{H}_{i,g}^{chp}$ Heat capacity of CHP g in MEMG i (kW)

$\eta_{i,g}^{chpe}$ Electricity energy conversion efficiency of CHP g in MEMG i (%)

$\eta_{i,g}^{chph}$ Heat energy conversion efficiency of CHP g in MEMG i (%)

$\overline{H}_{i,e}^{ehp}$ Heat capacity of EHP e in MEMG i (kW)

$\eta_{i,e}^{ehp}$ Heat energy conversion efficiency of EHP e in MEMG i (%)

$\overline{H}_{i,g}^{gb}$ Heat capacity of GB g in MEMG i (kW)

$\eta_{i,g}^{gb}$ Heat energy conversion efficiency of GB g in MEMG i (%)

$\overline{P}_{i,k}^{es}$ Power capacity of ES k in MEMG i (kW)

$\overline{E}_{i,k}^{es}$ Energy capacity of ES k in MEMG i (kWh)

$\underline{E}_{i,k}^{es}$ Minimum energy of ES k in MEMG i (kWh)

$\eta_{i,k}^{es}$ Charging/discharging coefficient of ES k in MEMG i (%)

D. Variables

$P_{i,g,t}^{gg}$ Active power output of GG g in MEMG i at time t (kW)

$Q_{i,g,t}^{gg}$ Reactive power output of GG g in MEMG i at time t (kVAR)

$G_{i,g,t}^{gw}$ Gas output of GW g in MEMG i at time t (Sm^3/h)

$P_{i,g,t}^{chp}$ Power output of CHP g in MEMG i at time t (kW)

$H_{i,g,t}^{chp}$ Heat output of CHP g in MEMG i at time t (kW)

$G_{i,g,t}^{chp}$ Gas input of CHP g in MEMG i at time t (Sm^3/h)

$H_{i,e,t}^{ehp}$ Heat output of EHP e in MEMG i at time t (kW)

$P_{i,e,t}^{ehp}$ Power input of EHP e in MEMG i at time t (kW)

$H_{i,g,t}^{gb}$ Heat output of GB g in MEMG i at time t (kW)

$G_{i,g,t}^{gb}$ Gas input of GB g in MEMG i at time t (Sm^3/h)

$P_{i,k,t}^{esc}$ Charging power of ES k in MEMG i at time t (kW)

$P_{i,k,t}^{esd}$ Discharging power of ES k in MEMG i at time t (kW)

$E_{i,k,t}^{es}$ Energy content of ES k in MEMG i at time t (kWh)

$P_{i,t}^{mg}$ Active power exchange of MEMG i with the IES at time t (kW)

$Q_{i,t}^{mg}$ Reactive power exchange of MEMG i with the IES at time t (kVAR)

$G_{i,t}^{mg}$ Gas exchange of MEMG i with the IES at time t (Sm^3/h)

$H_{i,t}^{mg}$ Heat exchange of MEMG i with the IES at time t (kW)

$P_{d,t}^{pd}$ Restored load d in the PDN at time t (kW)

$G_{d,t}^{gd}$ Restored load d in the GDN at time t (Sm^3/h)

$H_{d,t}^{hd}$ Restored load d in the DHN at time t (kW)

I. Introduction

A. Background and Motivation

EXTREME weather events, characterized by their high impact and low probability, can disrupt system components and cause severe damage. The increasing interdependencies between different energy sectors (e.g., power, gas, and heat) further exacerbate the consequences of extreme events. To tackle these challenges, the concept of *resilience* has been adopted in the field of integrated energy systems (IESs) [1]. In light of the potentially serious disruptions, the primary objective of a resilient IES after extreme events is to maintain the uninterrupted supply of essential loads across different energy sectors, thereby rendering a system-wise IES load restoration problem [2].

Decentralization and digitalization are rapidly transforming the energy industry and challenging the traditional top-down philosophy of energy systems [2]. As localized small energy systems, multi-energy microgrids (MEMGs) have been widely integrated into IESs for the effective coordination of various small-size energy sources [1]. Given their advanced control capabilities, utilizing multi-MEMGs to deliver resilience services for IESs via effective energy exchanges can be a promising research direction. Therefore, this paper suggests a real-time control strategy for the decentralized coordination of multi-MEMGs to achieve system-wise IES load restoration.

B. Literature Review

Previous work has investigated the coordination effect of multi-MGs on the load restoration problems of power distribution networks (PDNs), which can be classified into two categories depending on the control methods: 1) *centralized control* [3], [4] that involves a central entity for MGs to solve a global optimization, but requires intensive communication resources and can be prone to single-point failure; 2) *distributed control* [5], [6] that removes the central controller and allows MGs to make individual power schedules based on local information. However, the above work only focuses on the power sector without considering the interdependencies among multiple energy sectors. The flexibility of energy supply within only PDN is relatively limited, which can further limit the resilience of energy supply infrastructures [7].

To address this issue, there have been studies focused on developing resilient operation strategies for IESs, e.g., the load restoration of integrated power-gas networks based on centralized [8], [9] or distributed [10] control; nevertheless, the interactions with district heating networks (DHNs) are not considered. To capture a more comprehensive energy system, PDNs, DHNs, and gas distribution networks (GDNs) are fully modeled in a centralized [7] or distributed fashion [11] towards IES resilience. However, the above papers [7], [11] directly employ distributed devices as controllable components for IES resilience support, while it might be difficult for network operators to regulate and coordinate these large-scale and small-size devices as they may belong to different entities.

In this context, there have been studies focused on the resilience-oriented operation of MEMGs, due to their enhanced control capabilities in managing small-size distributed

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

devices [1]. For instance, planning or operation strategies have been developed in [12]–[16] to enhance the resilience of MEMGs themselves; nevertheless, these papers do not consider how to use MEMGs as resilience resources to support system-wise IES resilience. To address this problem, energy hubs have been included in IES operation towards overall resilience enhancement [17]–[21]. However, DHN operations are not considered in [17]–[19], [21], while deterministic models are employed in [17], [18], [20], [21] neglecting uncertainties and contingencies during extreme events. Additionally, it is worth noting that the above papers mainly consider energy hubs as energy conversion centers and corrective response between different energy networks. Compared with energy hubs, MEMGs with enhanced control capabilities can be realized as load restoration resources to provide sufficient energy supply for the IES load restoration after extreme events. Furthermore, all the above papers [3]–[21] adopting model-based optimization methods generally require a thorough understanding of the system-wise IES model, accurate uncertainty forecasting, and an effective optimization solver, which can be impractical, particularly given the potentially degraded communication infrastructures and the necessity for rapid response against extreme events.

Therefore, it is necessary to develop a real-time control strategy to realize local multi-MEMGs as the load restoration resource towards system-wise IES load restoration. Under this setting, MEMGs can interact with the IES environment and learn control policies rather than directly requiring the system-wise IES model, while the learned control policy will be capable of adapting to various system uncertainties and making realistic dispatch decisions without performing an optimization. In response to this requirement, *reinforcement learning* (RL) [22], as a data-driven and real-time control method, has been applied to address load restoration problems in PDNs, e.g., deep Q-network (DQN) [23], [24], double DQN (DDQN) [25], deep deterministic policy gradient (DDPG) [26], [27], and soft actor-critic (SAC) [28]. However, our focus is on the coordinated control of local multi-MEMGs for system-wise IES resilience. Utilizing the above single-agent RL methods for multi-MEMGs can raise serious scalability issues, due to the dimensionality curse from the exponential growth of state and action spaces with increasing agent size.

To this end, deploying a multi-agent RL (MARL) method for the studied system-wise IES load restoration problem with multi-MEMGs is reasonable; nevertheless, it has not yet been investigated. In the literature, studies have tried to apply various MARL methods to solve resilience-oriented coordination problems, e.g., MADQN [29], [30], MADPG [29], MASAC [29], [31], MADDQN [32], and multi-agent proximal policy optimization (MAPPO) [33], [34]. However, the above papers all realize one type of individual distributed devices as agents for resilience purposes, which may be difficult to achieve effective coordination among different types of devices and eventually lead to poor training performance, due to their large-scale nature. Additionally, none of the above papers consider a comprehensive IES model including PDN, GDN, and DHN. Furthermore, papers [29]–[31] only consider a fixed network topology during the RL training process, which can be impractical. This is because extreme events are characterized by high impact, leading to severe contingencies and dynamically changing network topology. Even though papers [32]–[34] realize contingencies (e.g., line faults) as uncertainties that can theoretically capture the dynamics of network topologies, the differences among them can be significant, making the training process very challenging and eventually leading to poor RL performance.

As such, it is necessary to develop a topology-aware MARL method for the resilient coordination of multi-MEMGs under changing network topologies towards IES resilience enhancement. Multi-task learning refers to a single-shared learning method that can perform multiple tasks with common features and offer the advantages of improved data efficiency, faster convergence, and reduced overfitting to a single task [35]. Specifically, multi-task learning can leverage the experiences gained from one task to improve the performance of others. It is reasonable to apply multi-task learning to the studied resilience-oriented coordination problem of multi-MEMGs; nevertheless, it has not yet been investigated. Inspired by recent advances in multi-task learning, this paper proposes a topology-aware MARL method with soft modularization to embed different network topologies caused by contingencies in the training procedure as different tasks. In this case, the well-trained RL policies can avoid overfitting in one contingency and better adapt to unanticipated network topologies. Furthermore, changing topologies in the multi-task learning framework can significantly influence the operating boundaries of MEMGs and easily cause constraint violations. To ensure secure MEMG operations, a physics-informed safety layer is extracted and cast on top of the multi-task learning framework for action corrections when detecting unsafe operations.

C. Research Gaps and Contributions

Although both model-based optimization and RL-based methods have been successfully applied to support the IES resilience, the following research gaps still remain:

1) Previous work either only focuses on the resilience of MEMGs themselves [12]–[16] or employs energy hubs as energy conversion centers for corrective response [17]–[21], which may not fully exploit the benefits of multi-MEMGs with enhanced control capabilities in supporting IES load restoration after extreme events.
2) Previous work [3]–[21] has developed various model-based optimization methods for the coordination of MGs or MEMGs towards resilience enhancement. However, model-based methods can be time-consuming and depend on accurate uncertainty forecasting, which may not meet the requirements of resilient IES operations for timely response and accurate decision making.
3) Previous work [23]–[33] has tried to employ various single-agent RL or MARL methods for PDN load restoration, while the system-wise IES resilience has not been investigated. Additionally, existing literature on MARL either considers a fixed network topology [29]–[31] or realizes contingencies as uncertainties [32]–[34], which may be incapable of adapting to different network topolo-

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

gies, lead to poor training performance, and even insecure system operations.

To fill in the above research gaps, this paper proposes a novel topology-aware MARL method for the resilience-oriented coordination of multi-MEMGs towards system-wise IES load restoration. In detail, the main contributions of this paper are summarized as follows:

1) Propose a decentralized framework for the individual energy schedules of local multi-MEMGs towards system-wise IES resilience enhancement. This framework addresses the following limitations of previous work: i) unlike [23]–[33] that only focus on PDN, this paper studies IES resilience; and ii) unlike [12]–[16] that focus on the resilience of a MEMG itself, this paper deploys MEMGs to provide energy supply for system-wise IES resilience; iii) unlike [17]–[21] that use energy hubs as energy conversion centers, this paper realizes MEMGs as load restoration resources after extreme events.
2) Formulate the resilience-oriented coordination problem of multi-MEMGs as a *Decentralized Partially Observable Markov Decision Process* (Dec-POMDP), where the operation model of system-wise IES is assumed as the environment and microgrid central controllers (MGCCs) are realized as agents. Unlike [3]–[21] that employs model-based optimization methods, this paper develops a real-time MEMG control strategy.
3) Suggest a novel topology-aware MARL method to effectively solve the Dec-POMDP by i) developing a multi-task learning framework to deal with changing network topologies caused by severe contingencies; ii) constructing an actor-critic architecture-based PPO algorithm to handle the high-dimensional observation and action spaces; iii) introducing a physics-informed safety layer on top of the MT learning framework that can respect physical constraints.
4) Validate the superiority of the proposed topology-aware MARL approach in system-wise IES load restoration, compared with existing MARL methods. The generalized scheduling policy is scalable and adaptable to different network sizes and topologies.

D. Paper Organization

The rest of the paper is organized as follows. Section II presents the proposed framework of multi-MEMGs in an IES as well as research challenges and solutions. Section III describes the general formulations of MEMGs and IESs. Section IV introduces the Dec-POMDP formulation, while Section V presents the proposed MARL method based on multi-task learning. In Section VI, case studies are carried out and discussed. Section VII draws the conclusions and future work of this paper.

II. Proposed Framework of Multi-MEMGs in an Integrated Energy Network

A. Problem Descriptions

Resilience refers to "*the ability of a system to limit the extent, system impact, and duration of degradation in order to sustain critical services following an extraordinary event*", where the key enablers for a resilient response include the capacity to anticipate, absorb, rapidly recover from, adapt to, and learn from such an event [36]. Specifically, these extreme events may be caused by natural threats, accidents, equipment failures, and deliberate physical or cyber-attacks [36]. As illustrated in Fig. 1, the chronological timeline of a resilient system through disruption includes five stages: 1) preventive stage from t_0 to t_1, 2) disturbance progress stage from t_1 to t_2, 3) post-disturbance degraded stage from t_2 to t_3, 4) restoration stage from t_3 to t_4, and 5) post-restoration stage after t_4 [37].

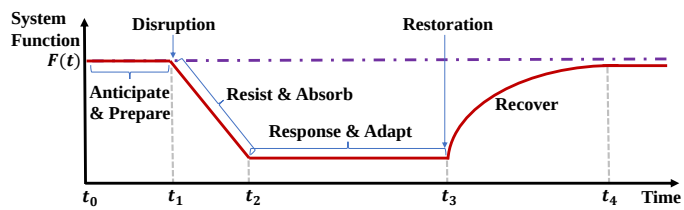


Fig. 1. Illustration of the chronological timeline of a resilient system through an extreme event.

In this paper, we focus on the resilience enhancement problem of IESs (including PDNs, GDNs, and DHNs) with multi-MEMGs during the restoration stage (e.g., $t_3 - t_4$ in Fig 1). In general, these MEMGs are connected to the system-wise IES for energy exchanges via specific electric buses, gas nodes, or heat nodes, as depicted in Fig. 2. Each MEMG is equipped with a MGCC that can manage the energy dispatches of controllable components (e.g., gas-fired generators (GGs), energy storages (ESs), combined heat and power (CHPs), electrical heat pumps (EHPs), gas wells (GWs), gas boilers (GBs), gas storages (GSs), thermal storages (TSs), etc.) within its own area. Then, these MEMGs coordinate with each other as the load restoration resource to support system-wise IES resilience, instead of serving as the corrective response.

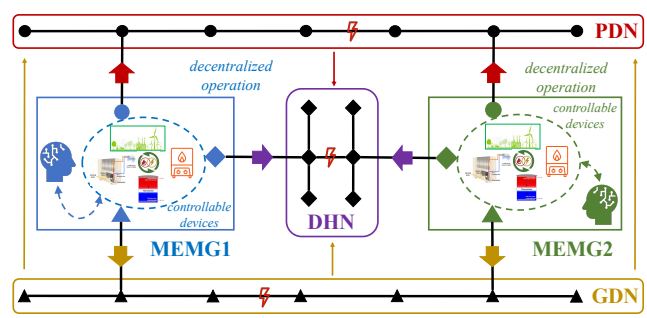


Fig. 2. Illustration of the coordination scheme of multi-MEMGs in the IES towards load restoration after extreme events.

B. Research Challenges and Solutions

To enhance IES resilience, especially when the communication infrastructure is damaged or degraded, two main challenges need to be appropriately addressed:

1) *How to coordinate these MEMGs in an effective way but without requiring intensive interactions among them when*

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

communication resources are limited? To reach this target, a decentralized operating paradigm is necessary, allowing these MEMGs to manage their own portfolios without relying on intensive communication resources. In other words, MEMGs can solely depend on the local information and then coordinate in a decentralized manner for system-wise IES load restoration.

2) *How to ensure the effectiveness of real-time MEMG operations under uncertain network topologies caused by severe contingencies?* To reach this target, a topology-aware multi-task learning framework with feature extraction capabilities is crucial, allowing these MEMGs to adapt to various network topologies and even make informed scheduling decisions against unanticipated contingencies.

In this context, the suggested topology-aware MARL approach can be efficiently deployed to solve the system-wise IES load restoration problem and overcome the above two challenges via the decentralized multi-MEMGs accounting for changing network topologies.

III. General Formulations of MEMGs in IESs

This section aims at providing the mathematical formulations of MEMGs as well as the IES network, including PDN, GDN, and DHN. In general, severe contingencies (e.g., line faults) can occur in the IES network after extreme events, causing energy shortage and essential load shedding. MEMGs can connect with the IES via certain distribution lines, gas pipelines, and heat pipelines, effectively manage their local energy resources, and then provide energy supply for the IES towards effective load restoration. Additionally, in the IES, PDN, GDN, and DHN are interconnected through various energy resources with energy conversion functions, e.g., GGs, CHPs, EHPs, etc.

A. Mathematical Models of MEMGs

The components inside MEMGs include GGs, ESs, GWs, GSs, CHPs, TSs, EHPs, GBs and renewable energy resources, where their operating models are described as below:

1) Gas-fired Generator: GGs consume natural gas and supply electricity, which can be expressed as

$$\underline{P}_{i,g}^{gg} \leq P_{i,g,t}^{gg} \leq \overline{P}_{i,g}^{gg}, \forall g \in \mathcal{GG}, \forall i \in \mathcal{I}, \tag{1}$$

$$\underline{Q}_{i,g}^{gg} \leq Q_{i,g,t}^{gg} \leq \overline{Q}_{i,g}^{gg}, \forall g \in \mathcal{GG}, \forall i \in \mathcal{I}, \tag{2}$$

$$P_{i,g,t}^{gg} = G_{i,g,t}^{gg} / \eta_{i,g}^{gg}, \forall g \in \mathcal{GG}, \forall i \in \mathcal{I}, \tag{3}$$

where $P_{i,g,t}^{gg}$ and $Q_{i,g,t}^{gg}$ are the active and reactive power outputs of GG g in MEMG i, restricted by (1)-(2), respectively. The energy conversion between power output $P_{i,g,t}^{gg}$ and gas consumption $G_{i,g,t}^{gg}$ is presented in (3), given coefficient $\eta_{i,g}^{gg}$.

2) Gas Well: GWs are used as the main source of the gas sector, which can be expressed as

$$\underline{G}_{i,g}^{gw} \leq G_{i,g,t}^{gw} \leq \overline{G}_{i,g}^{gw}, \forall g \in \mathcal{GW}, \forall i \in \mathcal{I}, \tag{4}$$

where constraint (4) limits the gas output $G_{i,g,t}^{gw}$ of GW g in MEMG i.

3) Combined Heat and Power: The coupled power and heat generation of a back-pressure CHP g in MEMG i can be expressed as

$$P_{i,g,t}^{chp} = \eta_{i,g}^{chpe} G_{i,g,t}^{chp}, \forall g \in \mathcal{CHP}, \forall i \in \mathcal{I}, \tag{5}$$

$$H_{i,g,t}^{chp} = \eta_{i,g}^{chph} G_{i,g,t}^{chp}, \forall g \in \mathcal{CHP}, \forall i \in \mathcal{I}, \tag{6}$$

$$0 \leq P_{i,g,t}^{chp} \leq \overline{P}_{i,g}^{chp}, \forall g \in \mathcal{CHP}, \forall i \in \mathcal{I}, \tag{7}$$

$$0 \leq H_{i,g,t}^{chp} \leq \overline{H}_{i,g}^{chp}, \forall g \in \mathcal{CHP}, \forall i \in \mathcal{I}, \tag{8}$$

where (5) and (6) correspond to the conversion from gas into power $P_{i,g,t}^{chp}$ and heat $H_{i,g,t}^{chp}$, respectively. The power and heat outputs of CHP g in MEMG i are limited by (7) and (8), respectively.

4) Electrical Heat Pump: EHPs consume electricity and then produce heat energy, which can be expressed as

$$H_{i,e,t}^{ehp} = \eta_{i,e}^{ehp} P_{i,e,t}^{ehp}, \forall e \in \mathcal{EHP}, \forall i \in \mathcal{I}, \tag{9}$$

$$0 \leq H_{i,e,t}^{ehp} \leq \overline{H}_{i,e}^{ehp}, \forall e \in \mathcal{EHP}, \forall i \in \mathcal{I}, \tag{10}$$

where $\eta_{i,e}^{ehp}$ and $\overline{H}_{i,e}^{ehp}$ are the energy conversion efficiency and the capacity of EHP e in MEMG i, respectively.

5) Gas Boilers: GBs are regarded as vessels that convert gas to heat, which can be expressed as

$$H_{i,g,t}^{gb} = \eta_{i,g}^{gb} G_{i,g,t}^{gb}, \forall g \in \mathcal{GB}, \forall i \in \mathcal{I}, \tag{11}$$

$$0 \leq H_{i,g,t}^{gb} \leq \overline{H}_{i,g}^{gb}, \forall g \in \mathcal{GB}, \forall i \in \mathcal{I}, \tag{12}$$

where $\eta_{i,g}^{gb}$ and $\overline{H}_{i,g}^{gb}$ are the energy conversion efficiency and the GB capacity, respectively.

6) Storage Units: Storage units are used to balance energy and demand. The operation model of ESs is expressed as

$$0 \leq P_{i,k,t}^{esc} \leq \overline{P}_{i,k}^{es}, \forall k \in \mathcal{ES}, \forall i \in \mathcal{I}, \tag{13}$$

$$-\overline{P}_{i,k}^{es} \leq P_{i,k,t}^{esd} \leq 0, \forall k \in \mathcal{ES}, \forall i \in \mathcal{I}, \tag{14}$$

$$\underline{E}_{i,k}^{es} \leq E_{i,k,t}^{es} \leq \overline{E}_{i,k}^{es}, \forall k \in \mathcal{ES}, \forall i \in \mathcal{I}, \tag{15}$$

$$E_{i,k,t+1}^{es} = E_{i,k,t}^{es} + P_{i,k,t}^{esc} \eta_{i,k}^{es} + P_{i,k,t}^{esd} / \eta_{i,k}^{es}, \quad \forall k \in \mathcal{ES}, \forall i \in \mathcal{I}, \tag{16}$$

where the charging and discharging power of ES k in MEMG i are limited by (13)-(14), respectively. The energy content of ES k is limited by (15), while the time-coupled property of ES k is presented in (16), given the efficiency $\eta_{i,k}^{es}$. The rates $G_{i,k,t}^{gsc}, G_{i,k,t}^{gsd}, H_{i,k,t}^{tsc}, H_{i,k,t}^{tsd}$ and energy contents $E_{i,k,t}^{gs}, E_{i,k,t}^{ts}$ of GS k and TS k in MEMG i can be formulated similarly as (13)-(16).

B. Energy exchanges with the IES

The MEMG i can exchange its energy with the system-wise IES subject to its own operation constraints, which are

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

expressed as

$$\xi_{i,t}^{mg} = \begin{cases} \underbrace{\{P_{i,t}^{mg}, Q_{i,t}^{mg}, G_{i,t}^{mg}, H_{i,t}^{mg}\}}_{\text{Energy balance constraints}}, \\ \underbrace{\{V_{i,b,t}, \rho_{i,b,t}, \varrho_{i,b,t}, \tau_{i,b,t}\}}_{\text{Nodal operation constraints}}, \forall b \in \mathcal{B}_i^{mg}, \quad \forall i \in \mathcal{I}, \\ \underbrace{\{P_{i,l,t}, Q_{i,l,t}, G_{i,l,t}, H_{i,l,t}\}}_{\text{Line capacity constraints}}, \forall l \in \mathcal{L}_i^{mg}, \end{cases} \tag{17}$$

where the constraint set $\xi_{i,t}^{mg}$ includes three types of constraints: i) energy balances, ii) nodal operation limits, and iii) line capacity limits. $\mathcal{B}_i^{mg}$ and $\mathcal{L}_i^{mg}$ correspond to the sets of buses/nodes and lines of MEMG i, which is categorized in set $\mathcal{I}$. $P_{i,t}^{mg}, Q_{i,t}^{mg}, G_{i,t}^{mg}, H_{i,t}^{mg}$ denote the energy exchanges of MEMG i with the IES. $V_{i,b,t}, \rho_{i,b,t}, \varrho_{i,b,t}, \tau_{i,b,t}$ refer to the voltage, gas pressure, water pressure, and temperature of bus/node b in MEMG i, respectively. $P_{i,l,t}, Q_{i,l,t}, G_{i,l,t}, H_{i,l,t}$ express the energy flow of line l in MEMG i.

C. Objective Function

The objective is to maximize the expectation of the weighted IES load restoration (i.e., the power-gas-heat network), which is expressed as

$$\max \mathbb{E}\Big[\sum_{d\in\mathcal{PD}} \lambda_d^p P_{d,t}^{pd} + \sum_{d\in\mathcal{GD}} \lambda_d^g G_{d,t}^{gd} + \sum_{d\in\mathcal{HD}} \lambda_d^h H_{d,t}^{hd}\Big], \tag{18}$$

where $P_{d,t}^{pd}, G_{d,t}^{gd}, H_{d,t}^{hd}$ represent the restored load d of PDN, GDN, and DHN in the IES, respectively, while $\lambda_d^p, \lambda_d^g, \lambda_d^h$ correspond to their individual load shedding costs, prioritizing the restoration of essential loads over non-essential loads. The IES operation can be modeled by an optimal energy flow (OEF) algorithm [11], [38], where detailed model formulations including PDN, GDN, and DHN can be found in constraints (43)-(59) of Appendix A.

D. Problem Challenges

Addressing the model-based optimization with the objective (18) subject to MEMG constraints (1)-(17) and the OEF constraints (43)-(59) is very challenging. First, it may be impractical to obtain the system-wise IES models and technical parameters for optimization purposes, due to the potentially degraded communication structures after extreme events. Second, considering the high-impact nature of extreme events, it can be very difficult to approximate the probability distributions of various system uncertainties (e.g., demand, renewable energy resources, etc.) and contingencies (e.g., line faults), leading to inaccurate and time-consuming decision-making. Third, model-based optimization methods lack the ability to adapt to dynamically changing system conditions, as the optimization needs to be solved for any new state. Thus, this paper proposes a data-driven and real-time MARL approach to solve the system-wise IES load restoration problem, which is capable of learning the features of uncertainties and contingencies during the training process. The well-trained MARL policies can be deployed in real-time decision-making processes of real-world applications.

IV. Problem Reformulation as a Dec-POMDP

Dec-POMDPs represent sequential decision-making problems for the coordination among multiple agents, where each agent can only receive a local observation, learn a local policy, and share a common reward function [39]. In the studied resilience-oriented coordination problem, MEMGs are formed as a multi-agent setup in a decentralized fashion with a dynamic decision-making process. Each MEMG can only access partial information about the system-wise IES (power-gas-heat network) and then make decentralized scheduling decisions towards overall IES load restoration. Thus, it is reasonable to formulate the studied resilience-oriented coordination problem as a Dec-POMDP [39]. Generally, Dec-POMDP can be realized as a 7-tuple $\langle \mathcal{I}, \mathcal{S}, \{\mathcal{O}_i\}, \{\mathcal{A}_i\}, \mathcal{R}, \mathcal{T}, \gamma \rangle$, where details of these components are given as follows:

A. Agent

Each MGCC $i \in \mathcal{I}$ is considered an agent that has direct control over the components within its own area.

B. Environment

The environment is defined as the operation model of the power-gas-heat network in Appendix A.

C. State and Observation

The environment state $s_t \in \mathcal{S}$ describes the observations $\{o_{1,t}, ..., o_{I,t}\}$ of MGCC agents $\mathcal{I}$ and the environment information (e.g., the IES models and parameters). The observation $o_{i,t}$ of agent i is expressed as

$$\begin{aligned} o_{i,t} =& [N_t^{ln}, P_t, Q_t, G_t, H_t, P_{i,t}^{pd}, Q_{i,t}^{pd}, P_{i,t}^{res}, \\ & G_{i,t}^{gd}, H_{i,t}^{hd}, E_{i,k,t}^{es}, E_{i,k,t}^{gs}, E_{i,k,t}^{ts}], \in \mathcal{O}_i, \end{aligned} \tag{19}$$

which comprises three parts: 1) the network topology N_t^{ln} presented by the time-varying statuses of lines (e.g., fault or not) in the IES; 2) the nodal energy injection information including nodal power injections $P_t = \{P_{b,t} : b \in \mathcal{PB}\}$ and $Q_t = \{Q_{b,t} : b \in \mathcal{PB}\}$, gas injection $G_t = \{G_{b,t} : b \in \mathcal{GB}\}$, and heat injection $H_t = \{H_{b,t} : b \in \mathcal{HB}\}$; 3) the local MEMG information, including demand information $P_{i,t}^{pd}, Q_{i,t}^{pd}, G_{i,t}^{gd}, H_{i,t}^{hd}$, renewable generation $P_{i,t}^{res}$, and battery energy content $E_{i,k,t}^{es}, E_{i,k,t}^{gs}, E_{i,k,t}^{ts}$.

D. Action

The action $a_{i,t}$ conducted by MGCC agent i includes the energy scheduling decisions for managing the controlled components, which is defined as

$$\begin{aligned} a_{i,t} =& [a_{i,g,t}^{gg}, a_{i,g,t}^{gw}, a_{i,g,t}^{chp}, a_{i,e,t}^{ehp}, \\ & a_{i,g,t}^{gb}, a_{i,k,t}^{es}, a_{i,k,t}^{gs}, a_{i,k,t}^{ts}] \in \mathcal{A}_i, \end{aligned} \tag{20}$$

where actions $a_{i,g,t}^{gg} \in [0,1]$ denote the power output magnitude of GG g in MEMG i as a ratio of the corresponding power capacities $[\underline{P}_{i,g}^{gg}, \overline{P}_{i,g}^{gg}]$, while actions $a_{i,g,t}^{gw} \in [0,1]$ denote the gas generation magnitude of GW g as a ratio of the gas capacities $[\underline{G}_{i,g}^{gw}, \overline{G}_{i,g}^{gw}]$. $a_{i,e,t}^{ehp}, a_{i,g,t}^{gb}, a_{i,g,t}^{chp} \in [0,1]$

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

represent the heat generation magnitude of EHP e, GB g, and CHP g as a percentage of their maximum energy limits $\overline{H}_{i,e}^{ehp}$, $\overline{H}_{i,g}^{gb}$, and $\overline{H}_{i,g}^{chp}$, respectively. Similarly, actions $a_{i,k,t}^{es}, a_{i,k,t}^{gs}, a_{i,k,t}^{ts} \in [-1, 1]$ represent the rate of charging (positive) and discharging (negative) of storage units as a percentage of their capacities.

E. State Transition

In the proposed MEMG coordination problem, Δt = 1 hour represents one time step. For each time step t, agent i observing its current observation $o_{i,t}$ can compute and execute an action $a_{i,t}$ to the environment by using a policy $\pi_i(a_{i,t}|o_{i,t})$. Afterward, following the state transition function $s_{t+1} = \mathcal{T}(s_t, a_{1:I,t}, \chi)$, the environment will transit into the next state s_{t+1}, which is influenced by state s_t, actions $a_{1:I,t}$, and the stochasticity of environment χ_t. For our proposed problem, $\chi_t = [P_{i,t}^{pd}, Q_{i,t}^{pd}, P_{i,t}^{res}, G_{i,t}^{gd}, H_{i,t}^{hd}]$ refers to the exogenous states, which are independent of agent actions and exhibit inherent variability. In machine learning area, RL can overcome this variability by adopting a data-driven fashion that does not depend on accurate probability distributions for uncertainties but instead learns state features from the data set [22]. To prove the effectiveness of RL in handling environment uncertainties, a test dataset (separate from training dataset) is typically utilized to verify the performance of the trained RL policy on generalization to different state conditions. Furthermore, once the RL policy is well trained, it can be directly deployed to the practical test process.

As an endogenous state, the transition of $E_{i,t}^{es}$ is influenced by the agent's action $a_{i,k,t}^{es}$ through the mutually exclusive quantities $P_{i,k,t}^{esc}$ and $P_{i,k,t}^{esd}$, which are limited by its min/max energy capacities $\underline{E}_{i,k}^{es}, \overline{E}_{i,k}^{es}$, power capacity $\overline{P}_{i,k}^{es}$, and charging/discharging efficiency $\eta_{i,k}^{es}$, which can be depicted as

$$P_{i,k,t}^{esc} = [\min(a_{i,k,t}^{es}\overline{P}_{i,k}^{es}, (\overline{E}_{i,k}^{es} - E_{i,k,t}^{es})/\eta_{i,k}^{es})]^+, \tag{21}$$

$$P_{i,k,t}^{esd} = [\max(a_{i,k,t}^{es}\overline{P}_{i,k}^{es}, (\underline{E}_{i,k}^{es} - E_{i,k,t}^{es})\eta_{i,k}^{es})]^-, \tag{22}$$

where the state transition of $E_{i,k,t}^{es}$ can be realized as (16). The transition of endogenous states $E_{i,k,t}^{gs}, E_{i,k,t}^{ts}$ for GSs and TSs can be deducted in the same matter as (21)-(22).

To this end, after obtaining the energy exchanges $P_{i,t}^{mg}, G_{i,t}^{mg}, H_{i,t}^{mg}$ of all MEMGs, the IES operator runs the OEF subject to (43)-(59) with the objective function (18), and then updates the nodal energy injections P_t, Q_t, G_t, H_t for next time step.

F. Resilience Index

As reviewed in [37], [40], resilience indices, such as restoration cost, restoration probability, time to restoration, resilience triangle, resilience trapezoid, etc., have been widely used to quantify the performance of resilience enhancement. Since this paper focuses on the system restoration stage of the resilience process, a resilience index considering the restoration costs of essential loads and non-essential loads is used, following the same practice in [32], [33], [41].

According to Fig. 1, the load restoration performance between t_3 and t_4 can be expressed as

$$RI = \frac{\int_{t_3}^{t_4} F(t)dt}{\int_{t_3}^{t_4} F_0(t)dt}, \tag{23}$$

where $F(t)$ and $F_0(t)$ denote the real-time and targeted load restoration performance, respectively. $F(t) = f_c F_c(t) + f_n F_n(t)$, where f_c and f_n are weighting factors of essential and non-essential load, respectively. In this context, the load restoration performance of the investigated IES during the scheduling horizon T can be detailed as

$$RI = \frac{\sum_{t\in T}(\sum_{d\in\mathcal{PD}} \lambda_d^p P_{d,t}^{pd} + \sum_{d\in\mathcal{GD}} \lambda_d^g G_{d,t}^{gd} + \sum_{d\in\mathcal{HD}} \lambda_d^h H_{d,t}^{hd})}{\sum_{t\in T}(\sum_{d\in\mathcal{PD}} \lambda_d^p \overline{P}_{d,t}^{pd} + \sum_{d\in\mathcal{GD}} \lambda_d^g \overline{G}_{d,t}^{gd} + \sum_{d\in\mathcal{HD}} \lambda_d^h \overline{H}_{d,t}^{hd})}, \tag{24}$$

where $P_{d,t}^{pd}, G_{d,t}^{gd}, H_{d,t}^{hd}$ represent the restored load d of PDN, GDN, and DHN, while $\overline{P}_{d,t}^{pd}, \overline{G}_{d,t}^{gd}, \overline{H}_{d,t}^{hd}$ represent the baseline of load d at time step t. $\lambda_d^p, \lambda_d^g, \lambda_d^h$ correspond to their individual load shedding costs, prioritizing the restoration of essential loads over non-essential loads [42].

G. Reward

The reward function design is responsible for motivating the agents' scheduling behaviors, and reward signals could be any unitless scalar values [22]. To improve the convergence rate and the optimality of control policies, it is better to scale the reward function between 0 and 1.

Since the objective (18) in Section III-C is the maximization of the weighted load restoration, the resilience index defined in (24) is modified as the reward function for each time step t, which can be expressed as

$$r_t = \frac{\sum_{d\in\mathcal{PD}} \lambda_d^p P_{d,t}^{pd} + \sum_{d\in\mathcal{GD}} \lambda_d^g G_{d,t}^{gd} + \sum_{d\in\mathcal{HD}} \lambda_d^h H_{d,t}^{hd}}{\sum_{d\in\mathcal{PD}} \lambda_d^p \overline{P}_{d,t}^{pd} + \sum_{d\in\mathcal{GD}} \lambda_d^g \overline{G}_{d,t}^{gd} + \sum_{d\in\mathcal{HD}} \lambda_d^h \overline{H}_{d,t}^{hd}}. \tag{25}$$

H. Objective

Every time step of the preceding process is repeated within the corresponding episode (e.g., one day), and each agent i creates a trajectory of its observations, actions, and rewards: $\tau_i = o_{i,1}, a_{i,1}, r_1, o_{i,2}, ..., r_T$ over $\mathcal{O}_i \times \mathcal{A}_i \times \mathcal{R} \to \mathbb{R}$. In the POMDP, each agent i seeks an optimal policy $\pi_i(a_{i,t}|o_{i,t})$ for the maximization of the discounted reward

$$G = \sum\nolimits_{t\in T} \gamma^t r_t, \tag{26}$$

where $\gamma \in [0, 1]$ refers to the discount factor that can determine the relative significance of immediate and future rewards.

V. Proposed Multi-Agent Learning Method

In this section, a topology-aware MARL method called MT-PIPPO is developed for resolving the proposed Dec-POMDP framework, where the overall architecture is illustrated in Fig. 3. Specifically, MT-PIPPO generates three crucial and insightful techniques: 1) utilizing a standard PPO algorithm

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

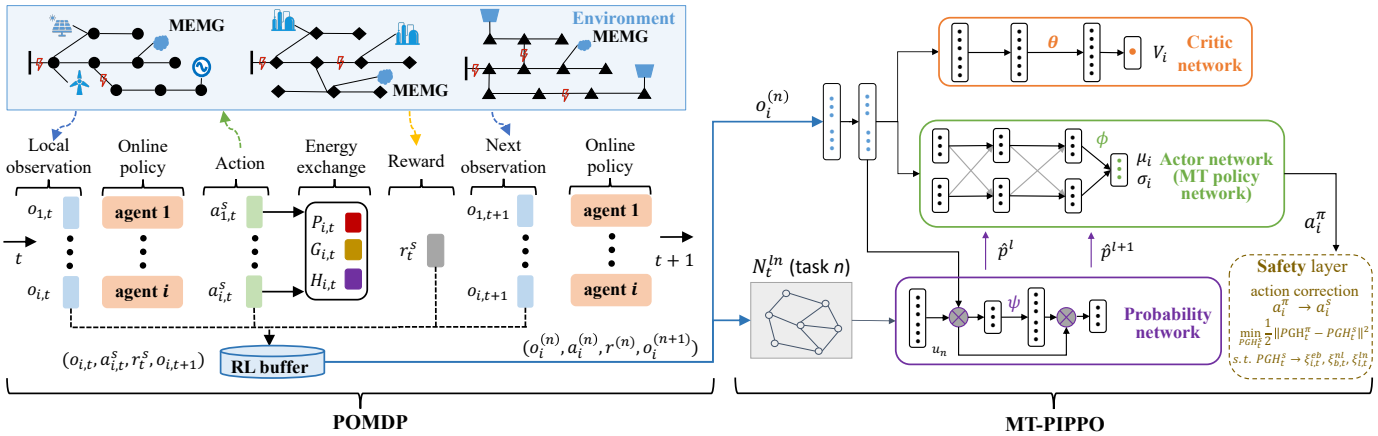


Fig. 3. Architecture of the proposed MT-PIPPO in solving Dec-POMDP of an IES load restoration problem.

[33] for the MARL policy update due to its easy hyperparameter tuning and stable learning performance; 2) developing a multi-task framework for a common representation, which can allow the algorithm to learn multiple tasks simultaneously and avoid overfitting in one contingency; and 3) casting a physics-informed safety layer that can automatically adjust the energy dispatches from multi-task learning for secure operations.

A. Proximal Policy Optimization

As a policy gradient algorithm, PPO has been applied in various control problems [33]. Overall, PPO is characterized by an actor-critic configuration that is capable of dealing with continuous spaces with high dimensions. For the actor network parameterized by ϕ_i, a stochastic policy $\pi_{\phi_i}(a|o)$ based on Gaussian distributions can be generated to capture the continuous action spaces in (20), which outputs the corresponding mean μ_i and standard deviation σ_i and samples the action $a_{i,t}$ conditioned on local observation $o_{i,t}$ for all energy dispatches of MGCC agent i. The policy gradient algorithm is employed to update the stochastic policy $\pi_{\phi_i}(a|o)$ to maximize the following clipped surrogate:

$$L_{i,t}(\phi_i) = \mathbb{E}_t\big[\min(\zeta_{i,t}\Lambda_{i,t}, \text{clip}(\zeta_{i,t}, 1-\epsilon, 1+\epsilon)\Lambda_{i,t})\big], \quad (27)$$

where $\zeta_{i,t}\Lambda_{i,t}$ refers to the policy gradient, while $\zeta_{i,t}$ is the probability ratio that is clipped by clip$(\cdot)$. $\epsilon \in [0,1]$ is a hyperparameter that is utilized to restrict the gradient update of the current policy from its old version, if the probability ratio $\zeta_{i,t}$ is over the range $[1-\epsilon, 1+\epsilon]$. This technique allows the policy gradient to be updated within a stable region, thereby overcoming the instability issue associated with the vanilla policy gradient algorithm.

Specifically, the probability proportion $\zeta_{i,t}$ in the PPO clipped policy (27) can be denoted as

$$\zeta_{i,t} = \frac{\pi_{\phi_i}(a_{i,t}|o_{i,t})}{\pi_{\phi_i^{\text{old}}}(a_{i,t}|o_{i,t})}, \quad (28)$$

where $\pi_{\phi_i}(a_{i,t}|o_{i,t})$ and $\pi_{\phi_i^{\text{old}}}(a_{i,t}|o_{i,t})$ correspond to the current policy and its old version, respectively. Additionally, the advantage function $\Lambda_{i,t}$ is defined as

$$\Lambda_{i,t} = \delta_{i,t} + (\gamma\lambda)\delta_{i,t+1} + \cdots + (\gamma\lambda)^{T-t+1}\delta_{i,T-1}, \quad (29)$$

$$\delta_{i,t} = r_{i,t} + \gamma V_{\theta_i}(o_{i,t+1}) - V_{\theta_i}(o_{i,t}), \quad (30)$$

where $V_{\theta_i}(o)$ is the state-value function that can be approximated by a critic network parameterized by θ_i, while $\gamma \in [0,1]$ and $\lambda \in [0,1]$.

B. Multi-Task Learning

In general, multi-task RL [35] is capable of jointly learning multiple tasks rather than learning them separately, given the assumption that these tasks involve common features. For our studied load restoration problem, the main focus of MEMG agents is to deal with the changing network topologies caused by severe contingencies after extreme events and provide energy supply for the IES towards overall resilience enhancement. It is worth noting that extreme events can normally affect a limited number of lines, while most lines will still remain intact under different contingency scenarios. In other words, except for the area affected by extreme events, the remaining network structure is same across different contingencies. Therefore, it is reasonable to realize different network topologies as the tasks ($\mathcal{N}^{ln}$) in the proposed multi-task learning framework, where the unaffected area can be generalized as common features. In this context, the multi-task learning method can extract features from different network topologies and generalize them as a common representation, which allows agents to learn multiple tasks simultaneously. Furthermore, the proposed multi-task framework can improve the learning performance of each individual task, compared with single-task learning methods.

In this section, we extend the above PPO algorithm from a single-task domain to a multi-task domain by learning a task-conditioned policy $\pi_{\phi_i}(a_{i,t}|o_{i,t}, u)$, where u represents a task embedding. These tasks can be sampled from a distribution $p(\mathcal{N}^{ln})$ and follow their own MDPs. Specifically, the proposed multi-task framework consists of two parts: 1) a policy network for action outputs; 2) a probability network for task connections. The detailed transformation process from single-task PPO to multi-task PPO with a probability network has been presented in Fig. 4.

1) Multi-task Policy Network: As depicted in Fig. 4, the policy network extends the actor network in Section V-A by

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

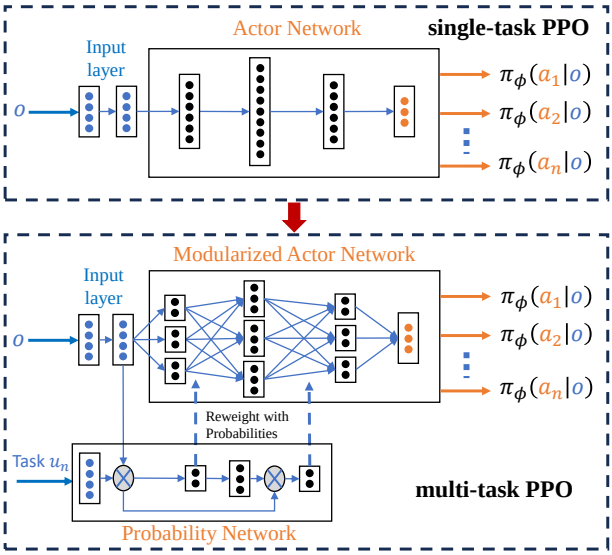


Fig. 4. Illustration of the transformation process from single-task PPO to multi-task PPO with a probability network.

involving multiple modules, taking $o_{i,t}$ as input and $a_{i,t}$ as output. The combinations of different modules in each layer are utilized to manage a specific task, denoted as u_n. The inter-module weights for task u_n are adjusted based on connection probabilities, which are learned from the probability network using u_n as input.

In the multi-task framework, the PPO policy is optimized to maximize the average expected return across all tasks sampled from $p(\mathcal{N}^{ln})$. As a result, the objective of PPO policy update (27) can be extended as

$$L_{i,t}(\phi_i) = \mathbb{E}_{t,n\sim p(\mathcal{N}^{ln})}\left[L_{i,t,n}(\phi_i)\right], \forall i \in \mathcal{I}, \tag{31}$$

where $L_{i,t,n}(\phi_i)$ is the policy update of agent i with respect to task n. We assume the policy network is constructed by $\mathcal{Z}$ layers, where each layer z contains a set of modules $m \in \mathcal{M}$. The input feature representation for m-th module in the $z+1$ layer is denoted by g_m^{z+1}, which can be calculated as

$$g_m^{z+1} = \sum\nolimits_{n\in\mathcal{M}} \hat{p}_{m,n}^z(\texttt{ReLU}(W_n^z g_n^z)), \forall z \in \mathcal{Z}, \tag{32}$$

where W_n^z corresponds to the parameters of module n in layer z. After the ReLU activation, a weighted sum of module z is calculated with $\hat{p}_{m,n}^z$, which refers to the connection probability between the n-th module in layer z and the m-th module in layer $z+1$. $\sum_{n\in\mathcal{M}} \hat{p}_{m,n}^z = 1$ holds true. Note that the connection probability $\hat{p}_{m,n}^z$ is different under different task embeddings, which will be learned from the probability network, as described in Section V-B2.

Given the module outputs of the final layer Z, the action's mean μ and variance σ can be calculated as

$$\mu, \sigma = \sum\nolimits_{m\in\mathcal{M}} W_n^Z g_n^Z, \tag{33}$$

where W_m^Z and g_n^Z refer to the parameters and feature representation of module m in the last layer Z.

2) Soft Modularization: One major difficulty in multi-task learning is how to effectively capture the mutual influence between tasks when they are trained jointly, since there is a trade-off between tasks, and optimizing one task may compromise some others. To address this issue, we introduce the soft modularization method suggested by [35] to automatically generate soft combinations between different modules without explicitly specifying the policy structure. Compared with 'hard modularization', the soft version does not introduce additional variance and can significantly stabilize RL training, leading to improved policy performance.

The function of soft modularization is achieved by a probability network, which consists of $Z-1$ layers, takes the task embedding u_n and the observation $o_{i,t}$ as input, and estimating the connection probability $p_{m,n}^z$ between module n in layer z and module m in layer $z+1$. Specifically, the probability matrix p^{z+1} for layer $z+1$ can be represented as

$$p^{z+1} = W_2^z(\sigma_\psi(W_1^z p^z \cdot (f^S(s_t) \cdot h^S(u_n)))), \tag{34}$$

where u_n is an one-hot vector to represent each task n. $f^S(\cdot)$ and $h^S(\cdot)$ represent a 2-layer Multi-layer Perceptrons (MLPs) and a fully-connected layer, respectively. W_1^z is a fully-connected layer that converts the probability matrix p^z to an embedding with the same dimension as the task embedding u_n. Then, the element-wise multiplication is conducted to obtain a new feature representation by combining the embeddings from p^z, $o_{i,t}$, and u_n. This feature is passed through the activation function σ_ψ with parameter ψ and forwarded to the fully-connected layer W_2^z, calculating the probability matrix p^{z+1} of layer $z+1$.

Finally, to weight the modules in the policy network, the softmax function is utilized to normalize p^z as

$$\hat{p}_{m,n}^z = \frac{exp(p_{m,n}^z)}{\sum_{n\in\mathcal{M}} exp(p_{m,n}^z)}, \tag{35}$$

where $\hat{p}_{m,n}^z$ is the connection probability between module n of layer z and module m of layer $z+1$, as defined in (32). In this context, given a specific task u_n, the probability network can estimate different connection strategies among modules and then reconfigure the policy network for task u_n via connection probabilities. As such, there will be different combinations of modules to specifically handle different tasks and ensure the solution quality.

C. Safety Layer

Training RL algorithms via deep neural networks is an unconstrained optimization that ignores system physical constraints. As a result, directly deploying the RL actions to the IES environment might cause constraint violations of the simulated MEMG energy flow. To address this issue, we introduce a safety layer to ensure the secure operations of MEMGs with minimum interference, which means the computed RL actions will be corrected as little as possible and only if they could endanger network safety.

In general, the module of the physics-informed safety layer includes two steps. The first step is to use the energy dispatches PGH_t^π estimated from the multi-task learning framework to check the feasibility of energy flow $\xi_{i,t}^{mg}$ in (17). If there is no constraint violation, the above energy dispatches can be

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

executed to the environment, and then the reward function (25) can be obtained. However, if constraints of energy flow $\xi_{i,t}^{mg}$ are not satisfied, the second-step correction optimization program will be performed. More specifically, a safety layer is cast on top of the multi-task learning framework to solve:

$$\underset{PGH_t^s}{\operatorname{argmin}} \frac{1}{2}\|PGH_t^{\pi} - PGH_t^s\|^2 \tag{36}$$

$$\text{s.t. } PGH_t^s \rightarrow \xi_{i,t}^{mg}, \tag{37}$$

where the objective function (36) is to find a set of secure energy dispatches PGH_t^s that perturb the original energy dispatches PGH_t^{π} in the Euclidean norm as little as possible to satisfy the physical constraints of energy flow in (37).

After the energy dispatches are corrected to PGH_t^s, the reward r_t and the next state s_{t+1} will be adjusted accordingly. In order to learn how to avoid constraint violations, we denote the deviation between PGH_t^{π} and PGH_t^s as c_t, and then reformulate the original reward function (25) as

$$r_t^s = r_t - \beta c_t, \tag{38}$$

where β denotes the significance of constraint violations with respect to load restoration. In this setting, this technique can allow agents to update secure actions straightforwardly.

It is worth noting that existing MARL methods on resilience-oriented operations [29]–[34] formulate constraint violations as penalty terms, which makes the reward function design a complicated and trivial task, and easily leads to clueless agent actions, slow convergence, and poor RL performance, especially as the number of constraints increases. Even though the reward function is well-designed, the secure MEMG operations in the training process still cannot be ensured, which is impractical in real-world applications, especially under extreme events when MEMGs are operating much closer to their stability limits due to various uncertainties and contingencies [43]. In this context, the proposed safety layer can allow agents to automatically detect the update directions that can significantly reduce constraint violations and ensure that only safe actions are executed into the environment.

D. Training Procedure

Regarding the training procedure of MT-PIPPO, each agent has its individual policy $\pi_{\phi_i}(a|o)$ that is simulated for T time steps through interactions with the environment. Then, the trajectory τ_i is collected and gathered, which can be used to estimate the discounted reward-to-go $\hat{R}_t = \sum_{h=t}^{T} \gamma^{h-t} r_h$. For each time step t, the generalized advantage function $\Lambda_{i,t}$ can also be calculated by the state-value function $V_\theta(o_{i,t})$ according to the trajectory τ_i. The actor network can be trained with the objective of maximizing the following function

$$\mathcal{L}^a(\phi_i) = \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\min\left(\zeta_{i,t}\Lambda_{i,t}, \text{clip}(\zeta_{i,t}, 1-\epsilon, 1+\epsilon)\Lambda_{i,t}\right), \tag{39}$$

Accordingly, the critic network of PPO can be trained by minimizing the following loss function of mean-squared error

$$\mathcal{L}^c(\theta_i) = \frac{1}{N}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(V_\theta(o_{i,t}) - \hat{R}_t\right)^2. \tag{40}$$

Using the objective functions mentioned above, the weights of the actor and critic networks are updated as below:

$$\phi_i \leftarrow \phi_i + \alpha^{\phi_i}\nabla_{\phi_i}\mathcal{L}^a(\phi_i), \tag{41}$$

$$\theta_i \leftarrow \theta_i + \alpha^{\theta_i}\nabla_{\theta_i}\mathcal{L}^c(\theta_i), \tag{42}$$

where $\alpha^{\phi_i}, \alpha^{\theta_i}$ denote the learning rates of actor and critic networks, respectively. The parameters ψ of the probability network can be updated according to the back-propagation approach within the learning process of parameters ϕ and θ. Finally, the pseudo-code of the proposed MT-PIPPO method is described in Algorithm 1.

Algorithm 1 MT-PIPPO for I agents

Initialize weights ϕ_i, θ_i for actor and critic networks
Set learning rates $\alpha^{\phi}, \alpha^{\theta}$
for episode (i.e., an operating day) $date = 1$ to E **do**
  Initialize both the local observation $o_{i,0}$ and global state s_0
  For MGCC agent $i \in \mathcal{I}$, create a new trajectory $\tau_i = []$
  for each time step (e.g., one hour) $t = 1$ to T **do**
    for MGCC agent $i = 1$ to I **do**
      Chooses PPO action $a_{i,t}$ according to observation $o_{i,t}$ via the task-conditioned policy $\pi_{\phi_i}(a|o,u)$
      MGCC runs the energy flow engine and corrects the energy dispatches PGH_t^s if energy flow constraints $\xi_{i,t}^{mg}$ are violated
      Execute the safe actions $a_{1:I,t}^s$ of all agents to system-wise IES environment
      Observes reward r_t^s and the next observation $o_{i,t+1}$
      Stores the sample experience into trajectory $\tau_i \mathrel{+}= [o_{i,t}, a_{i,t}^s, r_t^s]$
    end for
    Updates observation $o_{i,t} \leftarrow o_{i,t+1}$ for MGCC agent i
  end for
  for MGCC agent $i = 1$ to I **do**
    Approximates discounted reward-to-go $\hat{R}_t$ and advantage function $\Lambda_{i,t}$ utilizing trajectory τ_i
    Updates the parameters ϕ_i, θ_i of networks in (41)-(42)
  end for
end for

VI. Case Studies

A. Experimental Setup

1) Network and Data Descriptions: The proposed MT-PIPPO method is examined to solve the resilience-oriented coordination problem of multi-MEMGs towards the system-wise IES load restoration, where an integrated energy network including 33-bus PDN, 20-node GDN, and 20-node DHN with three MEMGs is deployed for experiments, as illustrated in Fig. 5. The technical parameters of controllable devices in MEMGs can be found in Table I.

To capture the environment stochasticity, we collect the yearly profiles of demand and RES generation from Ausgrid and RWTH Aachen University, then split them into the training (Jan.-Nov.) and test (Dec.) sets for MARL methods. To simulate the impact of extreme events, it is assumed that multiple

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

TABLE I
TECHNICAL PARAMETERS OF CONTROLLABLE DEVICES IN MEMGS

Device	Parameters	MEMG 1	MEMG 2	MEMG 3
GG	$\underline{P}_{i,g}^{gg}, \overline{P}_{i,g}^{gg}$	0-60 kW	0-130 kW	0-60 kW
ES	$\underline{E}_{i,k}^{es}, \overline{E}_{i,k}^{es}, \overline{P}_{i,i,k}^{es}$	0-120 kWh, 30 kW	0-100 kWh, 20 kW	0-150 kWh, 40 kW
EHP	$\overline{H}_{i,e}^{ehp}$	50 kW	50 kW	50 kW
GW	$\underline{G}_{i,g}^{gw}, \overline{G}_{i,g}^{gw}$	0-320 Sm^3/h	0-320 Sm^3/h	0-350 Sm^3/h
GS	$\underline{E}_{i,k}^{gs}, \overline{E}_{i,k}^{gs}, \overline{G}_{i,k}^{gs}$	0-160 Sm^3, 40 Sm^3/h	0-200 Sm^3, 50 Sm^3/h	0-200 Sm^3, 50 Sm^3/h
CHP	$\overline{P}_{i,g}^{chp}, \overline{H}_{i,g}^{chp}$	60 kW, 80 kW	90 kW, 120 kW	120 kW, 160 kW
TS	$\underline{E}_{i,k}^{ts}, \overline{E}_{i,k}^{ts}, \overline{H}_{i,k}^{ts}$	0-100 kWh, 20 kW	0-100 kWh, 20 kW	0-100 kWh, 20 kW
GB	$\overline{H}_{i,g}^{gb}$	100 kW	150 kW	150 kW

line faults can occur in the integrated power-gas-heat network. Specifically, a set of contingencies (i.e., multiple tasks $\mathcal{N}^{ln}$) can be generated by the Monte-Carlo sampling technique based on the fragility curve suggested by [41]. Since this paper focuses on the restoration stage of a resilience process after extreme events, the tail of distributions of extreme events is not considered, following the same practice in [38], [41]. For each episode, a task is randomly sampled from the contingency set to represent the network topology, where one example can be found in Fig. 5.

2) Benchmarks: Four benchmarks are constructed to compare with the proposed **MT-PIPPO**, including two MARL methods and two optimization methods: i) **ST-MAPPO with penalty**: MGCC agents are capable of accessing the global network information $\{P_b, Q_b, G_b, H_b : b \in \mathcal{B}\}$ and their local information, e.g., the statuses of controllable components defined in (19). The penalty terms of energy flow constraint violations are included in the reward function (25), while the multi-task learning framework is not included, i.e., network topologies will be treated as uncertainties in the training process. ii) **ST-PIPPO**: similar to the ST-MAPPO, MGCC agents can access their own local information as well as the global network information, while the multi-task learning framework is not included. However, instead of using penalty terms, a safety layer is employed. iii) **Central-Opt**: the operator can acquire all the mathematical models and technical parameters of the power-heat-gas network and multi-MEMGs and the perfect perfect perception of contingencies and system uncertainties, and then deploys a centralized and deterministic control method [17] to solve the optimization with the objective (18) subject to (1)-(17) and (43)-(59). iv) **Consensus**: MEMG agents and the IES employ an alternating direction method of multipliers [10] to solve the resilience-oriented coordination problem iteratively in a decentralized manner, assuming perfect information of the system models and technical parameters, while the system uncertainties and contingencies are captured via scenario generation and reduction techniques. In detail, the key features and the set of controllable variables/actions for the proposed MT-PIPPO and four benchmarks are summarized in Table II and Table III, respectively. The learning-based methods and the consensus-based method can only control the actions or variables of local devices in each MEMG, while the Central-Opt is capable of managing all the variables in MEMGs and system-wise IES.

TABLE II
COMPARISON OF THE PROPOSED MT-PIPPO AND FOUR BENCHMARKS

Method	Control Scheme	PPO	Safety Layer	Multi-Task Learning
MT-PIPPO	decentralized	✓	✓	✓
ST-PIPPO	decentralized	✓	✓	-
ST-MAPPO	decentralized	✓	-	-
Central-Opt	centralized	optimization	-	-
Consensus	decentralized	optimization	-	-

TABLE III
SUMMARY OF CONTROLLABLE VARIABLES/ACTIONS IN THE PROPOSED MT-PIPPO AND FOUR BENCHMARKS

Method	Controllable Variables/Actions
MT-PIPPO	$\{P_{i,g,t}^{gg}, P_{i,k,t}^{es}, G_{i,g,t}^{gw}, G_{i,k,t}^{gs}, H_{i,e,t}^{ehp}, H_{i,g,t}^{chp}, H_{i,g,t}^{gb}, H_{i,k,t}^{ts}\}$
ST-MAPPO	$\{P_{i,g,t}^{gg}, P_{i,k,t}^{es}, G_{i,g,t}^{gw}, G_{i,k,t}^{gs}, H_{i,e,t}^{ehp}, H_{i,g,t}^{chp}, H_{i,g,t}^{gb}, H_{i,k,t}^{ts}\}$
ST-PIPPO	$\{P_{i,g,t}^{gg}, P_{i,k,t}^{es}, G_{i,g,t}^{gw}, G_{i,k,t}^{gs}, H_{i,e,t}^{ehp}, H_{i,g,t}^{chp}, H_{i,g,t}^{gb}, H_{i,k,t}^{ts}\}$
Central-Opt	$\{P_{i,g,t}^{gg}, P_{i,k,t}^{es}, G_{i,g,t}^{gw}, G_{i,k,t}^{gs}, H_{i,e,t}^{ehp}, H_{i,g,t}^{chp}, H_{i,g,t}^{gb}, H_{i,k,t}^{ts}\}$, $\{P_{g,t}^{ps}, G_{g,t}^{gs}, H_{g,t}^{hs}, V_{b,t}, \rho_{b,t}, \varrho_{b,t}, \tau_{b,t}, P_{l,t}, G_{l,t}, H_{l,t}\}$
Consensus	$\{P_{i,g,t}^{gg}, P_{i,k,t}^{es}, G_{i,g,t}^{gw}, G_{i,k,t}^{gs}, H_{i,e,t}^{ehp}, H_{i,g,t}^{chp}, H_{i,g,t}^{gb}, H_{i,k,t}^{ts}\}$

3) Implementations: For all MARL methods, actor and critic networks are updated via the Adam optimizer, where the learning rates are $\alpha^{\phi} = 10^{-4}$ and $\alpha^{\theta} = 10^{-3}$, respectively. The discount factor is set as $\gamma = 0.99$ to expect a daily dynamic return of 24 time steps, while the clip rate ϵ is 0.2. MLPs are used to construct the two hidden layers (400 and 300 neurons) for both actor and critic networks. A Gaussian policy with mean and standard deviation is constructed by `Softplus` and `Tanh` activation functions for the actor network. We run 30,000 episodes with 10 random seeds, where each episode represents one operation day that is uniformly sampled from the training set.

B. Performance Evaluation

This section evaluates the training and test performance of the proposed MT-PIPPO and three benchmarks (ST-MAPPO, ST-PIPPO, and Central-Opt). Fig. 6(a) shows the episodic reward evolution of 3 MEMGs over 30,000 training episodes, where the shaded areas and the solid lines respectively correspond to the oscillations of 10 random seeds and the moving average over 100 episodes. Furthermore, we plot the cumulative resilience level of 3 MEMGs over the test month

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

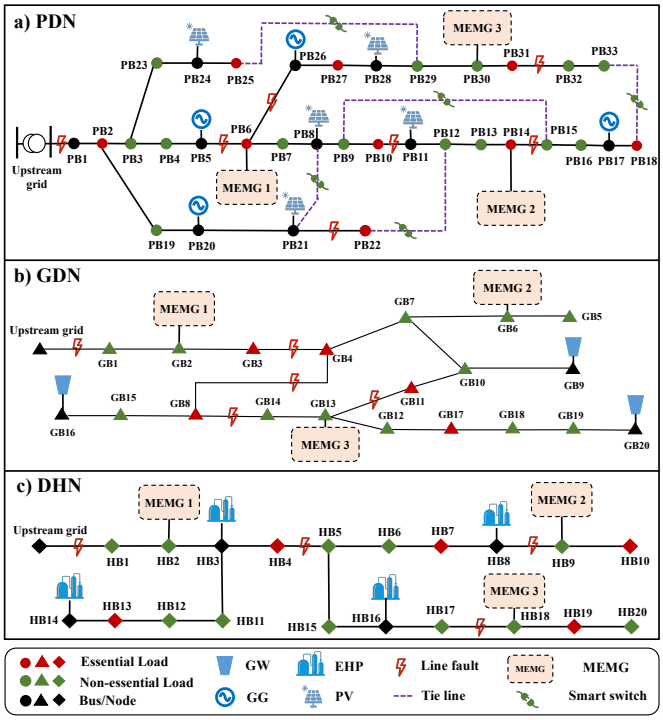
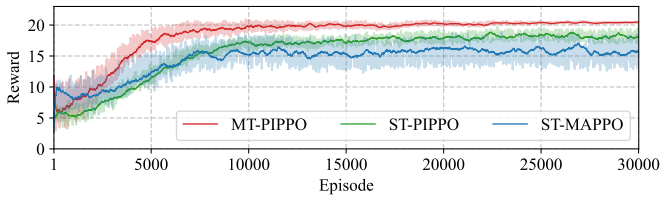
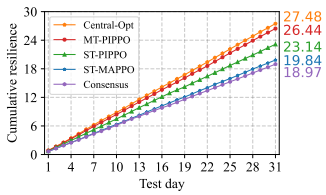


Fig. 5. The studied IES including 33-bus PDN, 20-node GDN, and 20-node DHN with three MEMGs.

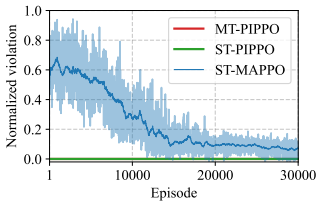
in Fig. 6(b) by deploying the trained MARL (actor) policies. Finally, we record the normalized constraint violations of three MARL methods in the RL training procedure in Fig. 6(c).



(a) Training reward



(b) Cumulative test resilience level



(c) Constraint violations

Fig. 6. Episodic training reward over 30,000 episodes (a), the cumulative test resilience level over 31 days (b), the episodic constraint violations (c) for different MARL methods.

The first finding from Fig. 6(a) is that ST-MAPPO (blue) and ST-PIPPO (green) are both relatively unstable compared to MT-PIPPO (red), and exhibit oscillatory training performance. This can be expected since single-task learning algorithms of ST-MAPPO and ST-PIPPO can not deal with changing network topologies caused by severe contingencies, which makes the learning very challenging. To this end, our proposed MT-PIPPO is capable of addressing this issue due to its efficient MT learning framework. Overall, it can be observed that MT-PIPPO achieves the highest reward level (20.46) and the most stable learning performance of three MARL methods.

TABLE IV
THE AVERAGED RESILIENCE LEVELS FOR DIFFERENT MARL AND OPTIMIZATION METHODS

Rank	Method	Resilience Index (%)	Percentage Variation (%)
1	Central-Opt	88.6	-
2	MT-PIPPO	85.3	3.7
3	ST-PIPPO	74.6	15.8
4	ST-MAPPO	64.0	27.8
5	Consensus	61.2	30.9

Similar to the training performance, as shown in Fig. 6(b), the proposed MT-PIPPO in the test process attains a close-to-optimal performance (3.78% lower than the benchmark Central-Opt), while significantly outperforming ST-PIPPO, ST-MAPPO, and Consensus by 12.48%, 24.96%, and 28.25% the cumulative resilience level over the test month (31 days), respectively. It is worth noting that the selected test month includes 31 days with different load levels, renewable penetration, and contingencies (line faults), which further verifies the effectiveness of the proposed MT-PIPPO in dealing with different network topologies, ensuring secure system operation, and achieving near-optimal solutions. Additionally, all three MARL methods can be implemented in real-time within 1 sec.; nevertheless, the Central-Opt and Consensus methods require 97.75 sec. and 876.12 sec. to solve a daily optimization problem, respectively. Furthermore, the averaged resilience levels and ranks over 31 test days for different MARL and optimization methods are summarized in Table IV, highlighting the advantages of the proposed MT-PIPPO in achieving near-optimal solutions compared with other decentralized methods.

Finally, it can be interesting to analyze the normalized constraint violations during training process. As depicted in Fig. 6(c), ST-MAPPO (blue) obtains large fluctuations of constraint violations, even though exhibiting a decreasing trend. Note that extreme events are characterized by high impact, leading to various uncertainties and contingencies. In this context, MEMG operations can be much closer to their security boundaries; therefore, using conventional RL methods with penalty terms can easily lead to large constraint violations and unsafe system operations. On the other hand, the constraint violations of ST-PIPPO (green) and MT-PIPPO (red) are completely zero. This is because the physics-informed safety layer employs a two-step correction optimization program, which can correct unsafe RL actions with minimum interference and ensure that only safe actions are executed in the RL training process.

To conclude, the advantages of the proposed MT-PIPPO method over conventional model-based optimization techniques can be summarized as below: 1) compared to model-based methods that are typically time-consuming, the well-trained RL policy allows for real-time decision making, which is a critical factor given the high-impact nature of extreme events and the requirement for fast response; 2) compared to model-based methods that normally require intensive communication resources, the proposed MARL method can operate

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

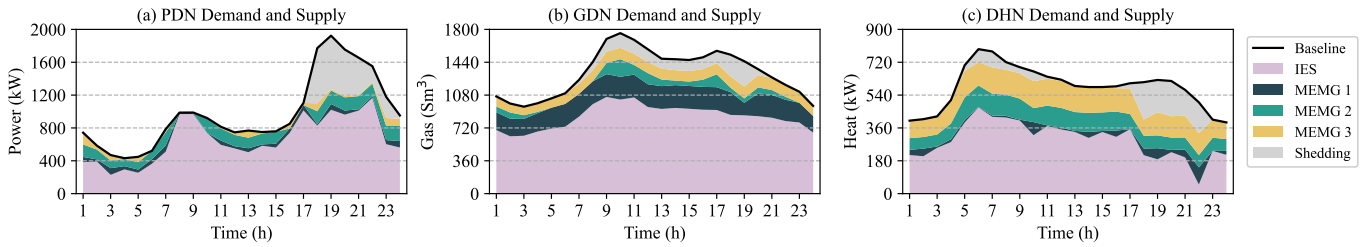


Fig. 7. Demand-supply balances of the integrated energy system with three MEMGs.

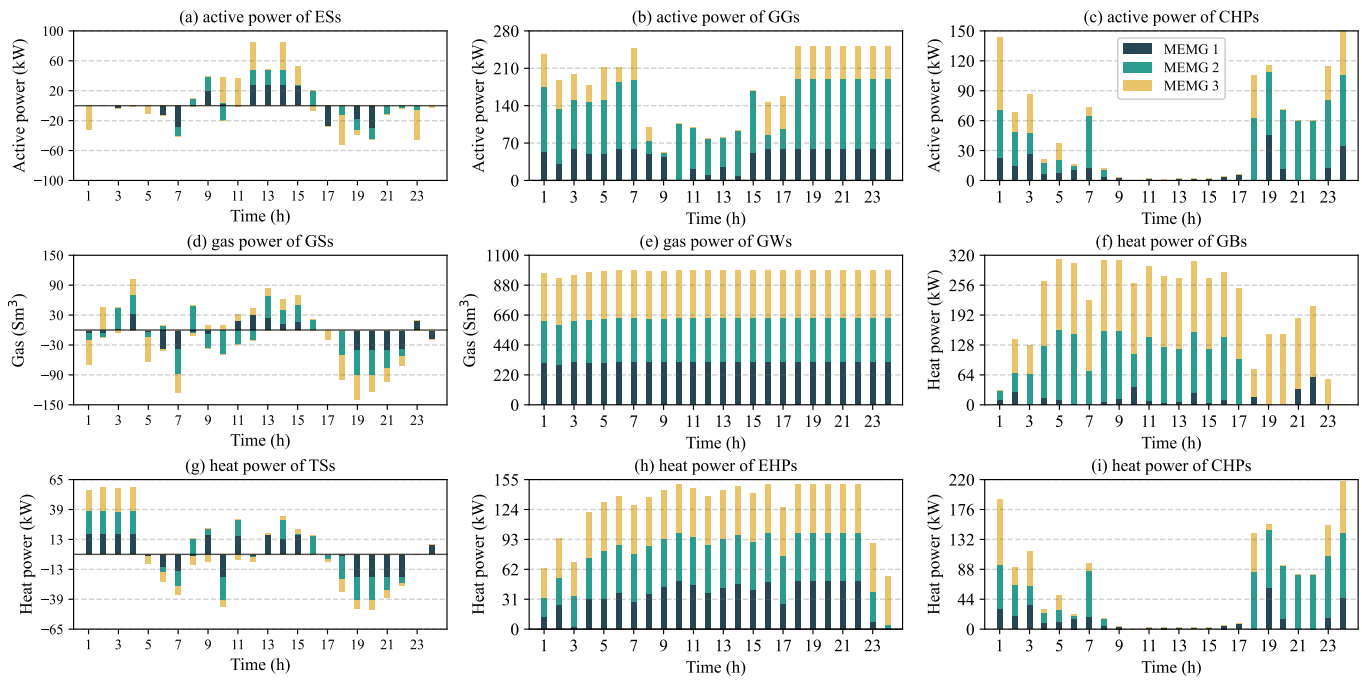


Fig. 8. Energy scheduling behaviors of controllable components in MEMGs.

MEMGs in a decentralized fashion without interactions among agents, which can ensure the uninterrupted load restoration even under potentially degraded communication structures after extreme events; 3) compared to model-based methods that lack of the ability to adapt to changing system conditions caused by uncertainties and contingencies, the proposed MT-PIPPO method employs a data-driven fashion and a multi-task learning framework, which can assist MEMG agents in learning and adapting to uncertain system features; 4) as a learning method, the proposed MT-PIPPO can achieve near-optimal solutions without constraint violations, compared to the benchmark Central-Opt.

C. MEMG Scheduling Analysis

This section focuses on validating the well-trained policy of MT-PIPPO for the operating characteristics of MEMGs, consisting of the energy exchanges with the network and the energy schedules of controllable components within the MEMGs, as depicted in Fig. 7 and 8, respectively. Furthermore, to estimate the contribution of each MEMG to the system load restoration, the demand-supply balances of PDN, GDN, and DHN are also presented in Fig. 7.

1) Energy Exchange with the Network: It can be observed from Fig. (7) that all three MEMGs send energy (i.e., power, gas, and heat) to the network towards effective load restoration. In particular, MEMG 1 mainly provides gas supply for GDN due to the severe line fault occurring at the gas pipeline $3-4$, as shown in Fig. 5. Note that this line fault causes the isolation of gas loads at nodes 1-3 from GDN, including one essential load that can only be supplied by MEMG 1. On the other hand, both MEMG 2 and MEMG 3 focus on DHN because of the line faults occurring at heat pipelines $8-9$ and $17-18$, respectively. Except for supplying DHN, MEMG 2 provides the most power supply for PDN than the two other MEMGs, since loads at buses 11-14 (including one essential load at bus 14) are isolated from the rest of PDN and MEMG 2 has the largest power capacity in this island. In this context, the contribution of each MEMG to the system-wise IES load restoration is presented in Fig. 7. An interesting finding is that the heat supply from DHN itself decreases during the evening. The reason is that the DHN is mainly supplied by EHPs, which are powered by PDN. When the peak demand occurs during the evening, PDN uses more energy for its own load supply rather than sending energy to DHN.

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

2) Energy Schedules of Controllable Components: In response to effective energy supply, these MEMGs optimally schedule their controllable components, as depicted in Fig. 8. Fig. 8(a), (d), and (g) show that battery units (ESs, GSs, and TSs) charge during the day and discharge in the evening to balance supply and demand towards load restoration. Specifically, ESs and GSs charge power in the midday, because of the high level of renewable generation. Note that a high RES level can provide much power supply for PDN in the midday and then reduce the gas consumption of GG units, allowing more inflation behaviors of GSs. As shown in Fig. 8(b) and (c), GGs and CHPs are providing power supply for PDN in the early morning and evening, especially in the evening due to the peak electric demand level. Accordingly, CHPs provide heat supply for DHN in the early morning and evening due to the coupling effect, as presented in Fig. 8(i). Furthermore, Fig. 8(f) and (h) present that GBs and CHPs supply heat energy to the DHN during the day, especially in the midday, as the high level of renewable generation in PDN allows more gas input for DHN. Finally, GWs in Fig. 8(e) are fully used during the day, due to the importance of gas energy in the IES, i.e., supplying all the PDN, GDN, and DHN.

D. Test Results in a Larger Integrated Energy Network

This section serves as a further demonstration of the proposed MT-PIPPO method on scalability through a larger IES including modified IEEE 69-bus PDN, 40-node GDN, and 62-node DHN [17], [43], [44], while 6 MEMGs are connected with the IES. The network structure of the IES with one selected event scenario (multiple line faults) is illustrated in Fig. 9.

The demand-supply balances of PDN, GDN, and DHN in the large-scale IES are presented in Fig. 10. Specifically, it can be found that MEMG 1 and MEMG 5 mainly focus on providing power supply for the PDN, especially for the essential loads on bus 18 and bus 61, respectively. This is because the severe line faults occurring on lines $PB14-PB15$ and $PB19-PB20$ as well as lines $PB60-PB61$ and $PB63-PB64$ cause two islanded areas without enough power supply, i.e., the area around buses 15-19 and the area around buses 61-63. In this case, it is necessary to employ MEMG 1 and MEMG 5 as load restoration resources to support the essential loads on bus 18 and bus 61, respectively. On the other hand, MEMG 2 and MEMG 3 focus on providing gas supply for the GDN. The main reason is that the severe line faults on pipelines $GB2-GB13$ and $GB13-GB14$ as well as $GB18-GB19$ cause two islanded areas that need gas supply, i.e., gas nodes 11-13 and 19-20, respectively. Similarly, MEMG 4 and MEMG 6 mainly focus on providing heat supply for the DHN, due to the severe line faults happening on heat pipelines $HB1-HB4$ and $HB4-HB47$ as well as $HB16-HB20$, respectively.

In summary, the well-trained MARL policy can guide these MEMGs to observe the contingency information, effectively schedule their local resources, and then provide sufficient energy supply for the system-wise IES in real time towards overall load restoration. These results further validate the effectiveness of the proposed MT-PIPPO method in supporting system-wise IES resilience as well as demonstrating the scalability and adaptability of the proposed method across varying network sizes and topologies.

VII. Conclusions

A novel MARL method called MT-PIPPO is developed in this paper to solve the resilience-oriented coordination problem of multi-MEMGs for system-wise IES load restoration. The proposed MT-PIPPO algorithm is characterized by a multi-task learning framework and a safety layer, which can effectively improve the topology awareness of MEMGs against unanticipated contingencies. The decentralized coordination paradigm among multi-MEMGs is realized in a Dec-POMDP framework, which takes into account various system dynamics and uncertainties, such as renewable generation and energy demand. Experiments are conducted on two large-scale integrated power-gas-heat networks (33-bus PDN, 20-node GDN, and 20-node DHN as well as 69-bus PDN, 40-node GDN, and 62-node DHN) to evaluate the superiority of the proposed MARL method in optimality and stability compared to optimization methods and state-of-the-art MARL methods. Finally, the coordination effects of multi-MEMGs on the system-wise IES resilience provision are analyzed.

In detail, case studies developed in this paper evaluate the performance of the proposed MT-PIPPO method with multi-task learning framework and safety layer and analyze the scheduling results of MEMGs for the load restoration of IESs after extreme events. Overall, the key findings can be summarized as below:

(1) The training and test performance of the proposed MARL method have been evaluated in Section VI-B. On one hand, compared with single-task learning method, the proposed multi-task learning framework can address the non-stability issue caused by changing network topologies and achieve better training performance. On the other hand, the proposed safety layer can significantly reduce constraint violations and ensure secure MEMG operations. Finally, MT-PIPPO can achieve near-optimal solutions (only 3.78% lower than the benchmark Central-Opt) and can be deployed in real time within 1 sec.

(2) The scheduling behaviors of various distributed components (e.g., ES, GS, TS, EHP, CHP, etc.) in each MEMG have been well-learned by the proposed MT-PIPPO, as presented in Section VI-C. The well-trained RL policy can guide MEMG agents to effectively schedule their local resources, coordinate with each other, and provide energy supply for the system-wise IES towards resilience enhancement.

(3) The scalability of the proposed MT-PIPPO has been demonstrated in two large-scale IESs, including 33-bus PDN, 20-node GDN, and 20-node DHN with 3 MEMGs as well as 69-bus PDN, 40-node GDN, and 62-node DHN with 6 MEMGs, as presented in Section VI-C and Section VI-D, respectively. Due to the superiority of the multi-task learning framework and safety layer, the proposed MT-PIPPO can still handle the uncertainties (renewable

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

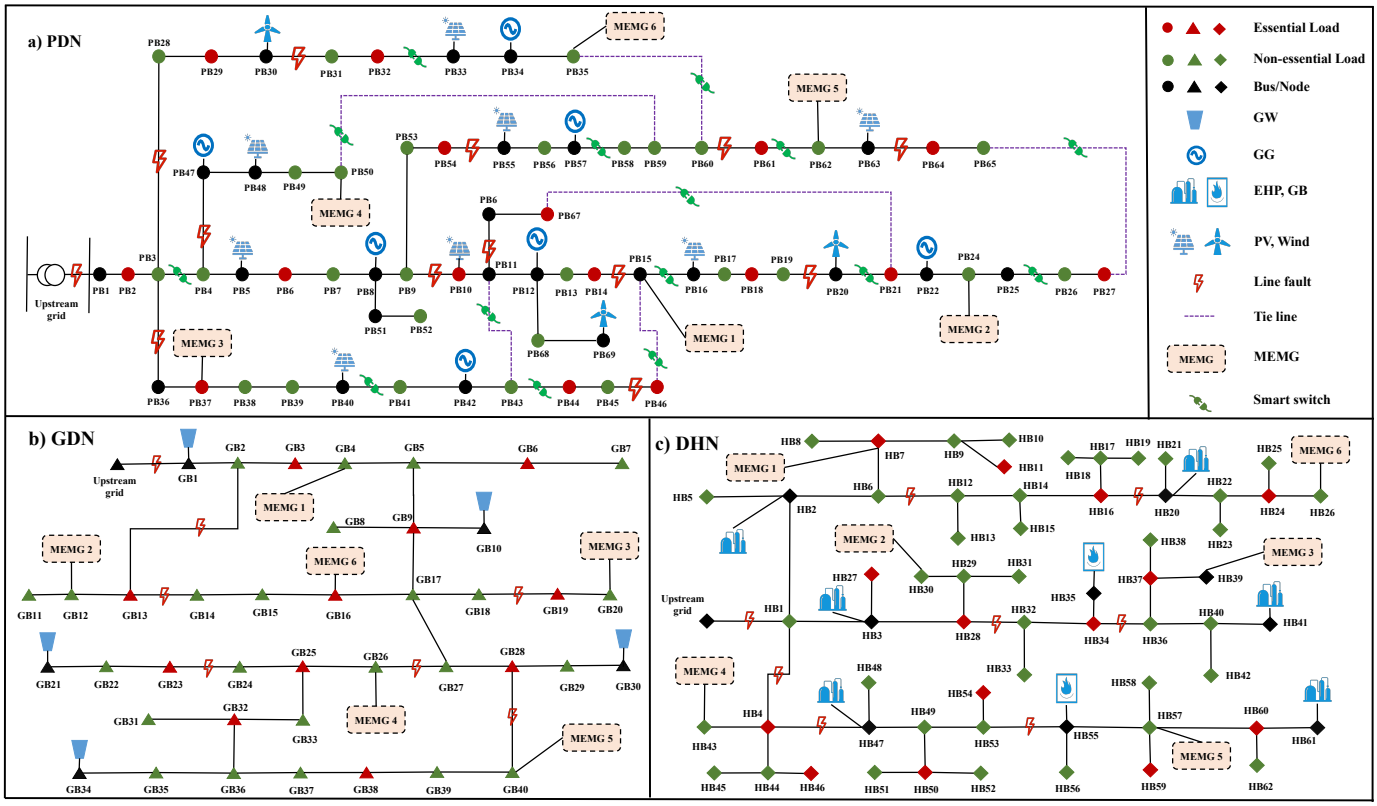


Fig. 9. The studied larger IES including 69-bus PDN, 40-node GDN, and 62-node DHN with 6 MEMGs.

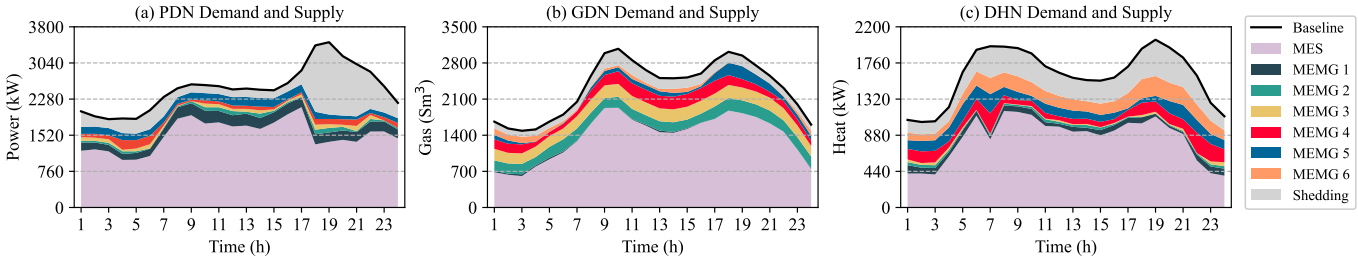


Fig. 10. Demand-supply balances of the larger integrated energy system with 6 MEMGs.

energy and demand) and contingencies (different network topologies) in a large-scale IES and then learn realistic MEMG scheduling behaviors.

As aforementioned, resilience-oriented system operations require fast response time, reduced dependence on communication infrastructures, and secure system operation. The proposed MT-PIPPO method can address the non-stability issues caused by uncertainties associated with renewable energy and demand as well as contingencies related to changing network topologies, while achieving real-time and decentralized decision making of multi-MEMGs without constraint violations. These aspects are very important for deploying MARL methods in real-world applications towards resilience enhancement. However, it is worth noting that RL-based methods normally require a large amount of data for effective training. Potential issues with data quantity, data quality, data availability, and data privacy can be challenges in the real-world applications of the proposed MARL method, requiring the incorporation of advanced blockchain and cyber security techniques. In addition, the communication delays caused by time-varying or asynchronous communication should also be considered for more realistic decision making. A detailed communication network can be integrated into the proposed real-time control method, facilitating data exchanges between MEMG agents and the environment, as well as enabling the execution process of MEMGs' scheduling behaviors. Finally, extreme events are featured by high impact and low probability, representing the tail of a probability distribution function [37]. To better account for this issue, future work will focus on developing a risk-averse planning approach based on Value at Risk or Conditional Value at Risk for the resilience-oriented coordination of multi-MEMGs.

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

APPENDIX A
INTEGRATED ENERGY SYSTEM OPERATION

The IES operation including PDN, GDN, and DHN can be modeled by the OEF algorithm, which is detailed as follows:

A. Power Network Model

The PDN operation is modeled by the linearized DistFlow method [38], which is realized as

$$\begin{aligned}\sum_{g\in B_{ps}} P_{g,t}^{ps} + \sum_{i\in B_{mg}} P_{i,t}^{mg} = \sum_{d\in B_{pd}} P_{d,t}^{pd} \\ - \sum_{(p,b)\in\mathcal{PL}} P_{pb,t} + \sum_{(b,p)\in\mathcal{PL}} P_{bp,t}, \forall b\in\mathcal{PB},\end{aligned} \tag{43}$$

$$\begin{aligned}\sum_{g\in B_{ps}} Q_{g,t}^{ps} + \sum_{i\in B_{mg}} Q_{i,t}^{mg} = \sum_{d\in B_{pd}} Q_{d,t}^{pd} \\ - \sum_{(p,b)\in\mathcal{PL}} Q_{pb,t} + \sum_{(b,p)\in\mathcal{PL}} Q_{bp,t}, \forall b\in\mathcal{PB},\end{aligned} \tag{44}$$

$$\underline{V}^2 \le V_{b,t}^2 \le \overline{V}^2, \forall b\in\mathcal{PB}, \tag{45}$$

$$P_{bp,t}^2 + Q_{bp,t}^2 \le y_{bp,t}\cdot \overline{S}_{bp}, \forall (b,p)\in\mathcal{PL}, \tag{46}$$

$$\begin{aligned}V_{b,t}^2 - V_{p,t}^2 \le 2\cdot(r_{bp}\,P_{bp,t} + x_{bp}\,Q_{bp,t}) \\ + (1-y_{bp,t})\cdot M, \forall (b,p)\in\mathcal{PL},\end{aligned} \tag{47}$$

$$\begin{aligned}V_{b,t}^2 - V_{p,t}^2 \ge 2\cdot(r_{bp}\,P_{bp,t} + x_{bp}\,Q_{bp,t}) \\ + (y_{bp,t}-1)\cdot M, \forall (b,p)\in\mathcal{PL},\end{aligned} \tag{48}$$

where active and reactive power balances at bus b are presented in (43)-(44). B_{pd}, B_{ps}, and B_{mg} refer to the sets of loads, power sources, and connected MEMGs at bus b. Nodal voltage levels and line power flows are limited by (45) and (46), respectively. The linearized power flow constraints are shown in (47)-(48), which are relaxed by the big-M method. PDN can be reconfigured by switch operations (i.e., binary $y_{bp,t}$ in (46)) respecting network radiality, where the detailed reconfiguration model is expressed in [38].

B. Gas Network Model

The GDN operation is modeled by the steady gas flow algorithm [11], which can be realized as

$$\begin{aligned}\sum_{g\in B_{gs}} G_{g,t}^{gs} + \sum_{i\in B_{mg}} G_{i,t}^{mg} = \sum_{d\in B_{gd}} G_{d,t}^{gd} \\ + \sum_{pb\in\mathcal{GL}} G_{pb,t} - \sum_{bp\in\mathcal{GL}} G_{bp,t}, \forall b\in\mathcal{GB},\end{aligned} \tag{49}$$

$$G_{bp,t}^2 = sign(G_{bp,t})\Pi_{bp}^2(\rho_{b,t}-\rho_{p,t}), \forall (b,p)\in\mathcal{GL}, \tag{50}$$

$$\underline{\rho}_b \le \rho_{b,t} \le \overline{\rho}_b,\ \forall b\in\mathcal{GB}, \tag{51}$$

$$\underline{G}_{bp} \le G_{bp,t} \le \overline{G}_{bp},\ \forall (b,p)\in\mathcal{GL}, \tag{52}$$

where the gas balance at node b is expressed in (49). B_{gd}, B_{gs}, and B_{mg} refer to the sets of gas loads, gas sources, and MEMGs located at node b, respectively. The relation between nodal gas pressure and gas flow at pipeline (b,p) is represented by (50), given the Weymouth constant Π_{bp}. Nodal gas pressure and gas flow are limited by (51)-(52), respectively.

C. Heat Network Model

The DHN operation is modeled by the variable temperature-variable flow algorithm [11], which can be realized as

$$\begin{aligned}\sum_{g\in B_{hs}} H_{g,t}^{hs} + \sum_{i\in B_{mg}} H_{i,t}^{mg} = \sum_{d\in B_{hd}} H_{d,t}^{hd} \\ + \sum_{bp\in\mathcal{HL}} c^w H_{bp,t}(\tau_{bp,t}^{in} - \tau_{bp,t}^{out}), \forall b\in\mathcal{HB},\end{aligned} \tag{53}$$

$$H_{bp,t}^2 = sign(H_{bp,t})\Phi_{bp}^2(\varrho_{b,t}^2 - \varrho_{p,t}^2), \forall (b,p)\in\mathcal{HL}, \tag{54}$$

$$\tau_{bp,t}^{out} - \tau^a = (\tau_{bp,t}^{in} - \tau^a)\Theta_{bp}, \forall (b,p)\in\mathcal{HL}, \tag{55}$$

$$\tau_{b,t}^{mix} \sum_{pb\in\mathcal{HL}} H_{pb,t} = \sum_{pb\in\mathcal{HL}} (H_{pb,t}\tau_{pb,t}^{out}), \forall b\in\mathcal{HB}, \tag{56}$$

$$\tau_{pb,t}^{in} = \tau_{b,t}^{mix}, \forall b\in\mathcal{HB}, \forall (p,b)\in\mathcal{HL}, \tag{57}$$

$$\underline{\varrho}_b \le \varrho_{b,t} \le \overline{\varrho}_b, \forall b\in\mathcal{HB}, \tag{58}$$

$$\underline{\tau}_b \le \tau_{b,t}^{mix} \le \overline{\tau}_b, \forall b\in\mathcal{HB}, \tag{59}$$

where heat balance and line water flow are modeled in (53)-(54). B_{hd}, B_{hs}, and B_{mg} refer to the sets of heat loads, heat sources, and MEMGs located at node b, respectively. Heat loss between hot water in the pipeline $\tau_{bp,t}^{in}/\tau_{bp,t}^{out}$ and ambient temperature τ^a is modeled in (55), while the nodal temperature $\tau_{b,t}^{mix}$ is calculated by (56). Φ_{bp} and Θ_{bp} are the Weymouth constant and loss coefficient of pipeline (b,p), respectively. Mass flowing temperature and nodal temperature are linked by (57), while water pressure and temperature are limited by (58) and (59), respectively.

REFERENCES

[1] T. Ding, W. Jia, M. Shahidehpour, O. Han, Y. Sun, and Z. Zhang, "Review of optimization methods for energy hub planning, operation, trading, and control," *IEEE Trans. Sustain. Energy*, vol. 13, no. 3, pp. 1802–1818, Jul. 2022.

[2] Y. Wang, A. O. Rousis, and G. Strbac, "On microgrids and resilience: A comprehensive review on modeling and operational strategies," *Renew. Sust. Energ. Rev.*, vol. 134, p. 110313, Dec. 2020.

[3] Y. Wang, Y. Xu, J. He, C.-C. Liu, K. P. Schneider, M. Hong, and D. T. Ton, "Coordinating multiple sources for service restoration to enhance resilience of distribution systems," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5781–5793, Sept. 2019.

[4] S. Yao, P. Wang, and T. Zhao, "Transportable energy storage for more resilient distribution systems with multiple microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3331–3341, May 2019.

[5] Z. Wang, B. Chen, J. Wang, and C. Chen, "Networked microgrids for self-healing power systems," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 310–319, Jan. 2016.

[6] F. Shen, Q. Wu, J. Zhao, W. Wei, N. D. Hatziargyriou, and F. Liu, "Distributed risk-limiting load restoration in unbalanced distribution systems with networked microgrids," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 4574–4586, Nov. 2020.

[7] W. Huang, X. Zhang, K. Li, N. Zhang, G. Strbac, and C. Kang, "Resilience oriented planning of urban multi-energy systems with generalized energy storage sources," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2906–2918, Jul. 2022.

[8] R.-P. Liu, S. Lei, C. Peng, W. Sun, and Y. Hou, "Data-based resilience enhancement strategies for electric-gas systems against sequential extreme weather events," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5383–5395, Nov. 2020.

[9] Y. Lin, B. Chen, J. Wang, and Z. Bie, "A combined repair crew dispatch problem for resilient electric and natural gas system considering reconfiguration and dg islanding," *IEEE Trans. Power Syst.*, vol. 34, no. 4, pp. 2755–2767, Jul. 2019.

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

[10] G. Li, K. Yan, R. Zhang, T. Jiang, X. Li, and H. Chen, "Resilience-oriented distributed load restoration method for integrated power distribution and natural gas systems," *IEEE Trans. Sustain. Energy*, vol. 13, no. 1, pp. 341–352, Jan. 2022.

[11] V. S. Saravi, M. Kalantar, and A. Anvari-Moghaddam, "Resilience-constrained expansion planning of integrated power–gas–heat distribution networks," *Appl. Energy*, vol. 323, p. 119315, Oct. 2022.

[12] H. Masrur, M. Shafie-Khah, M. J. Hossain, and T. Senjyu, "Multi-energy microgrids incorporating ev integration: Optimal design and resilient operation," *IEEE Trans. Smart Grid*, vol. 13, no. 5, pp. 3508–3518, Sept. 2022.

[13] M. Z. Gargari, M. T. Hagh, and S. G. Zadeh, "Preventive scheduling of a multi-energy microgrid with mobile energy storage to enhance the resiliency of the system," *Energy*, vol. 263, p. 125597, Jan. 2023.

[14] M. H. Amirioun, F. Aminifar, and M. Shahidehpour, "Resilience-promoting proactive scheduling against hurricanes in multiple energy carrier microgrids," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 2160–2168, May 2019.

[15] S. E. Ahmadi, M. Marzband, A. Ikpehai, and A. Abusorrah, "Optimal stochastic scheduling of plug-in electric vehicles as mobile energy storage systems for resilience enhancement of multi-agent multi-energy networked microgrids," *J. Energy Storage*, vol. 55, p. 105566, Nov. 2022.

[16] H. Masrur, M. M. Gamil, M. R. Islam, K. M. Muttaqi, M. H. Lipu, and T. Senjyu, "An optimized and outage-resilient energy management framework for multicarrier energy microgrids integrating demand response," *IEEE Trans. Ind. Appl.*, vol. 58, no. 3, pp. 4171–4180, May-Jun. 2022.

[17] M. Yan, Y. He, M. Shahidehpour, X. Ai, Z. Li, and J. Wen, "Coordinated regional-district operation of integrated energy systems for resilience enhancement in natural disasters," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4881–4892, Sept. 2018.

[18] R. Hemmati, H. Mehrjerdi, and S. M. Nosratabadi, "Resilience-oriented adaptable microgrid formation in integrated electricity-gas system with deployment of multiple energy hubs," *Sustainable Cities and Society*, vol. 71, p. 102946, Aug. 2021.

[19] Q. Sun, Z. Wu, Z. Ma, W. Gu, X.-P. Zhang, Y. Lu, and P. Liu, "Resilience enhancement strategy for multi-energy systems considering multi-stage recovery process and multi-energy coordination," *Energy*, vol. 241, p. 122834, Feb. 2022.

[20] X. Li, X. Du, T. Jiang, R. Zhang, and H. Chen, "Coordinating multi-energy to improve urban integrated energy system resilience against extreme weather events," *Appl. Energy*, vol. 309, p. 118455, Mar. 2022.

[21] E. A. Javadi, M. Joorabian, and H. Barati, "A bi-level optimization framework for resilience enhancement of electricity and natural gas networks with participation of energy hubs," *Int. J. Electr. Power Energy Syst.*, vol. 142, p. 108312, Nov. 2022.

[22] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[23] J. C. Bedoya, Y. Wang, and C.-C. Liu, "Distribution system resilience under asynchronous information using deep reinforcement learning," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4235–4245, Sept. 2021.

[24] Y. Huang, G. Li, C. Chen, Y. Bian, T. Qian, and Z. Bie, "Resilient distribution networks by microgrid formation using deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4918–4930, Nov.

[25] J. Zhao, F. Li, S. Mukherjee, and C. Sticht, "Deep reinforcement learning based model-free on-line dynamic multi-microgrid formation to enhance resilience," *IEEE Trans. Smart Grid*, vol. 13, no. 4, pp. 2557–2567, Jul. 2022.

[26] M. M. Hosseini and M. Parvania, "Resilient operation of distribution grids using deep reinforcement learning," *IEEE Trans. Industr. Inform.*, vol. 18, no. 3, pp. 2100–2109, Mar. 2021.

[27] Y. Du and D. Wu, "Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids," *IEEE Trans. Sustain. Energy.*, vol. 13, no. 2, pp. 1062–1072, Apr. 2022.

[28] Y. Gao, W. Wang, J. Shi, and N. Yu, "Batch-constrained reinforcement learning for dynamic distribution network reconfiguration," *IEEE Trans. Smart Grid*, vol. 11, no. 6, pp. 5357–5369, Nov. 2020.

[29] T. Wu, J. Wang, X. Lu, and Y. Du, "Ac/dc hybrid distribution network reconfiguration with microgrid formation using multi-agent soft actor-critic," *Appl. Energy*, vol. 307, p. 118189, Feb. 2022.

[30] T. Zhao and J. Wang, "Learning sequential distribution system restoration via graph-reinforcement learning," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1601–1611, Mar. 2022.

[31] M. Kamruzzaman, J. Duan, D. Shi, and M. Benidris, "A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5525–5536, Nov. 2021.

[32] Y. Wang, D. Qiu, and G. Strbac, "Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems," *Appl. Energy*, vol. 310, p. 118575, Mar. 2022.

[33] D. Qiu, Y. Wang, T. Zhang, M. Sun, and G. Strbac, "Hybrid multiagent reinforcement learning for electric vehicle resilience control towards a low-carbon transition," *IEEE Trans. Industr. Inform.*, vol. 18, no. 11, pp. 8258–8269, Nov. 2022.

[34] ——, "Hierarchical multi-agent reinforcement learning for repair crews dispatch control towards multi-energy microgrid resilience," *Appl. Energy*, vol. 336, p. 120826, Apr. 2023.

[35] R. Yang, H. Xu, Y. Wu, and X. Wang, "Multi-task reinforcement learning with soft modularization," *Adv. Neural Inf. Process Syst.*, vol. 33, pp. 4767–4777, 2020.

[36] I. P. T. Force, A. Stanković, K. Tomsovic, F. De Caro, M. Braun, J. Chow, N. Äukalevski, I. Dobson, J. Eto, B. Fink *et al.*, "Methods for analysis and quantification of power system resilience," *IEEE Trans. Power Syst.*, 2022.

[37] R. Moreno, M. Panteli, P. Mancarella, H. Rudnick, T. Lagos, A. Navarro, F. Ordonez, and J. C. Araneda, "From reliability to resilience: Planning the grid against the extremes," *IEEE Power Energy Mag.*, vol. 18, no. 4, pp. 41–53, Jul.-Aug. 2020.

[38] T. Ding, Z. Wang, M. Qu, Z. Wang, and M. Shahidehpour, "A sequential black-start restoration model for resilient active distribution networks," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 3133–3136, Jul. 2022.

[39] F. A. Oliehoek and C. Amato, *A concise introduction to decentralized POMDPs*. Springer, 2016.

[40] Z. Bie, Y. Lin, G. Li, and F. Li, "Battling the extreme: A study on the power system resilience," *Proc. IEEE*, vol. 105, no. 7, pp. 1253–1266, Nov. 2017.

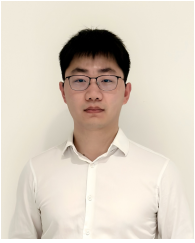
[41] S. Lei, C. Chen, H. Zhou, and Y. Hou, "Routing and scheduling of mobile power sources for distribution system resilience enhancement," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5650–5662, Sept. 2019.

[42] Y. Wang, A. O. Rousis, and G. Strbac, "A three-level planning model for optimal sizing of networked microgrids considering a trade-off between resilience and cost," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5657–5669, Apr. 2021.

[43] Y. Wang, D. Qiu, F. Teng, and G. Strbac, "Towards microgrid resilience enhancement via mobile power sources and repair crews: A multi-agent reinforcement learning approach," *IEEE Trans. Power Syst.*, Jan. 2023.

[44] H. Tan, W. Yan, Z. Ren, Q. Wang, and M. A. Mohamed, "A robust dispatch model for integrated electricity and heat networks considering price-based integrated demand response," *Energy*, vol. 239, p. 121875, Jan. 2022.

Yi Wang (Member, IEEE) received the B.Eng. degree and the M.Eng. degree from Tianjin University in 2015 and 2018, and the Ph.D. degree from Imperial College London in 2022. He is currently employed as a Research Associate in the Department of Electrical and Electronic Engineering at Imperial College London. His research interests include mathematical programming and learning approaches applied to the planning and operation of networked microgrids, the resilience enhancement of future power systems, and multi-energy system integration.



Dawei Qiu (Member, IEEE) received the B.Eng. degree from Northumbria University in 2014, the M.Sc. degree from University College London in 2015, and the Ph.D. degree from Imperial College London in 2020.

He is currently employed as a Research Associate in the Department of Electrical and Electronic Engineering at Imperial College London. His research focuses on the development and application of decentralized and market-driven approaches to electricity market, peer-to-peer energy trading, multi-energy system, microgrid resilience, and vehicle-to-grid flexibility.

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for publication in IEEE Transactions on Sustainable Energy. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2023.3317133

Xiaotian Sun (Student Member, IEEE) received the B.S. degree from the School of Electrical Engineering in Xi'an Jiaotong University, Xi'an, China, in 2020, where he is currently pursuing the Ph.D. degree. His major research interests include the electricity market, game theory, and energy system resilience.

Zhaohong Bie (Fellow, IEEE) received the B.S. and M.S. degrees in Electric Power from Shandong University, Jinan, China, in 1992 and 1994, respectively, and the Ph.D. degree in Electrical Engineering from Xi'an Jiaotong University, in 1998. She is currently a Professor with the School of Electrical Engineering, Xi'an Jiaotong University. Her research interests include power system reliability evaluation, integration of renewable energy, grid resilience, energy internet, and microgrids.

Goran Strbac (Member, IEEE) is a Professor of Energy Systems at Imperial College London, London, U.K. He led the development of novel advanced analysis approaches and methodologies that have been extensively used to inform industry, governments, and regulatory bodies about the role and value of emerging new technologies and systems in supporting cost effective evolution to smart low carbon future. He is currently the Director of the joint Imperial-Tsinghua Research Centre on Intelligent Power and Energy Systems, Leading Author in IPCC WG 3, Member of the European Technology and Innovation Platform for Smart Networks for the Energy Transition, and Member of the Joint EU Programme in Energy Systems Integration of the European Energy Research Alliance.

Authorized licensed use limited to: Imperial College London. Downloaded on February 19,2024 at 03:57:35 UTC from IEEE Xplore. Restrictions apply.
© 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.