# Resilience-Oriented Coordination of Networked Microgrids: a Shapley Q-Value Learning Approach

Dawei Qiu, *Member, IEEE,* Yi Wang, *Member, IEEE,* Jianhong Wang, *Student Member, IEEE,* Ning Zhang, *Senior Member, IEEE,* Goran Strbac, *Member, IEEE,* and Chongqing Kang, *Fellow, IEEE*

***Abstract*—High-impact and low-probability extreme events have occurred more frequently than before because of rapid climate change, which can seriously damage distribution systems. However, conventional distribution management can be dysfunctional after an event, destroying its centralized supervision towards resilience enhancement. In this context, networked microgrids (NMGs) with distributed energy resources provide a viable solution for the resilience enhancement of distribution systems. Existing literature tends to employ model-based optimization approaches for resilient operations of NMGs, which require complete system models and can be time-consuming. To address these challenges, this paper suggests a decentralized framework for resilience-oriented coordination of NMGs and proposes a novel multi-agent reinforcement learning (MARL) method to solve it. Specifically, the proposed MARL method develops an efficient credit assignment scheme for NMGs to learn their contributions to the distribution system resilience via the Shapley Q-value technique with more efficient resilience enhancement. Case studies based on two modified IEEE 15- and 69-bus distribution networks are conducted to validate the effectiveness of the proposed MARL method in enabling effective coordination among NMGs and providing a high resilience level.**

***Index Terms*—Multi-agent reinforcement learning, Networked microgrids, Resilience, Shapley Q-value.**

## NOMENCLATURE

### A. Abbreviation

| | |
|---|---|
| DER | Distributed energy resource |
| DG | Diesel generator |
| ES | Energy storage |
| HILP | High impact and low probability |
| MG | Microgrid |
| NMG | Networked microgrid |
| PV | Photovoltaic |
| RES | Renewable energy resource |
| SoC | State-of-the-charge |
| WT | Wind turbine |

### B. Indices and Sets

| | |
|---|---|
| $i \in \mathcal{I}$ | Index and set of MGs |
| $t \in \mathcal{T}$ | Index and set of time steps |
| $g \in \mathcal{DG}$ | Index and set of DGs |
| $g \in \mathcal{RES}$ | Index and set of RESs |
| $k \in \mathcal{ES}$ | Index and set of ESs |
| $b \in \mathcal{B}$ | Index and set of buses |
| $d \in \mathcal{D}$ | Index and set of loads |
| $(b,p) \in \mathcal{Y}$ | Index and set of lines |

### C. Parameters

| | |
|---|---|
| $\Delta t$ | Time resolution (1 hour) |
| $\lambda_{i,d}$ | Shedding cost of load $d$ in MG $i$ (£/kWh) |
| $\overline{P}^{ed}_{i,d,t}$ | Active power of baseline load $d$ in MG $i$ at time $t$ (kW) |
| $\overline{P}^{dg}_{i,g}$ | Maximum limit of active power of DG $g$ in MG $i$ (kW) |
| $\underline{P}^{dg}_{i,g}$ | Minimum limit of active power of DG $g$ in MG $i$ (kW) |
| $\overline{Q}^{dg}_{i,g}$ | Maximum limit of reactive power of DG $g$ in MG $i$ (kVAR) |
| $\underline{Q}^{dg}_{i,g}$ | Minimum limit of reactive power of DG $g$ in MG $i$ (kVAR) |
| $\delta^{dg}_{i,g}$ | Rated power factor of DG $g$ in MG $i$ |
| $\overline{P}^{res}_{i,g,t}$ | Power realization of RES $g$ in MG $i$ at time $t$ (kW) |
| $\overline{P}^{es}_{i,k}$ | Power capacity of ES $k$ in MG $i$ (kW) |
| $\overline{S}^{es}_{i,k}$ | Maximum limit of SoC of ES $k$ in MG $i$ (%) |
| $\underline{S}^{es}_{i,k}$ | Minimum limit of SoC of ES $k$ in MG $i$ (%) |
| $\overline{E}^{es}_{i,k}$ | Energy capacity of ES $k$ in MG $i$ (kWh) |
| $\eta^{esc}_{i,k}$ | Charging coefficient of ES $k$ in MG $i$ (%) |
| $\eta^{esd}_{i,k}$ | Discharging coefficient of ES $k$ in MG $i$ (%) |
| $\overline{V}$ | Maximum permissible voltage (p.u.) |
| $\underline{V}$ | Minimum permissible voltage (p.u.) |
| $r_{i,bp}$ | Resistance of line $(b,p)$ in MG $i$ (p.u.) |
| $x_{i,bp}$ | Reactance of line $(b,p)$ in MG $i$ (p.u.) |
| $\overline{S}_{i,bp}$ | Capacity limit of line $(b,p)$ in MG $i$ (kVA) |
| $\overline{P}_{ij}$ | Maximum limit of active power exchange between MG $i$ and MG $j$ (kW) |
| $\underline{P}_{ij}$ | Minimum limit of active power exchange between MG $i$ and MG $j$ (kW) |

This work was supported by the UK EPSRC project: 'Integrated Development of Low-Carbon Energy Systems (IDLES): A Whole-System Paradigm for Creating a National Strategy' (project code: EP/R045518/ 1), the Horizon Europe project 'Reliability, Resilience and Defense technology for the griD' (Grant agreement ID: 101075714), and the 'Imperial - Tsinghua Research Centre on Intelligent Power and Energy Systems'. Chongqing Kang and Ning Zhang are supported by Tsinghua University Initiative Scientific Research Program 20223080032. *(Corresponding author: Yi Wang.)*

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

Jianhong Wang is with the Department of Computer Science, University of Manchester, Manchester, M13 9PL, U.K. (e-mail: jianhong.wang@manchester.ac.uk).

Ning Zhang, and Chongqing Kang are with the State Key Laboratory of Power Systems, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: ningzhang@tsinghua.edu.cn, cqkang@tsinghua.edu.cn).

$\overline{Q}_{ij}$ Maximum limit of reactive power exchange between MG $i$ and MG $j$ (kVAR)
$\underline{Q}_{ij}$ Minimum limit of reactive power exchange between MG $i$ and MG $j$ (kVAR)
$\overline{S}_{ij}$ Maximum limit of apparent power exchange between MG $i$ and MG $j$ (kVA)
$\overline{F}_{i,bp}$ Maximum limit of virtual power flow through branch $(b,p)$ in MG $i$

*D. Variables*

$P^{ed}_{i,d,t}$ Active power of restored load $d$ in MG $i$ at time $t$ (kW)
$Q^{ed}_{i,d,t}$ Reactive power of restored load $d$ in MG $i$ at time $t$ (kVAR)
$P^{dg}_{i,g,t}$ Active power output of DG $g$ in MG $i$ at time $t$ (kW)
$Q^{dg}_{i,g,t}$ Reactive power output of DG $g$ in MG $i$ at time $t$ (kVAR)
$P^{res}_{i,g,t}$ Active power output of RES $g$ in MG $i$ at time $t$ (kW)
$P^{esc}_{i,k,t}$ Charging power of ES $k$ in MG $i$ at time $t$ (kW)
$P^{esd}_{i,k,t}$ Discharging power of ES $k$ in MG $i$ at time $t$ (kW)
$S^{es}_{i,k,t}$ SoC level of ES $k$ in MG $i$ at time $t$ (%)
$V_{i,b,t}$ Voltage of bus $b$ in MG $i$ at time $t$ (p.u.)
$P^{ex}_{ij,t}$ Active power exchange between MG $i$ and MG $j$ at time $t$ (kW)
$Q^{ex}_{ij,t}$ Reactive power exchange between MG $i$ and MG $j$ at time $t$ (kVAR)
$P_{i,bp,t}$ Active power flow of line $(b,p)$ in MG $i$ at time $t$ (kW)
$Q_{i,bp,t}$ Reactive power flow of line $(b,p)$ in MG $i$ at time $t$ (kVAR)
$F_{i,bp,t}$ Virtual power flow through branch $(b,p)$ in MG $i$ at time $t$
$y^{ln}_{i,bp,t}$ Binary indicating the status of line $(b,p)$ in MG $i$ at time $t$ (1 if closed, 0 otherwise)
$e^{bn}_{i,bp,t}$ Binary indicating the status of branch $(b,p)$ in MG $i$ at time $t$ (1 if closed, 0 otherwise)
$y^{ex}_{ij,t}$ Binary indicating the connection status between MG $i$ and MG $j$ at time $t$ (1 if connected, 0 otherwise)
$u^{es}_{i,k,t}$ Binary indicating the status of ES $k$ in MG $i$ at time $t$ (1 if charging, 0 otherwise)

## I. Introduction

### A. Background and Motivation

EXTREME weather events are characterized by high impact and low probability (HILP), which can affect the status of power system components and cause severe power outages [1]. Recently, decarbonization of the power industry has led to the increasing penetration of renewable energy resources (RESs), whose intermittent and fluctuating nature may further exacerbate the impact of extreme weather events. To deal with these HILP events, the concept of *resilience* has been introduced into power systems [2]. Given the serious disruptions, the main goal of a resilient power system is to maintain the continuous supply of essential loads, posing a system load restoration problem [3].

Microgrids (MGs), as localized small power systems with enhanced control capabilities, are regarded as an effective solution to integrate and coordinate different types of distributed energy resources (DERs) (e.g., diesel generators (DGs), wind turbines (WTs), photovoltaics (PVs), energy storages (ESs), etc.) for resilience enhancement [4]. In this context, it can be anticipated that MGs will become widespread in the coming decades and play a significant role in the development of future power systems globally, due to the crucial benefits (e.g., self-controlling, self-protecting, and self-healing) they offer for the enhancement of resilience in decentralized operation paradigms [4]. Going further, several MGs can even connect with each other as networked MGs (NMGs), where the energy sharing among them can effectively increase the survivability of essential loads during extreme events [5].

### B. Literature Review

Recently, much research has applied various control schemes to operate NMGs for the load restoration problem, which can be classified into the following three categories. First, *centralized control* that involves a central controller making energy schedules on behalf of all MGs. In [6], the NMGs are centrally optimized to provide reliable power supply to maximize load continuity. In [7], a group of mobile sources are centrally optimized to enhance the resiliency of NMGs. In [8], a linear integer program is introduced to solve the load restoration problem using NMGs, accounting for the factors of stability, frequency, voltage, and current. In [9], [10], a two-stage module enabling the evaluation of NMGs' restoration actions and their feasibility by unbalanced three-phase optimal power flow (3Ph-OPF) is proposed. In [11], a novel robust dual dynamic integer programming approach is firstly proposed to efficiently tackle the multi-stage resilient scheduling problems. The load restoration under tropical cyclones is maximized with guaranteed robustness to unanticipated outage uncertainties. However, the centralized control method may raise privacy concerns and can be prone to single-point failure, in particular when extreme weather events cause physical damage and the lose efficacy of information and communication technologies (ICTs) [12]. Second, *hierarchical control* that builds up a hierarchical structure between the central controller taking the role of enhancing the whole system resilience and the local MGs managing their individual energy schedules. In [13], a hierarchical outage management scheme is proposed to enhance the resilience of a distribution system comprised of multi-MGs against extreme events. In [14], a nested energy management system is proposed to hierarchically control the NMGs to enhance resilient performance. In [15], a two-stage heuristic is developed for the critical load restoration problem of NMGs through a chance-constrained stochastic program, accounting for the uncertainties of renewable and load. In [16], a heuristic rule-based algorithm is proposed to determine the optimal nodes to be restored within distribution systems. In this setting, privacy concerns can be addressed properly, the issue of single-point failure however still remains due

to the requirement for a central controller. Third, *distributed control* that removes the central controller and makes MGs' individual energy schedules. In [17], [18], a consensus-based algorithm is proposed to schedule various dispatchable DERs to maximize the supply adequacy of each MG. In [19], an alternating direction method of multipliers (ADMM) algorithm is proposed to make load restoration decisions individually for each MG, while ensuring the risk-limiting constraints of the overall system to be satisfied. In this setting, MGs rely on local information exchanges between neighbors with privacy perseverance. However, the iterative algorithm of distributed control may not guarantee an optimal or even a feasible solution [5].

Although the above model-based optimization approaches have been developed for a series of resilience enhancement problems under the NMG concept, the following challenges must be addressed when taking the real-world environment into account. First, exact operation models and accurate technical parameters of MGs are normally unavailable in practice due to aging and privacy concerns [5]. Second, because the distribution network environment is highly dynamic and stochastic, generalizing an adaptive and resilient control scheme that accounts for various system dynamics and uncertainties is difficult. Third, the resilience-oriented decisions of NMGs require a fast response time, while solving a complex optimization problem is normally time-consuming.

In this context, *reinforcement learning* (RL) [20], a data-driven and model-free approach, can solve dynamic decision-making problems by learning optimal control policies through repeated interactions with the environment without any *prior* knowledge. As an online learning approach, RL can effectively utilize the growing amount of data from the environment, capture various uncertainties, and adjust to different state conditions. Furthermore, the well-trained control policies can be directly deployed to the practical test process in milliseconds without solving an optimization problem. However, research work using RL to solve load restoration problems with NMGs has not yet been thoroughly investigated. In contrast, various RL methods have been successfully applied to the distribution system or a single MG environment, e.g., deep Q-network (DQN) [21]–[23] and soft actor-critic (SAC) [24] for reconfiguring a distribution network via smart switches after a major outage; deep deterministic policy gradient (DDPG) for optimal load restoration in a distribution network [25] and an islanded MG [26] via DER schedules.

However, the problem studied in this paper focuses on the coordination of NMGs for resilience enhancement. Using the above *single-agent reinforcement learning* (SARL) methods [21], [22], [24]–[26] to assist a cluster of NMGs in making decisions at the same time may cause scalability issues and further suffer from excessive computation time, because both state and action spaces grow proportionally with agent size, making it hard to train a large-scale neural network. To this end, it is reasonable to introduce the other category of RL, i.e., *multi-agent reinforcement learning* (MARL). Previous work [27] has successfully applied multi-agent DQN, multi-agent DPG, and multi-agent SAC to learn the optimal reconfiguration decisions for NMGs in distribution networks. However, it is assumed that each MG can acquire the local information of the active load in each bus and the private information of other MGs' actions, which may raise privacy concerns. To address this issue, a graph convolutional network is proposed in [28] to capture the graph-structured distribution system measurement patterns without complete information. Furthermore, the flexibility of DGs in providing resilience services is also investigated in [28] through multi-agent DQN. Finally, multi-agent SAC is proposed in [29] to learn the dispatched power from shunts for system resilience enhancement while only using the local information of bus voltages in each region.

Despite the aforementioned achievements in the applications of MARL methods to the distribution system resilience enhancement problems [27]–[29], a significant research gap has been identified, i.e., *a shared global reward function (e.g., the sum of weighted load restorations [27]) is designed for each individual agent in a coordinated manner that does not reflect each agent's contribution to the group, which may lead to unfair and consequentially inefficient learning performance.* Thus, this motivates us to investigate a credit assignment problem [30] that targets the distribution of global rewards to these agents according to their individual contributions. Inspired by the Shapley value [31], i.e., an efficient payoff distribution scheme, we propose a Shapley Q-value [32] that extends the Shapley value for the credit assignment of a MARL method. In this setting, the control policy of NMG operation can be effectively learned, while the contribution of each MG to the resilience of the distribution system can also be accurately quantified.

### C. Research Questions and Contributions

Although the aforementioned studies have shown that both model-based optimization and model-free MARL methods can be successfully applied to the NMGs' load restoration problem, the following critical research questions are still worth discussing:

1) Previous work [6]–[11], [13]–[19] has tried to employ model-based optimization approaches to operate NMGs for the load restoration problem. What if the exact operation models as well as the accurate technical and uncertain parameters of MGs cannot be obtained in practice? Are model-based optimization approaches fast enough to satisfy the response time requirement of a resilience-oriented problem?
2) Previous work [21], [22], [24]–[26] has tried to employ various SARL methods to operate smart switches and DER schedules to a resilient MG. What about the coordination effect of NMGs on providing load restorations? If the coordination of NMGs is introduced, what kinds of methods can be proposed to efficiently solve the load restoration problem with decentralized NMGs?
3) Previous work [27]–[29] has tried to employ various MARL methods to operate reconfiguration decisions, DG power schedules, and shunt power schedules for NMGs. How to identify individual contributions of each resource to the system-wise resilience performance?

To this end, this paper tries to answer the above discussed questions and makes the following contributions:

1) Develop a decentralized framework for the coordination effect of NMGs towards resilience enhancement. Formulate the coordinated operation problem of NMGs as a *Decentralized Partially Observable Markov Decision Process* (Dec-POMDP) [33], wherein each microgrid central controller (MGCC) is defined as an agent who can regulate its own DERs and power exchanges with other connected MGs in real-time; while the power flow model and load restoration process of the distribution network are together simulated as the environment. The objective of this Dec-POMDP is finding an optimal control policy for each agent to maximize the restored load condition of a distribution network.
2) Derive a novel MARL method named Shapley Q-value Deep Deterministic Policy Gradient (SQDDPG) to solve the formulated Dec-POMDP by i) introducing the concept of Shapley value [31] for accurate credit assignment in the load restoration process, resulting in a fair distribution of each agent's contribution to the system overall resilience; ii) learning a decentralized policy with a centralized critic by using an actor-critic architecture-based Deep Deterministic Policy Gradient (DDPG) [34] algorithm that can handle the continuous state and action spaces. In this setting, the system's operation models as well as the technical and uncertain parameters are unnecessary. Additionally, the well-trained control policy can be directly deployed to the resilience-oriented problem in milliseconds without solving any optimization problem.

In case studies, we validate the superior performance of the proposed SQDDPG method over the existing model-based and model-free methods in enhancing resilience. Specifically, the learned Shapley Q-value can efficiently reflect the contribution of each agent to the system overall resilience. Furthermore, a generalized control policy for the coordinated NMG operation problem is learned and can adapt to various system uncertainties, such as load profiles, RES generation, and contingencies. Finally, the scalability of the proposed SQDDPG method is evaluated on two modified IEEE 33-bus and 69-bus distribution networks. To the best of the authors' knowledge, this is the first time that Shapley Q-value has been applied to power systems to address the coordinated operation problem of NMGs towards a resilient distribution network.

### D. Paper Organization

The rest of this paper is organized as follows. Section II discusses the challenges and potential solutions of a resilient distribution system via NMGs. Section III presents the general formulations of NMG operations, while Section IV and Section V introduce the proposed Dec-POMDP formulation and SQDDPG method, respectively. In Section VI, case studies are carried out and analyzed on two modified IEEE 33-bus and 69-bus distribution networks. Section VII draws the conclusions and future work of this paper.

## II. Proposed Framework of Networked Microgrids Towards Resilience Enhancement

We focus on the coordination problem of an NMG cluster involving multi-interconnected MGs towards overall resilience enhancement, as illustrated in Fig. 1. In general, these MGs can be connected with each other and regulate their own energy resources for power sharing to facilitate load restoration after a major power outage in the distribution network.

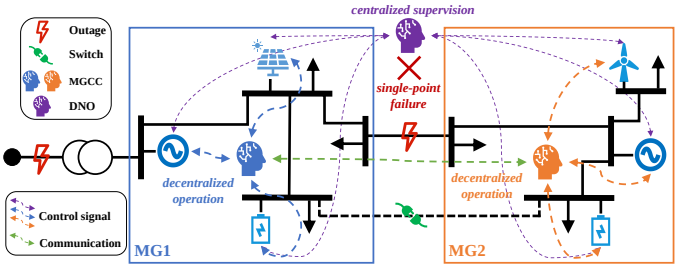



Fig. 1. Scheme of NMGs towards resilience enhancement. It also illustrates the transition from centralized supervision (purple) via DNO to decentralized operation (blue and orange) via NMGs due to the single-point failure and line outages.

### A. Challenges of Resilient Distribution System

The urgent need for resilience enhancement requires distribution systems to be managed in an efficient and secure manner. However, conventional distribution management featuring solely centralized supervision by the distribution network operator (DNO) may suffer from a reduction in operational reliability due to single-point failure, which can be caused by large-scale blackouts and damaged communication infrastructures [3].

As a result, a paradigm shift from a centralized to a decentralized manner is required, e.g., by dividing the distribution network into two regions, as illustrated in Fig. 1. As an emerging operation paradigm recently, MGs are able to provide a viable solution by constructing a hierarchical infrastructure to manage DERs in distribution networks [4]. In contrast to the centralized supervision, MGs can operate in a decentralized manner to manage regional operations in emergency conditions. However, the capacity and functionality of a single MG are still limited to achieve effective resilience enhancement of the distribution network, since the other MGs may not be capable of being restored by themselves [5].

### B. Resilience-Oriented Coordination of Networked MGs

NMGs can be implemented to enable additional flexibility for resilient operations by cooperatively sharing extra power resources [5]. Furthermore, each MG can have a heterogeneous topological and operational design that is more flexible to target local power supplies. More importantly, MGs operate in a decentralized manner that can prevent the system from a single-point failure. Finally, these MGs operate in parallel, which can expedite the restoration process.

The proposed framework of NMGs for resilience enhancement is illustrated in Fig. 1. After the power outages (e.g., main connection, distribution line) and also the single-point failure of centralized supervision, each MG is equipped with a MGCC that can regulate the power dispatches of controllable resources (e.g., PVs, WTs, DGs, and ESs) inside its own region, utilize tie-lines or smart switches, and then manage power exchanges with each other. In this context, these two MGs have their

own controllability and can operate within a decentralized framework without central commands from DNO, enabling a fast response time. As such, the overall resilience performance can be improved significantly.

## III. General Formulations of NMG Operation

In this section, we provide the general mathematical formulations of the considered DERs (including DGs, RESs, and ESs), the power exchanges, and the network operation constraints inside the NMGs as well as the objective function of the resilient distribution network. Finally, the practical challenges and potential solutions for solving a NMG coordination problem towards resilience enhancement are discussed.

### A. Operation Behaviors of DERs

*1) Diesel Generator:* The DG's operation model is simplistically constrained by its active and reactive power output limits

$$\underline{P}^{dg}_{i,g} \le P^{dg}_{i,g,t} \le \overline{P}^{dg}_{i,g}, \forall i \in \mathcal{I}, \forall g \in \mathcal{DG}, \tag{1}$$

$$\underline{Q}^{dg}_{i,g} \le Q^{dg}_{i,g,t} \le \overline{Q}^{dg}_{i,g}, \forall i \in \mathcal{I}, \forall g \in \mathcal{DG}, \tag{2}$$

$$|Q^{dg}_{i,g,t}| \le P^{dg}_{i,g,t} \tan(\cos^{-1} \delta^{dg}_{i,g}), \forall i \in \mathcal{I}, \forall g \in \mathcal{DG}, \tag{3}$$

where constraints (1)-(3) represent the active and reactive power limits of DG $g$ as well as its rated power factor $\delta^{dg}_{i,g}$ in MG $i$, respectively. It is noted that the operation constraints of ramping, start-up, and shunt-down are not considered in this model, which follow the same setting in [14], [35], [36], since this paper focuses on the resilience enhancement problem of distribution networks with small-size DGs.

*2) Renewable Generator:* The active power limit of RES $g$ (including WTs and PVs) in MG $i$ is represented by

$$0 \le P^{res}_{i,g,t} \le \overline{P}^{res}_{i,g,t}, \forall i \in \mathcal{I}, \forall g \in \mathcal{RES}, \tag{4}$$

where $\overline{P}^{res}_{i,g,t}$ represents the power realization of RES $g$ in MG $i$ at time step $t$, related to wind speed or solar irradiation. It is noted that the reason to consider the active power output of RESs as a dispatchable variable is to make renewable curtailment possible and allow the distribution network to have more flexibility, which follow the same setting in [13], [36].

*3) Energy Storage:* The operation model of ES $k$ in MG $i$ is formulated as

$$0 \le P^{esc}_{i,k,t} \le u^{es}_{i,k,t} \overline{P}^{es}_{i,k}, \forall i \in \mathcal{I}, \forall k \in \mathcal{ES}, \tag{5}$$

$$-(1 - u^{es}_{i,k,t}) \overline{P}^{es}_{i,k} \le P^{esd}_{i,k,t} \le 0, \forall i \in \mathcal{I}, \forall k \in \mathcal{ES}, \tag{6}$$

$$\underline{S}^{es}_{i,k} \le S^{es}_{i,k,t} \le \overline{S}^{es}_{i,k}, \forall i \in \mathcal{I}, \forall k \in \mathcal{ES}, \tag{7}$$

$$S^{es}_{i,k,t+1} = S^{es}_{i,k,t} + \frac{P^{esc}_{i,k,t} \Delta t \, \eta^{esc}_{i,k} + P^{esd}_{i,k,t} \Delta t / \eta^{esd}_{i,k}}{\overline{E}^{es}_{i,k}}, \quad \forall i \in \mathcal{I}, \forall k \in \mathcal{ES}, \tag{8}$$

where constraints (5) and (6) restrict charging and discharging power limits of ES $k$ in MG $i$ at time step $t$; the binary variable $u^{es}_{i,k,t} \in \{0, 1\}$ indicates its charging ($u^{es}_{i,k,t} = 1$) or discharging ($u^{es}_{i,k,t} = 0$) status. Constraint (7) limits the minimum and maximum battery state-of-the-charge (SoC) level of ES $k$, while its dynamic transition between two consecutive time steps is presented in (8), given the charging/discharging power $P^{es}_{i,k,t} = P^{esc}_{i,k,t} + P^{esd}_{i,k,t}$ as well as the charging/discharging coefficients $\eta^{esc}_{i,k}/\eta^{esd}_{i,k}$.

*4) Power Exchanges:* The active and reactive power exchanges between the NMGs are limited by

$$y^{ex}_{ij,t} \underline{P}_{ij} \le P^{ex}_{ij,t} \le y^{ex}_{ij,t} \overline{P}_{ij}, \forall i \in \mathcal{I}, \forall j \in \mathcal{I} \setminus \{i\}, \tag{9}$$

$$y^{ex}_{ij,t} \underline{Q}_{ij} \le Q^{ex}_{ij,t} \le y^{ex}_{ij,t} \overline{Q}_{ij}, \forall i \in \mathcal{I}, \forall j \in \mathcal{I} \setminus \{i\}, \tag{10}$$

$$(P^{ex}_{ij,t})^2 + (Q^{ex}_{ij,t})^2 \le y^{ex}_{ij,t} \overline{S}_{ij}, \forall i \in \mathcal{I}, \forall j \in \mathcal{I} \setminus \{i\}, \tag{11}$$

where constraints (9)-(11) correspond to the active, reactive, and apparent power exchange limits between MG $i$ and its interconnected MG $j$ at time step $t$, respectively. The binary $y^{ex}_{ij,t} \in \{0, 1\}$ indicating the status of connection line between MG $i$ and MG $j$ at time step $t$ (1 if closed, 0 if open or damaged) is determined by their reconfiguration switch.

### B. Network Constraints

For NMG operations, the AC power flow model with network constraints is employed, which can be realized as

$$\sum_{g \in B_{dg}} P^{dg}_{i,g,t} + \sum_{g \in B_{res}} P^{res}_{i,g,t} = \sum_{k \in B_{es}} P^{es}_{i,k,t} + \sum_{d \in B_{ed}} P^{ed}_{i,d,t} + \sum_{j \in B_{mg}} P^{ex}_{ij,t} - \sum_{(p,b) \in \mathcal{Y}} P_{i,pb,t} + \sum_{(b,p) \in \mathcal{Y}} P_{i,bp,t}, \forall i \in \mathcal{I}, \forall b \in \mathcal{B}, \tag{12}$$

$$\sum_{g \in B_{dg}} Q^{dg}_{i,g,t} = \sum_{d \in B_{ed}} Q^{ed}_{i,d,t} + \sum_{j \in B_{mg}} Q^{ex}_{ij,t} - \sum_{(p,b) \in \mathcal{Y}} Q_{i,pb,t} + \sum_{(b,p) \in \mathcal{Y}} Q_{i,bp,t}, \forall i \in \mathcal{I}, \forall b \in \mathcal{B}, \tag{13}$$

$$\underline{V}^2 \le V^2_{i,b,t} \le \overline{V}^2, \forall i \in \mathcal{I}, \forall b \in \mathcal{B}, \tag{14}$$

$$P^2_{i,bp,t} + Q^2_{i,bp,t} \le y^{ln}_{i,bp,t} \overline{S}_{i,bp}, \forall i \in \mathcal{I}, \forall (b,p) \in \mathcal{Y}, \tag{15}$$

$$V^2_{i,b,t} - V^2_{i,p,t} \le 2 (r_{i,bp} P_{i,bp,t} + x_{i,bp} Q_{i,bp,t}) + (1 - y^{ln}_{i,bp,t}) M, \forall i \in \mathcal{I}, \forall (b,p) \in \mathcal{Y}, \tag{16}$$

$$V^2_{i,b,t} - V^2_{i,p,t} \ge 2 (r_{i,bp} P_{i,bp,t} + x_{i,bp} Q_{i,bp,t}) + (y^{ln}_{i,bp,t} - 1) M, \forall i \in \mathcal{I}, \forall (b,p) \in \mathcal{Y}, \tag{17}$$

where the nodal active and reactive power balances at a certain bus $b$ in MG $i$ are presented by (12)-(13), respectively. $B_{ed}$, $B_{dg}$, $B_{res}$, $B_{es}$, and $B_{mg}$ correspond to the bus sets of loads, DGs, RESs, ESs, and connected MGs located at bus $b$, respectively. The nodal voltage and power flow limits are constrained in (14) and (15) respectively, while the linearized power flow constraints are expressed in (16)-(17) [37], which can be relaxed by the big-M method with a large positive value $M$. It is crucial to emphasize the importance of carefully selecting the big-M value in the model. If the big-M value is chosen too small, the resulting solution may violate the constraints of the model, leading to sub-optimal solutions. On the other hand, if the big-M value is chosen too large, it can result in a model with weak relaxations, making it challenging for the solver to converge efficiently. A study in [38] high-

lights these considerations. In this context, we have taken a thoughtful approach to select a reasonable value for the big-M parameter. Our chosen value ensures that the linearized power flow model yields optimal solutions while also facilitating efficient convergence of the solver. This careful selection of the big-M value enhances the accuracy and effectiveness of the power flow model in our study.

Furthermore, the MG network can be dynamically reconfigured through smart switch operations respecting the system radiality, subject to a set of virtual network constraints (18)-(21) according to the single-commodity flow model [37]:

$$\sum_{(b,p)\in\mathcal{Y}} e^{bn}_{i,bp,t} = |\mathcal{B}| - 1, \forall i \in \mathcal{I}, \tag{18}$$

$$\sum_{(p,b)\in\mathcal{Y}} F_{i,pb,t} - \sum_{(b,p)\in\mathcal{Y}} F_{i,bp,t} = 1, \forall i \in \mathcal{I}, \forall b \in \mathcal{B}/\{b^{sub}\}, \tag{19}$$

$$-e^{bn}_{i,bp,t}\,\overline{F}_{i,bp} \le F_{i,bp,t} \le e^{bn}_{i,bp,t}\,\overline{F}_{i,bp}, \forall i \in \mathcal{I}, \forall (b,p) \in \mathcal{Y}, \tag{20}$$

$$y^{ln}_{i,bp,t} \le e^{bn}_{i,bp,t}, \forall i \in \mathcal{I}, \forall (b,p) \in \mathcal{Y}, \tag{21}$$

where constraints (18)-(20) formulate a virtual network [37] with its structure being the same as the power network but without damages, ensuring that i) the virtual network has $|\mathcal{B}| - 1$ closed branches; ii) all virtual nodes are connected. Constraint (21) models the connections between the power network and the virtual network through their corresponding switches. $b^{sub}$ refers to the substation node inside the virtual network. Binary $e^{bn}_{i,bp,t} \in \{0, 1\}$ represents the status of branch $(b, p)$ in the virtual network (1 if closed, 0 if open), while $F_{i,bp,t}$ indicates the virtual power flow through branch $(b, p)$.

### C. Objective Function

The objective of this problem is to maximize the expectation of the weighted load restoration in the entire distribution network (comprised of multi-NMGs) over the daily horizon, which can be expressed as

$$\max \mathbb{E}_{\omega_t}\big[\sum\nolimits_{i\in\mathcal{I}}\sum\nolimits_{d\in\mathcal{D}}\sum\nolimits_{t\in\mathcal{T}} \lambda_{i,d} P^{ed}_{i,d,t}\big], \tag{22}$$

where $P^{ed}_{i,d,t}$ represents the amount of restored load $d$ in MG $i$ at time step $t$. Given the limited resource capacity and operational constraints of NMGs, restoring all electric loads in the distribution network is difficult; thus, load shedding cost $\lambda_{i,d}$ is introduced to prioritize the restoration of essential loads (e.g., hospitals and data centers) over non-essential loads (e.g., washing machines and office lighting). Finally, the load restoration process takes various uncertainties $\omega_t$ (e.g., load profiles, RES generation, and contingencies) into consideration when an outage occurs.

### D. Problem Challenges and Solutions

Solving the above optimization problem for NMGs towards load restoration is challenging. Firstly, the MGCC faces difficulties when the mathematical models and technical parameters of DERs and networks are unknown. In such cases, the optimization problem (1)-(22) cannot be formulated at all, leaving the MGCC without a clear solution strategy. Secondly, even the optimization problem (1)-(22) can be extracted, for example by the digital-twin technologies, it is subject to various uncertainties, making it difficult to obtain accurate probability distributions for these uncertainties. Thirdly, solving a time-coupled optimization problem with a large number of high-dimensional stochastic variables is computationally intensive and time-consuming. The load restoration process requires prompt decision-making and actions, but the lengthy computation times may impede the timely response needed for load restoration. Fourthly, developing a generalized control scheme that can adapt to any system state condition is difficult. Each new state condition requires resolving an independent optimization problem, which may not be feasible or efficient in practical scenarios with rapidly changing system conditions. Finally, achieving coordinated load restoration across all NMGs presents challenges in assessing the individual contribution of each MG to the overall resilience task. It is crucial to accurately reflect and evaluate the individual behaviors in relation to their impact on the overall load restoration objectives.

To address the challenges mentioned earlier, this paper proposes an alternative data-driven and model-free MARL method to solve the coordination problem of NMGs, which is followed by two practical implementations: 1) reformulating the NMG coordination problem into a Dec-POMDP, while NMGs can operate in a decentralized manner without any prior knowledge; and 2) deriving a novel MARL method based on Shapley value and DDPG algorithm that can learn a generalized control policy in real time with fair credit assignment in the load restoration process. Overall, the two implementation details are presented in the following Section IV and Section V, respectively.

## IV. Reformulation as a Dec-POMDP

Since NMGs are operating in a decentralized manner with a dynamic decision-making process and can only observe partial information of the distribution network (while the energy portfolios and dispatch behaviors of other NMGs are unknown), it is reasonable to formulate the examined NMG coordination problem as a *Decentralized Partially Observable Markov Decision Process* (Dec-POMDP) [33]. In general, Dec-POMDP is a 7-tuple $\langle \mathcal{I}, \mathcal{S}, \{\mathcal{O}_i\}, \{\mathcal{A}_i\}, \mathcal{R}, \mathcal{P}, \gamma \rangle$, including a set of agents $i \in \mathcal{I}$, a collection of global states $s \in \mathcal{S}$, a set of individual observations $\{o_i \in \mathcal{O}_i\}$, a set of individual actions $\{a_i \in \mathcal{A}_i\}$, and a global reward function $r \in \mathcal{R}$, as well as a state transition function $\mathcal{P}(s, a_{\mathcal{I}}, \omega)$, here $\omega$ represents the environment stochasticity as mentioned in objective function (22). The time interval of the load restoration process $\Delta t$ = 1 hour. Specifically, the components and dynamic process of Dec-POMDP are as follows.

### A. Agent and Environment Interactions

*1) Definitions of Agents and Environment:* The agents are defined as the MGCCs, who can regulate their individual DERs and power exchanges with the connected MGs (Section III-A). The environment is defined as the power flow model and load restoration process of the distribution network (Section III-B).

*2) Dynamic Decision-Making Process:* For each agent $i$ at time step $t$, an action $a_{i,t}$ is computed using the policy $\pi_i(a|o)$ conditioned on the current local observation $o_{i,t}$. Then, the environment transits to the next state given the transition function $\mathcal{P}$, while each agent $i$ is rewarded $r_t$ and receives a new local observation $o_{i,t+1}$. Following this process, each agent $i$ emits a trajectory of local observations, actions, and rewards: $\tau_i = o_{i,1}, a_{i,1}, r_1, o_{i,2}, ..., r_T$ over $\mathcal{O}_i \times \mathcal{A}_i \times \mathcal{R} \to \mathbb{R}$. The objective of this Dec-POMDP is to find an optimal policy $\pi_i(a|o)$ for each agent $i$ that can maximize the cumulative discounted global reward

$$G_t = \sum_{t=0}^{T} \gamma^t r_t, \tag{23}$$

where $\gamma \in [0, 1)$ and $T$ = 24 hours refer to the discount factor and restoration horizon, respectively.

### B. State and Local Observation

The environment state $s_t = \{o_{\mathcal{I},t}, x_t\}$ describes the configurations of all agents' local observations $o_{\mathcal{I},t}$ and NMGs' power exchanges $x_t$, which can be respectively defined as

$$o_{i,t} = [P_{i,b,t}, Q_{i,b,t}, y^{ex}_{ij,t}, S^{es}_{i,k,t}, S^{ln}_{i,t}] \in \mathcal{O}_i, \tag{24}$$

$$x_t = [P^{ex}_{ij,t}, Q^{ex}_{ij,t}] \in \mathcal{S}, \tag{25}$$

including 1) the active and reactive power injections $P_{i,b,t}, Q_{i,b,t}$ of each bus $b$ inside MG $i$; 2) the switch status $y^{ex}_{ij,t}$ of connection line between MG $i$ and its connected MG $j$; 3) the battery SoC $S^{es}_{i,k,t}$ of ES $k$ inside MG $i$; 4) the location of damaged lines $S^{ln}_{i,t}$ inside MG $i$; and 5) the active and reactive power exchanges $P^{ex}_{ij,t}, Q^{ex}_{ij,t}$ between all MGs $i \in \mathcal{I}$ and their corresponding connected MGs $j \in \mathcal{I} \backslash \{i\}$.

### C. Action

The action $a_{i,t}$ of each agent $i$ at time step $t$ is defined as

$$a_{i,t} = [a^{dg,p}_{i,g,t}, a^{dg,q}_{i,g,t}, a^{res}_{i,g,t}, a^{es}_{i,k,t}, a^{ex,p}_{ij,t}, a^{ex,q}_{ij,t}, a^{sw}_{ij,t}] \in \mathcal{A}_i, \tag{26}$$

including 1) the magnitude of active power dispatch $a^{dg,p}_{i,g,t} \in [0, 1]$ of DG $g$ inside MG $i$ as a percentage of its active power capacity $P^{dg}_{i,g,t} \in [\underline{P}^{dg}_{i,g}, \overline{P}^{dg}_{i,g}]$; 2) the magnitude of reactive power dispatch $a^{dg,q}_{i,g,t} \in [-1, 1]$ of DG $g$ inside MG $i$ as a percentage of its reactive power limit $Q^{dg}_{i,g,t} \in [\underline{Q}^{dg}_{i,g}, \overline{Q}^{dg}_{i,g}] \cap [-P^{dg}_{i,g,t} \tan(\cos^{-1} \delta^{dg}_{i,g}), P^{dg}_{i,g,t} \tan(\cos^{-1} \delta^{dg}_{i,g})]$; 3) the magnitude of active power generation $a^{res}_{i,g,t} \in [0, 1]$ of RES $g$ inside MG $i$ as a percentage of its power realization $P^{res}_{i,g,t} \in [0, \overline{P}^{res}_{i,g,t}]$; 4) the magnitude of charging (positive) and discharging (negative) power $a^{es}_{i,k,t} \in [-1, 1]$ as a percentage of its power capacity $P^{es}_{i,k,t} \in [-\underline{P}^{es}_{i,k}, \overline{P}^{es}_{i,k}]$; 5) the magnitude of active power exchange $a^{ex,p}_{ij,t} \in [-1, 1]$ between MG $i$ and MGs $j$ as a percentage of the active power limits $P^{ex}_{ij,t} \in [\underline{P}_{ij}, \overline{P}_{ij}]$; 6) the magnitude of reactive power exchange $a^{ex,q}_{ij,t} \in [-1, 1]$ between MG $i$ and MGs $j$ as a percentage of the reactive power limits $Q^{ex}_{ij,t} \in [\underline{Q}_{ij}, \overline{Q}_{ij}] \cap [-\sqrt{(\overline{S}^{ex}_{ij})^2 - (P^{ex}_{ij,t})^2}, \sqrt{(\overline{S}^{ex}_{ij})^2 - (P^{ex}_{ij,t})^2}]$; and 7) the switch action $a^{sw}_{ij,t} \in \{0, 1\}$ indicating the status of connection line between MG $i$ and its connected MG $j$, i.e., $a^{sw}_{ij,t} = 1$ for closed ($y^{ex}_{ij,t} = 1$) and $a^{sw}_{ij,t} = 0$ for open or damaged ($y^{ex}_{ij,t} = 0$).

### D. State Transition

The state transition process from time step $t$ to $t + 1$ is governed by $s_{t+1} = \mathcal{P}(s_t, a_{\mathcal{I},t}, \omega_t)$, which is influenced by a combination of the environment current state $s_t$, all agents' actions $a_{\mathcal{I},t}$, and environment stochasticity $\omega_t$.

At time step $t$, each MGCC agent $i$ executes its action $a_{i,t}$ and consequently can calculate the power dispatches of all its own DERs (i.e., $P^{dg}_{i,g,t}, Q^{dg}_{i,g,t}, P^{res}_{i,g,t}, P^{es}_{i,k,t}$), the switch operations $y^{ex}_{ij,t}$, and the power exchanges $P^{ex}_{ij,t}, Q^{ex}_{ij,t}$ with other connected MGs $j$ (Section IV-C). Afterwards, each MGCC agent $i$ runs its individual AC power flow and consequently can calculate the nodal power injections $P_{i,b,t}, Q_{i,b,t}$ of all buses inside its own region (Section III-B).

In related to the power exchanges between NMGs, the detailed calculation steps should be explained, which can be classified into three scenarios: 1) MG $i$ and MG $j$ have conflicting power exchanging intentions, i.e., $P^{ex}_{ij,t} \times P^{ex}_{ji,t} > 0$, therefore, there is no power exchanging activity between them; 2) MG $i$ and MG $j$ have complementary power exchanging intentions, i.e., $P^{ex}_{ij,t} \times P^{ex}_{ji,t} < 0$, then, the actual power exchanging amount between MG $i$ and MG $j$ is selected as the lower absolute value of their intended quantities ($\min\{|P^{ex}_{ij,t}|, |P^{ex}_{ji,t}|\}$), and the power exchanging direction depends on the sign of $P^{ex}_{ij,t}$, power flows from MG $i$ to MG $j$ if $P^{ex}_{ij,t} > 0$ and from MG $j$ to MG $i$ else if $P^{ex}_{ij,t} < 0$; and 3) one or both of MG $i$ and MG $j$ has or have no power exchanging intentions, i.e., $P^{ex}_{ij,t} \times P^{ex}_{ji,t} = 0$, then there is no power exchanging activity between them.

Since a storage unit cannot behave charging and discharging simultaneously at the same time, the state feature of ES battery SoC $S^{es}_{i,k,t}$ of MG $i$ is managed by the mutually exclusive quantities $P^{esc}_{i,k,t}, P^{esd}_{i,k,t}$ together with the minimum and maximum SoC levels $\underline{S}^{es}_{i,k}, \overline{S}^{es}_{i,k}$, energy and power capacities $\overline{E}^{es}_{i,k}, \overline{P}^{es}_{i,k}$, as well as the charging and discharging coefficients $\eta^{esc}_{i,k}, \eta^{esd}_{i,k}$, depicted as

$$P^{esc}_{i,k,t} = [\min(a^{es}_{i,k,t} \overline{P}^{es}_{i,k}, (\overline{S}^{es}_{i,k} - S^{es}_{i,k,t}) \overline{E}^{es}_{i,k} / (\Delta t\, \eta^{esc}_{i,k})]^+, \tag{27}$$

$$P^{esd}_{i,k,t} = [\max(a^{es}_{i,k,t} \overline{P}^{es}_{i,k}, (\underline{S}^{es}_{i,k} - S^{es}_{i,k,t})\, \overline{E}^{es}_{i,k}\, \Delta t\, \eta^{esd}_{i,k}]^-, \tag{28}$$

where $[\cdot]^{+/-} = \max / \min\{\cdot, 0\}$. Then, the state transition of $S^{es}_{i,k,t}$ from time step $t$ to $t+1$ can be written as equation (8).

On the other hand, the state feature of line outages $S^{ln}_{i,t}$ is not determined by the MGCC agents. Instead, it is influenced by the increasing frequency and severity of weather-related events. In this paper, we assume $S^{ln}_{i,t}$ corresponds to the exogenous state that is independent of agent actions and has intrinsic variability. RL can overcome this variability by adopting a data-driven fashion that directly learns its characteristics from the data set itself [20].

### E. Reward Function

The reward function design is responsible for motivating the agents' scheduling behaviors, and the reward signals could be

any unitless scalar values [20]. In section III-C, the objective of the problem is to maximize the weighted load restorations of the distribution network. However, directly using equation (22) as the reward function may raise serious convergence and optimality issues. This is because of the large fluctuations of weighted load restorations for different state conditions, possibly learning the unbalanced distributions of weight and bias values of the control policies [32]. To combat this, we scale the weighted load restorations and introduce an unitless resilience index [35] as the reward function

$$r_t = \frac{\sum_{i\in\mathcal{I}}\sum_{d\in\mathcal{D}}\lambda_{i,d}P^{ed}_{i,d,t}}{\sum_{i\in\mathcal{I}}\sum_{d\in\mathcal{D}}\lambda_{i,d}\overline{P}^{ed}_{i,d,t}} \in [0, 100\%], \tag{29}$$

where $\overline{P}^{ed}_{i,d,t}$ represents the baseline of load $d$ in MG $i$ at time step $t$. It is expected that the higher value of $r_t$ indicates more restoration of the weighted loads $P^{ed}_{i,d,t}$ and consequently better performance of the overall resilience enhancement. Furthermore, the scaled reward $r_t$ between 0 and 1 can improve the convergence rate and the final optimality of control policies.

## V. Proposed MARL Method

As explained in Section IV, the MGCC agents in the formulated Dec-POMDP aim to maximize the cumulative discounted global reward that is shared among all agents. However, the shared global reward function does not give individuals the accurate contribution, in other words, the incentives to adjust individual behaviors to reach optimality. As a result, the performance could be corrupted in the studied coordination problem of NMGs. This motivates us to apply the concept of credit assignment [30] that can efficiently distribute the global reward to agents according to their contributions. Specifically, we derive a novel MARL method called Shapley Q-value deep deterministic policy gradient (SQDDPG) to solve the above Dec-POMDP, with its overall architecture being depicted in Fig. 2. In SQDDPG, there are two practical implementation details that are insightful and crucial: 1) generalizing a Shapley Q-value [32] (i.e., an efficient payoff distribution scheme in RL) for credit assignment of the coordinated NMGs towards resilience enhancement; and 2) learning decentralized policies and centralized critics via the conventional deep deterministic policy gradient (DDPG) algorithm [34] to update the MARL policy with privacy perseverance.

### A. Shapley Q-value

*1) Shapley Value:* Shapley value [31] is one of the most popular methods to solve the payoff distribution problem for effective coordination. Given a set of NMGs $\mathcal{MG} = (\mathcal{I}, V)$, for any subset (coalition) of NMGs $\mathcal{G} \subseteq \mathcal{I}\backslash\{i\}$ let

$$\Phi_i(\mathcal{G}) = V(\mathcal{G}\cup\{i\}) - V(\mathcal{G}) \tag{30}$$

be a marginal contribution of agent $i$ with respect to coalition $\mathcal{G}$, where $V(\cdot)$ represents the value function to measure the payoffs earned by a NMG group. Then, the Shapley value of each agent $i$ can be written as

$$\mathrm{Sh}_i(\mathcal{MG}) = \sum_{\mathcal{G}\subseteq\mathcal{I}\backslash\{i\}} \frac{|\mathcal{G}|!(|\mathcal{I}| - |\mathcal{G}| - 1)!}{|\mathcal{I}|!}\cdot\Phi_i(\mathcal{G}). \tag{31}$$

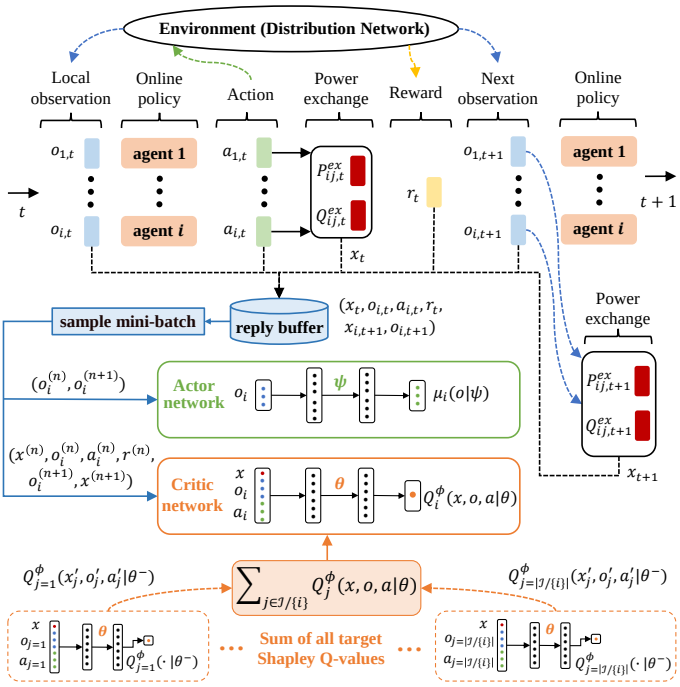



Fig. 2. The structure of the proposed SQDDPG method.

Literally, Shapley value takes the weighted average of marginal contributions of all possible coalitions where the weight follows a bell-shape distribution, so that it satisfies [32]: 1) *efficiency*: the group value equals to the sum of all local shapley values $\max V(\mathcal{I}) = \sum_{i\in\mathcal{I}}\max V(i)$; 2) *dummy*: if agent $i$ has no contribution, then $V(i) = 0$; and 3) *symmetry*: if agent $i$ and another agent $j$ make the same marginal contribution to the group, then $V(i) = V(j)$.

*2) Shapley Q-Learning:* In the era of RL concept, it is common to use Q-value (i.e., action-value) function $Q(s, a)$ to equivalently represent the value function $V(o)$ [20]. In this context, we can reformulate (31) as a Shapley Q-value

$$Q^{\phi}_i(s, a_i) = \sum_{\mathcal{G}\subseteq\mathcal{I}\backslash\{i\}} \frac{|\mathcal{G}|!(|\mathcal{I}| - |\mathcal{G}| - 1)!}{|\mathcal{I}|!}\cdot\Phi_i(s, a_i|\mathcal{G}), \tag{32}$$

where $\Phi_i(s, a_i|\mathcal{G})$ is defined as the marginal contribution of agent $i$ taking action $a_i$ in observing $s$, which is transferred from (30) using Q-value function $Q(s, a)$ for representing value function $V(o)$:

$$\Phi_i(s, a_i|\mathcal{G}) = \max_{a_{\mathcal{G}}} Q^{\pi^*_{\mathcal{G}}}(s, a_{\mathcal{G}\cup\{i\}}) - \max_{a_{\mathcal{G}}} Q^{\pi^*_{\mathcal{G}}}(s, a_{\mathcal{G}}), \tag{33}$$

where $Q^{\pi^*_{\mathcal{G}}}(s, a_{\mathcal{G}})$ represents the Q-value of coalition $\mathcal{G}$ taking actions $a_{\mathcal{G}}$ in observing $s$ via the optimal policy $\pi^*_{\mathcal{G}}$.

To effectively approximate the Shapley Q-value in (32), we take the principle of Bellman optimality equation [20] for the optimal Shapley Q-value such that

$$Q^{\pi^*}(s, a_{\mathcal{I}}) = \mathbb{E}_{s'}\left[r + \gamma\max_{a'_{\mathcal{I}}} Q^{\pi^*}(s', a'_{\mathcal{I}})\right]. \tag{34}$$

Then, by the property of *efficiency* defined in Section V-A1, we can get the optimal global Q-value by the sum of all

individual Shapley Q-values

$$\max_{a'_{\mathcal{I}}} Q^{\pi^*}(s', a'_{\mathcal{I}}) = \sum_{i\in\mathcal{I}} \max_{a'_{\mathcal{I}}} Q_i^{\phi^*}(s', a'_{\mathcal{I}}). \tag{35}$$

Furthermore, it is assumed that for each agent $i$ there exists bounded weights $w_i(s, a_i) \in \mathbb{R} > 0$ and biases $b_i(s) \in \mathbb{R} \geq 0$ that can project $Q^{\pi^*}(s, a_{\mathcal{I}})$ onto the space of $Q_i^{\phi^*}(s', a'_i)$ such that

$$Q_i^{\phi^*}(s, a_i) = w_i(s, a_i)\ Q^{\pi^*}(s, a_{\mathcal{I}}) - b_i(s). \tag{36}$$

Afterwards, by substituting (35) into (34) and further merging (36), we can obtain

$$Q_i^{\phi^*}(s, a_i) = w_i(s, a_i)\mathbb{E}_{s'}\big[r + \gamma \sum_{i\in\mathcal{I}} \max_{a'_i} Q_i^{\phi^*}(s', a'_i)\big] - b_i(s). \tag{37}$$

Finally, given the Bellman optimality equation and based on the conditions such that $\sum_{i\in\mathcal{I}} w_i(s, a_i)^{-1} b_i(s) = 0$ and $w_i(s, a_i) = 1/|\mathcal{I}|$ [39], we can update the Shapley Q-learning via its temporal-difference (TD) error

$$\Delta = r + \gamma \sum_{i\in\mathcal{I}} \max_{a'_i} Q_i^{\phi}(s', a'_i) - \sum_{i\in\mathcal{I}} Q_i^{\phi}(s, a_i) \tag{38}$$

until to find the optimal $Q_i^{\phi^*}(s, a_i)$ for each agent $i$.

It is worthwhile to note that the fundamental assumption of Shapley value is convexity, which is named as convexity in the context of cooperative game theory by Shapley [31]. In [32], it is proved that the global reward game is equivalent to a convex game (the game with convexity condition) with the grand coalition. For the studied problem in this paper, it can be categorized into the convex game, since NMGs are forming into a cooperative game with a common goal by equation (29) (that is defined as a global reward game). Also, Shapley value is a solution to convex game. That is, if a problem can be modelled as a global reward game, then Shapley value is a solution for the problem and therefore the value can be converged given sufficient number of samples.

### B. Shapley Q-value Deep Deterministic Policy Gradient

However, the value-based Shapley Q-learning method cannot deal with the continuous state and action spaces, referring to the setup of our studied problem in Section IV. To this end, we would like to propose a policy-based method by deploying the technique of Shapley Q-value to the conventional deep deterministic policy gradient (DDPG) algorithm [34] in a multi-agent domain, namely SQDDPG.

In detail, SQDDPG is constructed based on an actor-critic architecture that has two networks with different purposes. Regarding the actor network $\mu_i(o_i|\psi)$ parameterized by $\psi$, the input corresponds to the local observation $o_i$ with privacy perseverance rather than the environment state $s$, the output refers to the action value $a_i$ via the deterministic policy gradient theorem. Regarding the critic (i.e., Shapley Q-value) network $Q_i^{\phi}(x, o_i, a_i|\theta)$ parameterized by $\theta$, the input corresponds to the concatenation of NMG's power exchanges $x$, local observation $o_i$ and executed action $a_i$ of agent $i$, the output refers to a scalar estimate of the Shapley Q-value to perform the policy evaluation task. Based on the derivation of TD error expressed in (38), the critic network can be optimized via the TD learning

$$\mathcal{L}(\theta) = \mathbb{E}_{o'_i}\big[(r + \gamma \sum_{i\in\mathcal{I}} Q_i^{\phi}(x', o'_i, a'_i|\theta^-) - Q_i^{\phi}(x, o_i, a_i|\theta))^2\big], \tag{39}$$

where the parameters $\theta^-$ for the target critic network $Q_i^{\phi}(\cdot|\theta^-)$ are updated by copying the online parameters $\theta$ softly, to stabilize the learning performance. In target critic network $Q_i^{\phi}(\cdot|\theta^-)$, $a'_i$ represents the action generated from the target actor network $\mu_i(o'_i|\phi^-)$ according to the next local observation $o'_i$. Similarly, the target actor parameters $\phi^-$ are softly updated with its online actor parameters $\phi$. Here, the actor network is optimized via the deterministic policy gradient

$$\nabla_{\psi} J(\mu_i) = \nabla_{\psi}\mu_i(o_i|\psi)\nabla_{a_i} Q_i^{\phi}(x, o_i, a_i|\theta)|_{a_i=\mu_i(o_i|\phi)}. \tag{40}$$

### C. Implementation

SQDDPG is an off-policy MARL method that requires the past experiences to update the networks. Thus, a replay buffer $\mathcal{F}_i$ storing the past experiences acquired from the environment is required for each agent $i$. In detail, an experience is a transition tuple that contains $e_{i,t} = (x_t, o_{i,t}, a_{i,t}, r_t, x_{t+1}, o_{i,t+1})$. To update the actor and critic networks, a minibatch of $N$ mixed experiences are uniformly sampled from the replay buffer $\{e_i^{(n)}\}_{n=1}^{N} \sim \mathcal{F}_i$. Then, the mean-squared TD error of online critic network can be calculated as

$$\mathcal{L}(\theta) = \frac{1}{N}\sum_{n=1}^{N} \min\Big[\big(y_i^{(n)} - Q_i^{\phi}(x^{(n)}, o_i^{(n)}, a_i^{(n)}|\theta)\big)^2\Big], \tag{41}$$

where the target Shapley Q-value

$$y_i^{(n)} = r^{(n)} + \gamma \sum_{i\in\mathcal{I}} Q_i^{\phi}\Big(x^{(n+1)}, o_i^{(n+1)}, \mu_i(o_i^{(n+1)}|\psi^-)|\theta^-\Big), \tag{42}$$

The online actor network employing deterministic policy gradient theorem can be expressed as

$$\begin{aligned}\nabla_{\psi} J(\mu_i) = \frac{1}{N}\sum_{n=1}^{N}\Big[&\nabla_{\psi}\mu_i(o_i^{(n)}|\psi) \\ &\nabla_{a_i^{(n)}} Q_i^{\phi}(x^{(n)}, o_i^{(n)}, a_i^{(n)}|\theta)|_{a_i^{(n)}=\mu_i(o_i^{(n)}|\psi)}\Big].\end{aligned} \tag{43}$$

Afterward, the parameters of both online and target (actor and critic) networks are updated as

$$\theta \leftarrow \theta - \alpha^{\theta}\nabla_{\theta}\mathcal{L}(\theta) \quad \text{and} \quad \theta^- \leftarrow \tau\theta + (1-\tau)\theta^-, \tag{44}$$

$$\psi \leftarrow \psi + \alpha^{\psi}\nabla_{\psi} J(\mu_i) \quad \text{and} \quad \psi^- \leftarrow \tau\psi + (1-\tau)\psi^-. \tag{45}$$

where $\alpha^{\theta}, \alpha^{\psi}$ are the learning rates of the gradient descent algorithm for the online critic and actor networks, and $\tau$ is the soft update rate for the target critic and actor networks.

Additionally, to assist the MGCC agents in exploring the environment and acquiring more extensive experiences, a Gaussian noise $\mathcal{N}(0, \sigma_{i,t}^2)$ can be added to the online actor network (policy) $\mu_i(o_{i,t}|\psi_i)$, forming an exploration policy

$$\hat{\mu}(o_{i,t}) = \mu_i(o_{i,t}|\psi) + \mathcal{N}(0, \sigma_{i,t}^2). \tag{46}$$

Finally, the pseudo-code of SQDDPG is organized and presented in Algorithm 1:

**Algorithm 1** SQDDPG for $\mathcal{I}$ MGCC agents

Initialize actor $\mu_i(\cdot|\psi)$ with shared weights $\psi$ among all agents
Initialize critic $Q_i(\cdot|\theta)$ with shared weights $\theta$ among all agents
Set learning rates $\alpha^{\psi}, \alpha^{\theta}$ and an empty buffer $\mathcal{F}_i = \{\}$
**for** episode (i.e., day) $epi = 1$ to $E$ **do**
  Initialize the global state $s_0, x_0$ and local observation $o_{i,0}$
  Initialize a Gaussian exploration noise $\sigma_{i,t}$.
  **for** time step (i.e., 1 hour) $t = 1$ to $T$ **do**
    For each agent $i$, select action $a_{i,t}$ in observing $o_{i,t}$ via the exploration policy $\hat{\mu}(o_{i,t})$ in (46)
    Execute all agents' actions $a_{\mathcal{I},t}$ to the environment
    Calculate DERs' power dispatches $P^{dg}_{i,g,t}$, $Q^{dg}_{i,g,t}$, $P^{res}_{i,g,t}$, $P^{es}_{i,k,t}$ and NMGs' power exchanges $P^{ex}_{ij,t}, Q^{ex}_{ij,t}$
    Run power flow and get power injections $P_{i,b,t}, Q_{i,b,t}$
    Obtain reward $r_t$ and next observation $o_{i,t+1}$ and state $x_{t+1}$
    For each agent $i$, store one sample experience to buffer $e_{i,t} = (x_t, o_{i,t}, a_{i,t}, r_t, x_{t+1}, o_{i,t+1}) \rightarrow \mathcal{F}_i$
    **for** MGCC agent $i \in \mathcal{I}$ **do**
      Sample uniformly a minibatch of experiences $\{e^{(n)}_i = (x^{(n)}_i, o^{(n)}_i, a^{(n)}_i, r^{(n)}_i, x^{(n+1)}_i, o^{(n+1)}_i)\}^J_{j=1}$ from $\mathcal{F}_i$
      Updates network weights $\psi, \theta$ and $\psi^-, \theta^-$ in (44)-(45)
    **end for**
  **end for**
  Update state $s_t \leftarrow s_{t+1}$, $x_t \leftarrow x_{t+1}$, and local observation $o_{i,t} \leftarrow o_{i,t+1}$
**end for**

## VI. Case Studies

### A. Experimental Setup

*1) Networked Microgrids:* To assess the effectiveness of the proposed MARL method for the coordination of NMGs towards resilience enhancement, a modified IEEE 15-bus distribution network containing 3 NMGs, 2 tie-lines, and 4 smart switches is utilized for our experiments, as shown in Fig. 3. It is noted that the NMGs in this work are used to support load restorations for line outages. To achieve this target, MGs can use both generation resources (e.g., DGs, PVs, WTs, ESs) to directly support its own loads and network reconfigurations to allow power exchanges with their connected MGs. In this context, the modified IEEE 15-bus distribution network can be divided into three regions (MGs), of which each MG owns 1 DG, 1 PV or WT, and 1 ES. To capture the impact of extreme events, multiple line outages can happen in distribution network, where the potential outage locations are depicted in Fig. 3.

*2) Data Descriptions:* To capture uncertainties associated with demand and RES generation, a real-world open-source yearly dataset from [40] capturing various data characteristics is utilized, where the mean and samples of the daily system demand as well as PV and WT power generation over the year are illustrated in Fig. 4. Afterward, we split it into two pieces, with the first 11 months being the training data and the last month being the test data, for the purpose of MARL method evaluation.

*3) Benchmarks:* The proposed SQDDPG is compared with four benchmarks, including two MARL methods and two optimization methods: i) **IDDPG** [34]: each agent independently employs a conventional DDPG to provide resilience for its own region by constructing a local reward function $r_{i,t} = \sum_{d\in\mathcal{D}_i} \lambda_d P^{ed}_{d,t} / \sum_{d\in\mathcal{D}_i} \lambda_d \overline{P}^{ed}_{d,t}$, where $\mathcal{D}_i$ is the load set of MG $i$; ii) **MADDPG** [41]: each agent concatenates all

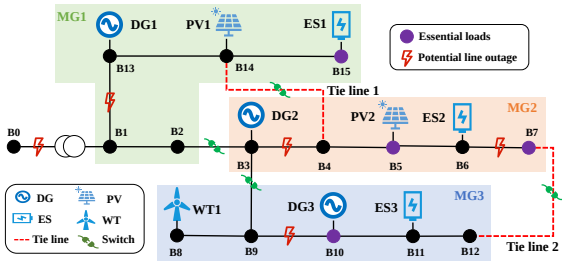



Fig. 3. The modified IEEE 15-bus distribution network with 3 NMGs.

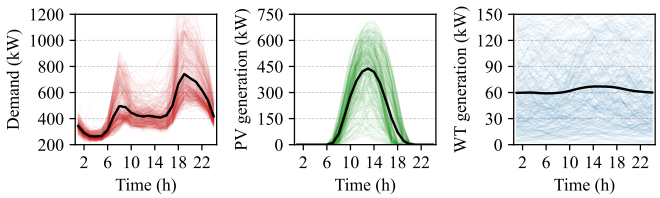



Fig. 4. Means (black lines) and samples (colored lines) of daily system demand as well as PV and WT power generation over one year.

others' local observations and actions into the critic network $Q_i(o_{\mathcal{I}}, a_{\mathcal{I}}|\theta)$, but neglects its marginal contribution to the group; iii) **Consensus** [19], [42], [43]: each agent employs an ADMM to solve a distributed coordination problem of NMGs, assuming perfect information of the system models and technical parameters, handling the system uncertainties via scenario generation and reduction techniques [35], where the binary variables (e.g., $y^{ln}_{bp,t}, y^{ex}_{ij,t}, e^{bn}_{bp,t}, u^{es}_{k,t}$) can be relaxed by their continuous versions, i.e., $y, e \in \{0,1\} \rightarrow 0 \leq y, e \leq 1$, suggested by [44]; and iv) **Centralization** [35]: DNO employs a central deterministic mixed-integer linear programming (MILP) for the daily optimization with the objective function (22) and constraints (1)-(21), assuming the perfect knowledge of the system models, technical parameters, and system uncertainties.

*4) Hyperparameters:* To build up the actor and critic networks, we use Multilayer Perceptrons with two hidden layers with 400 and 300 units, respectively. Two Adam optimizers with learning rates $\alpha^{\theta} = 10^{-3}$ and $\alpha^{\psi} = 10^{-4}$ are employed to update the critic and actor networks, respectively. We set the replay buffer size $|\mathcal{F}_i| = 10^5$ and minibatch size $N = 64$. The discount factor $\gamma = 0.99$ and soft update rate $\tau = 10^{-2}$ are implemented. For all MARL methods, we run 5,000 episodes with the same 10 random seeds. For each episode, a certain amount of lines with a relatively high probability can be potentially damaged, as shown in Fig. 3.

### B. Performance Evaluation

In this section, the training and test performance of the proposed SQDDPG and four benchmark methods are evaluated. Fig. 5(a) depicts the evolution of episodic rewards of three MARL methods over 5,000 training episodes, where the solid lines and the shaded areas indicate the moving average over 100 episodes and the oscillations of the original reward, respectively. Fig. 5(b) shows the evolution of episodic Shapley Q-values of 3 NMGs over 5,000 training episodes. We also collect the averaged resilience index (i.e., global reward) and

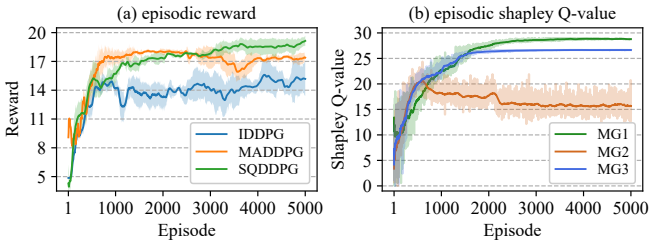



Fig. 5. (a) episodic reward for three MARL methods and (b) episodic Shapley Q-values for 3 NMGs in SQDDPG method over 5,000 episodes.

TABLE I
AVERAGED RESILIENCE INDEX AND COMPUTATION TIME OVER 31 TEST DAYS FOR DIFFERENT MARL AND OPTIMIZATION METHODS

| Method | Resilience index (%) | Computation (sec.) |
|---|---|---|
| IDDPG | 63.56 | 0.63 |
| MADDPG | 72.36 | 0.59 |
| SQDDPG | 76.42 | 0.56 |
| Consensus | 67.98 | 488.93 |
| Centralization | 79.54 | 9.56 |

computation time over the 31 test days for three MARL and two optimization methods in Table I.

The first observation we notice from Fig. 5(a) is that IDDPG (blue) has the most unstable and oscillatory training performance, thereby obtaining the lowest reward level. The independent learning method of IDDPG, which concentrates on local observation while disregarding others, is the major cause of this instability issue, making the environment non-stationary. As a result, by concatenating all agents' local observations and actions, MADDPG (orange) can effectively mitigate such a non-stationarity issue and thus display better stability performance. However, in the absence of the NMGs' power exchange information, agents in MADDPG become devoid of knowledge on the status of NMG coordination, resulting in sub-optimal performance. Additionally, using a common reward without credit assignment means each agent cannot acquire its accurate contribution to the group, which may result in ineffective learning of Q-value and consequently poor performance on policy evaluation tasks. In this case, the proposed SQDDPG (green) can address the above issues by 1) utilizing power exchange information x to learn NMG coordination dynamics; and 2) employing Shapley Q-value to effectively reflect each agent's contribution to the system's overall resilience, as depicted in Fig. 5(b), i.e., MG 1 (green) and MG 3 (blue) learn higher values of Shapley Q than MG 2 (orange), indicating their higher contributions to the group. Going further, it can be observed that MG 2 has a higher deviated Shapley Q-value than MG 1 and MG 3. This is because MG 1 and MG 3 have learned to supply power to MG 2 and make significant contributions to the group. The Shapley Q-values of MG 1 and MG 3 are very critical to the system's resilience performance, resulting in very stable learning curves under optimal solutions. On the other hand, the behavior of MG 2 is relatively less important; MG 2 learns a vibrative Shapley Q-value, but its effect on the system's resilience performance is relatively irrelevant.

Regarding the test performance in Table I, the proposed SQDDPG achieves a near-to-optimal performance (3.92% lower than Centralization), and outperforms IDDPG, MADDPG, and Consensus in terms of the averaged resilience over 31 test days by 20.23%, 5.61%, and 12.42%, respectively. On the other hand, all three MARL methods can be deployed in real-time at around 0.6 sec., while the optimization-based Consensus and Centralization methods require around 500 and 10 sec., respectively, to compute solutions. It is worth noting that real-time control is important to the resilient NMG coordination problem due to the demand for a fast response time.

### C. Analysis of Coordinated NMG Operation

After evaluating the MARL performance, this section validates the learned policy of SQDDPG for the coordination effect of 3 NMGs under a particular event scenario with 3 line outages (lines $0-1$, $6-7$, and $9-10$ in Fig. 3). Specifically, we analyze the 3 NMGs' corresponding behaviors of smart switches, power exchanges, and DER dispatches, which are illustrated in Fig. 6 and Fig. 7, respectively. Furthermore, the power supplies of 3 NMGs are shown in Fig. 8. Finally, the load restorations and the individual contributions of 3 NMGs to resilience enhancement are presented in Table II.

*1) Power Exchanges among NMGs and Switch Operations:* We first examine the power exchanges among the 3 NMGs. It can be observed from Fig. 6 that both MG 1 (green) and MG 3 (blue) are learned to supply power to MG 2. This is mainly driven by the following three reasons. First, MG 1 and MG 3 are characterized by their abundant resources (e.g., large DG capacity in Fig. 7(b)), resulting in the energy surplus supplied to MG 2 with the energy deficit of relatively high demand levels. Second, the smart switch operations prompt the transmission channel to enhance the capability of power exchanges among the 3 NMGs. Third, the particular (middle) location of MG 2 allows it to connect with both MG 1 and MG 3, leading to more options for power supply.

In particular, the smart switch operations of lines 2-3 and 4-14 are learned to close in the morning and evening, as depicted in Fig. 6(a)-(b). This allows MG 1 to be capable of supplying power to MG 2 (green) through these two lines. The reason why MG 1 does not supply power to MG 2 at midday is that MG 2 is characterized by abundant PV resources (Fig. 7(c)), which is enough to supply its own midday demand. Similarly, as shown in Fig. 6(c)-(d), MG 3 (blue) is learned to supply MG 2 via lines 3-9 and 7-12 with the close switch operations. However, MG 2 (orange) is learned to supply MG 3 through line 3-9 at midday because of its excessive PV resources. The interesting result in Fig. 6(d) is that MG 3 is also learned to supply power back to MG 2 at midday via line 7-12. This is because of the outage occurring at line 6-7. The essential load at bus 7 is isolated from MG 2, while the tie line is closed to allow the power supply from MG 3 to MG 2.

The physical behaviors of power exchanges and switch operations for 3 NMGs also reflect the elaborated theory of Shapley Q-value as discussed in Section VI-B. Specifically, MG 1 and MG 3 with abundant resources supplying power to MG 2 make significant contributions to the system resilience enhancement, therefore learning the higher levels of Shapley

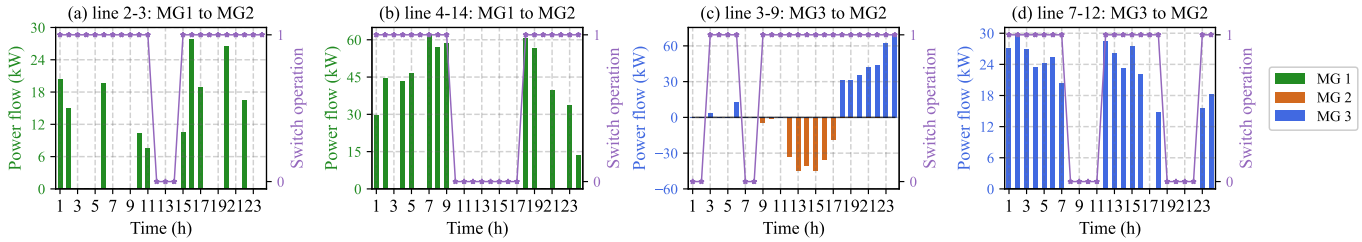



Fig. 6. Switch operations and power exchanges among 3 NMGs via 4 connected lines (a)-(d).

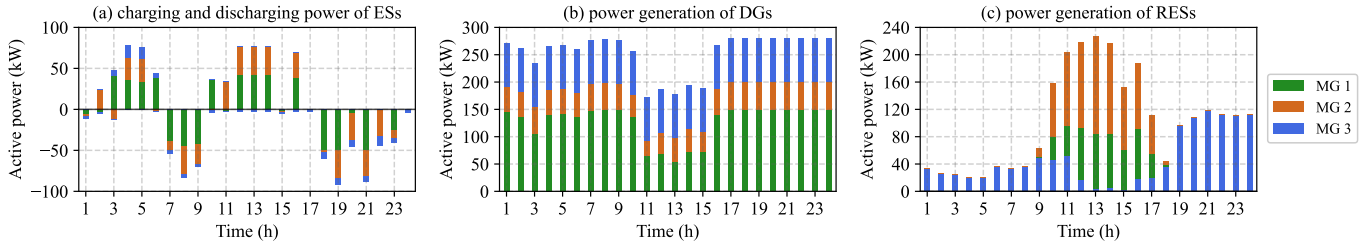



Fig. 7. Power dispatches of (a) ESs, (b) DGs, and (c) RESs of 3 NMGs.

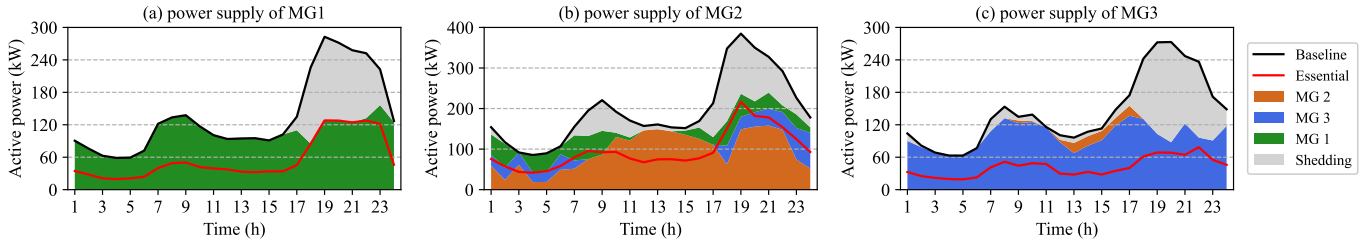



Fig. 8. Load profiles, load shedding, and power supplies of 3 NMGs (a)-(c).

TABLE II
SHEDDING QUANTITIES AND CONTRIBUTIONS OF 3 NMGS TO LOAD RESTORATION PROCESS IN THE MODIFIED IEEE 15-BUS POWER NETWORK

| Agent | MG 1 | MG 2 | MG 3 | Total |
|---|---|---|---|---|
| Essential load (kW) | 0 | 0 | 0 | 0 |
| Non-essential load (kW) | 786 | 1,111 | 988 | 2,885 |
| Contribution (%) | 41 | 23 | 36 | 100 |

Q-values in Fig. 5(b). On the other hand, MG 2 contributes less since it receives power from MG 1 and MG 3, therefore learning a lower Shapley Q-value in Fig. 5(b).

*2) Dispatches of DERs:* We next analyze the power dispatches of DERs inside 3 NMGs. It can be observed from Fig. 7(a) that the flexibility of ESs in both MG 1 and MG 2 is fully explored via the significant charging and discharging power magnitudes, while ES in MG 3 behaves inactively over the day. Specifically, the first charging behaviors of ESs in MG 1 and MG 2 occur at the beginning of the day (hours 1-6) for the purpose of discharging to supply the secondary demand peak in the early morning (hours 7-9). Meanwhile, the second charging and discharging cycle occurs at midday (hours 10-16) and at night (hours 17-24), respectively. This is because ESs in MG 1 and MG 2 are learned to shift PV generation from midday to supply the primary demand peak at night. The reason why ES in MG 3 behaves inactively is that MG 3's two demand peaks can be mostly met by its installed DG and WT, with no need for ES.

In terms of the dispatches of DGs and RESs in Fig. 7(b)-(c), their aggregated power generation exhibits a complementary effect, which can effectively supply the overall demand requirements of 3 NMGs.

*3) Power Supply and Load Restoration:* We third analyze the power supplies of 3 NMGs together with their load restorations and individual contributions. It can be found that MG 1 in Fig. 8(a) solely uses its own resources (green) to supply itself; while MG 2 in Fig. 8(b), apart from itself (orange), also receives a significant amount of power supply from both MG 1 (green) and MG 3 (blue) in the morning and at night; and lastly, MG 3 in Fig. 8(c) mainly relies on its own resources (blue) but also receives a certain level of power supply from MG 2 (orange) in the midday. Those power exchanges among the 3 NMGs are caused by i) the abundant resources in MG 1 and MG 3; ii) the high demand requirements in MG 2; iii) the severe power outages occurring in MG 2; and iv) the smart switch operations making the power transmission available. More importantly, all the essential loads (red lines) in the 3 NMGs are fully supplied, while the load shedding (gray area) belonging to non-essential loads happens during the two peak demand periods.

In summary, the proposed SQDDPG is capable of learning an effective coordination policy for 3 NMGs towards the load restoration problem of the modified IEEE 15-bus distribution network. As shown in Table II, the resilience enhancement for essential loads outperforms that for non-essential loads,

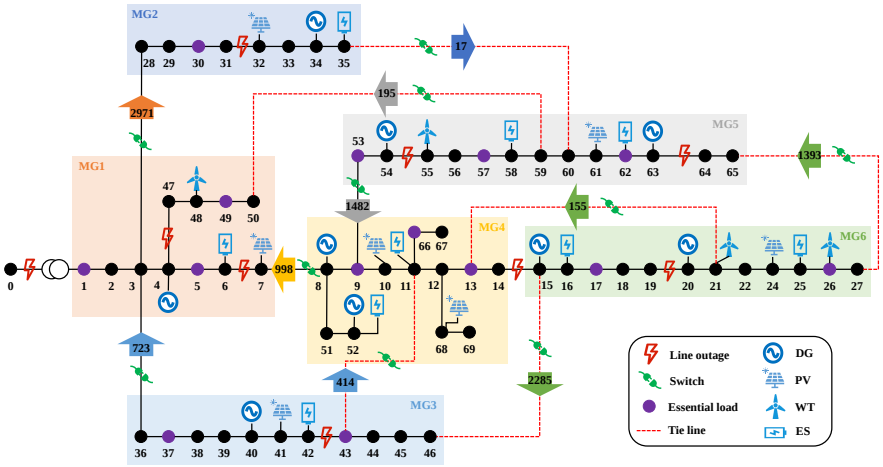



Fig. 9. Network topology and daily power exchanges of the modified IEEE 69-bus distribution network.

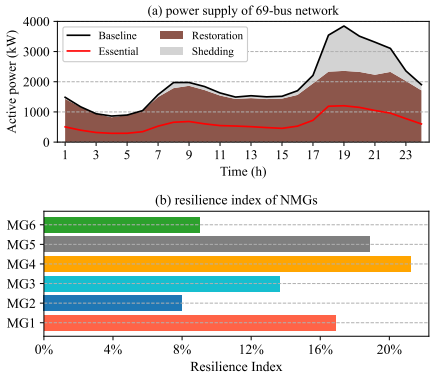



Fig. 10. (a) Aggregated load profile, shedding and restoration of the modified IEEE 69-bus distribution network; (b) Contributions of 6 NMGs to the resilient IEE 69-bus distribution network.

respectively, resulting in completely zero and 2,885 kWh of total load shedding quantity, yielding a 75% resilience index of the entire distribution network. Furthermore, MG 1 contributes the most (41%), which is nearly 1.78 and 1.14 times MG 2 and MG 3, respectively. Such outstanding performance further verifies the effectiveness of the proposed SQDDPG method in providing resilience enhancement for the distribution network.

### *D. Test Results in Modified IEEE 69-Bus Distribution Network*

This section serves as a further demonstration of the proposed SQDDPG method on scalability through a modified IEEE 69-bus distribution network that is operated via 6 NMGs, as depicted in Fig. 9. Inside each MG, multiple DGs, ESs, and RESs are appropriately deployed as emergency resources together with the smart switches for the overall load restoration. Multiple line outages can occur in an extreme event, including the disconnection from the main grid.

Similar to the 33-bus system, the performance of resilience enhancement for essential loads is much better than that for non-essential loads, i.e., no essential load shedding, as shown in Fig. 10(a). Specifically, the power exchanges among the 6 examined NMGs are presented in the colored arrows in Fig. 9. For instance, a line outage happens in line 31-32, causing the isolation of loads at buses 28-31, including an essential load at bus 30. To restore the load, MG 1 (orange) is sending a certain amount of power (2971 kW) to MG 2 (dark blue) through line 3-28. A similar effect can be found between MG 5 (gray) and MG 6 (green) through tie-line 27-65, MG 1 (orange) and MG 5 (gray) through tie-line 50-59, and MG 3 (light blue) and MG 6 (green) through tie-line 15-46.

Furthermore, the contribution of each MG can be found in Fig. 10(b), indicating the highest three levels of MGs 4-6 that together make 73% contributions to the 6-NMG group. This is because i) these 3 MGs are equipped with abundant generation and storage resources (2 DGs for all 3 MGs and 3 RESs for MG 6) to supply the system load restoration, e.g., MG 6 provides 3,833 kWh (sum of 3 green arrows) energy exports; ii) these 3 MGs are located in the center of the distribution network, allowing more lines (4 lines for both MG 4 and MG 5, and 3 lines for MG 6) for more effective power supplies with other NMGs. In summary, the proposed SQDDPG method learns an 88% resilience index of the entire distribution network, further validating its scalability in a large-scale system.

## VII. Conclusions and Future Work

This paper proposes a novel MARL method called SQDDPG to solve the resilience-oriented coordination problem of NMGs in a distribution network. The proposed SQDDPG method features a Shapley Q-value and DDPG algorithm that can accurately learn each agent's contribution to the system's resilience while also handling continuous state and action spaces. The NMG coordination problem is formulated as a Dec-POMDP, rendering it in a decentralized fashion and capturing various system dynamics and uncertainties, such as demand, renewable generation, and line outages. Experiment results based on two radial distribution networks (IEEE 33-bus and IEEE 69-bus) evaluate the superior performance of the proposed SQDDPG method in optimality, stability, and scalability compared to the state-of-the-art MARL and optimization methods. Finally, the coordination effects of NMGs on the resilience provision of two distribution networks are analyzed.

This paper only focuses on the implementation of two experiments: 3 NMGs in a modified IEEE 15-bus distribution network and 6 NMGs in a modified IEEE 69-bus distribution network. The scalability issue of using the SARL method (mentioned in Section I-B) can be improved by employing the proposed MARL-based SQDDPG method. In this setting, each MGCC agent only needs to manage its own region, which both state and action spaces reduce proportionally with agent size. However, the implementation of a large-scale NMG coordination problem is not considered, for example, hundreds of NMGs. Generally, there are two key challenges to scaling up the number of agents in the proposed SQDDPG method: 1) the dimensions of concatenating all agents' power exchanges, local observations, and actions will increase proportionally with the agent number, causing the curse of dimensionality

and making it impractical to train the neural networks; and 2) agent-agent interactions are critical in multi-agent systems while the number of interactions grows quadratically with the agent number, causing the non-stationary issue and difficulty in policy stabilization. Future work will try to address the above two challenges and develop a scalable MARL method for the large-scale NMG coordinated problem towards a resilient distribution network. Since the NMGs are characterized by similar behaviors, their scalability issue can be potentially addressed by the techniques of mean-filed approximation [45] and parameter-sharing [46].

Furthermore, distribution systems are inherently three-phase unbalanced because of the random demand changes in each phase. The existence of various uncertainties and contingencies under extreme weather events may even cause larger unbalances in distribution networks [9]. In this context, this paper employs a three-phase balanced distribution network as the RL environment, which may be impractical. Future work will focus on the resilience-oriented coordination problems of networked MGs in three-phase unbalanced distribution networks, while the unbalanced dynamic control of voltage and frequency can also be further investigated.

Finally, generalization and security are two key factors that limit the real-world applications of RL in power systems. This paper tries to demonstrate the generalization of the proposed SQDDPG method through two experiments on the modified IEEE 15-bus and 69-bus distribution networks. However, the model (SQDDPG) needs to be retrained when the later 69-bus distribution network is implemented. Additionally, the safety of power system operations must always be guaranteed. However, unsafe exploration actions can cause system infrastructure damage, resulting in more severe outages. Future work aims to address the generalization and security issues to develop a more generalized and safer RL algorithm that can quickly adapt to the real-world environment and also ensure secure operations, for example, by making use of the double critic networks [47], the integrated safety layer [48], [49], or the primal-dual gradient update [50].

## REFERENCES


[1] A. Poudyal, S. Poudel, and A. Dubey, "Risk-based active distribution system planning for resilience against extreme weather events," *IEEE Trans. Sustain. Energy*, Nov. 2022.

[2] IEEE PES Task Force, "Methods for analysis and quantification of power system resilience," *IEEE Trans. Power Syst.*, 2022.

[3] 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.

[4] G. Strbac, N. Hatziargyriou, J. P. Lopes, C. Moreira, A. Dimeas, and D. Papadaskalopoulos, "Microgrids: Enhancing the resilience of the european megagrid," *IEEE Power Energy Mag.*, vol. 13, no. 3, pp. 35–43, May-Jun. 2015.

[5] Z. Li, M. Shahidehpour, F. Aminifar, A. Alabdulwahab, and Y. Al-Turki, "Networked microgrids for enhancing the power system resilience," *Proc. IEEE*, vol. 105, no. 7, pp. 1289–1310, Jul. 2017.

[6] Z. Wang and J. Wang, "Self-healing resilient distribution systems based on sectionalization into microgrids," *IEEE Trans. Power Syst.*, vol. 30, no. 6, pp. 3139–3149, Nov. 2015.

[7] B. Papari, C. S. Edrington, M. Ghadamyari, M. Ansari, G. Ozkan, and B. Chowdhury, "Metrics analysis framework of control and management system for resilient connected community microgrids," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 704–714, Nov. 2021.

[8] Y. Xu, C.-C. Liu, K. P. Schneider, F. K. Tuffner, and D. T. Ton, "Microgrids for service restoration to critical load in a resilient distribution system," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 426–437, Jan. 2018.

[9] J. C. Bedoya, J. Xie, Y. Wang, X. Zhang, and C.-C. Liu, "Resiliency of distribution systems incorporating asynchronous information for system restoration," *IEEE Access*, vol. 7, pp. 101 471–101 482, Aug. 2019.

[10] 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.

[11] H. Qiu, W. Gu, W. Sheng, L. Wang, Q. Sun, and Z. Wu, "Resilience-oriented multistage scheduling for power grids considering nonanticipativity under tropical cyclones," *IEEE Trans. Power Syst.*, Aug. 2022.

[12] National Academies of Sciences, Engineering, and Medicine, *Enhancing the resilience of the nation's electricity system*. National Academies Press, 2017.

[13] H. Farzin, M. Fotuhi-Firuzabad, and M. Moeini-Aghtaie, "Enhancing power system resilience through hierarchical outage management in multi-microgrids," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2869–2879, Nov. 2016.

[14] A. Hussain, V.-H. Bui, and H.-M. Kim, "A resilient and privacy-preserving energy management strategy for networked microgrids," *IEEE Trans. Smart Grid*, vol. 9, no. 3, pp. 2127–2139, May. 2018.

[15] H. Gao, Y. Chen, Y. Xu, and C.-C. Liu, "Resilience-oriented critical load restoration using microgrids in distribution systems," *IEEE Trans. Smart Grid*, vol. 7, no. 6, pp. 2837–2848, Nov. 2016.

[16] A. Sharma, D. Srinivasan, and A. Trivedi, "A decentralized multiagent system approach for service restoration using dg islanding," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2784–2793, Nov. 2015.

[17] 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.

[18] A. Hussain, V.-H. Bui, and H.-M. Kim, "Resilience-oriented optimal operation of networked hybrid microgrids," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 204–215, Jan. 2019.

[19] 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.

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

[21] 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.

[22] Y. Huang, G. Li, C. Chen, Y. Bian, T. Qian, and Z. Bie, "Resilient distribution networks by microgrid formation using deep reinforcement learning," *IEEE Trans. Smart Grid*, Jun. 2022.

[23] J. Zhao, F. Li, H. Sun, Q. Zhang, and H. Shuai, "Self-attention generative adversarial network enhanced learning method for resilient defense of networked microgrids against sequential events," *IEEE Trans. Power Syst.*, 2022.

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

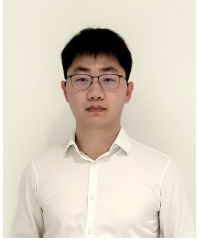
[29] M. Kamruzzaman, J. Duan, D. Shi, and M. Benidris, "A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources," *IEEE Trans. Power Syst.*, vol. 36, no. 6, pp. 5525–5536, Nov. 2021.

[30] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Auton. Agent Multi. Agent Syst.*, vol. 11, no. 3, 2005.

[31] L. S. Shapley, "A value for n-person games," *Contrib. Game Theory Manag.*, vol. 2, no. 28, pp. 307–317, 1953.

[32] J. Wang, Y. Zhang, T.-K. Kim, and Y. Gu, "Shapley q-value: A local reward approach to solve global reward games," in *Proc. Conf. AAAI Artif. Intell.*, vol. 34, no. 05, 2020, pp. 7285–7292.

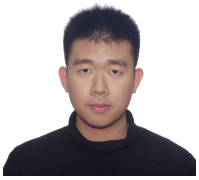
[33] F. A. Oliehoek and C. Amato, *A concise introduction to decentralized POMDPs*. Springer, 2016.
[34] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2016, pp. 1–14.
[35] 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.
[36] L. Bai, J. Wang, C. Wang, C. Chen, and F. Li, "Distribution locational marginal pricing (dlmp) for congestion management and voltage support," *IEEE Trans. Power Syst.*, vol. 33, no. 4, pp. 4061–4073, Jul. 2017.
[37] S. Lei, C. Chen, Y. Li, and Y. Hou, "Resilient disaster recovery logistics of distribution systems: Co-optimize service restoration with repair crew and mobile power source dispatch," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6187–6202, Nov. 2019.
[38] M. Cococcioni and L. Fiaschi, "The big-m method with the numerical infinite m," *Optim. Lett.*, vol. 15, no. 7, pp. 2455–2468, Sept. 2021.
[39] J. Wang, Y. Zhang, Y. Gu, and T.-K. Kim, "Shaq: Incorporating shapley value theory into multi-agent q-learning," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5941–5954, 2022.
[40] E. L. Ratnam, S. R. Weller, C. M. Kellett, and A. T. Murray, "Residential load and rooftop pv generation: an australian distribution network dataset," *Int. J. Sustain. Energy*, vol. 36, no. 8, pp. 787–806, 2017.
[41] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 30, pp. 6379–6390, 2017.
[42] J. Xie, C.-C. Liu, M. Sforna, and Y. Xu, "Consensus weighting of a multi-agent system for load shedding," *Int. J. Electr. Power Energy Syst.*, vol. 117, p. 105615, May 2020.
[43] K. P. Schneider, J. Glass, C. Klauber, B. Ollis, M. J. Reno, M. Burck, L. Muhidin, A. Dubey, W. Du, L. Vu *et al.*, "A framework for coordinated self-assembly of networked microgrids using consensus algorithms," *IEEE Access*, vol. 10, pp. 3864–3878, Jan. 2022.
[44] C. He, L. Wu, T. Liu, and M. Shahidehpour, "Robust co-optimization scheduling of electricity and natural gas systems via admm," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 658–670, Apr. 2016.
[45] D. Qiu, J. Wang, Z. Dong, Y. Wang, and G. Strbac, "Mean-field multi-agent reinforcement learning for peer-to-peer multi-energy trading," *IEEE Trans. Power Systems*, Oct. 2022.
[46] D. Qiu, Y. Ye, D. Papadaskalopoulos, and G. Strbac, "Scalable coordinated management of peer-to-peer energy trading: A multi-cluster deep reinforcement learning approach," *Appl. energy*, vol. 292, p. 116940, Jun. 2021.
[47] D. Qiu, Z. Dong, X. Zhang, Y. Wang, and G. Strbac, "Safe reinforcement learning for real-time automatic control in a smart energy-hub," *Appl. Energy*, vol. 309, p. 118403, Mar. 2022.
[48] Y. Wang, D. Qiu, M. Sun, G. Strbac, and Z. Gao, "Secure energy management of multi-energy microgrid: A physical-informed safe reinforcement learning approach," *Appl. Energy*, vol. 335, p. 120759, Apr. 2023.
[49] Q. Zhang, K. Dehghanpour, Z. Wang, and Q. Huang, "A learning-based power management method for networked microgrids under incomplete information," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1193–1204, Mar. 2020.
[50] Q. Zhang, K. Dehghanpour, Z. Wang, F. Qiu, and D. Zhao, "Multi-agent safe policy learning for power management of networked microgrids," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1048–1062, Mar. 2021.



**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.

**Yi Wang** (Member, IEEE) received 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.



**Jianhong Wang** (Student Member, IEEE) received B.Eng. degree from University of Liverpool, UK in 2016, M.Sc. degree from Imperial College London, UK in 2017, and M.Res. degree from University College London, UK in 2018. He has submitted thesis for the Ph.D. degree in Electrical Engineering Research at Imperial College London, UK.

He is now doing research associate at the Department of Computer Science, University of Manchester. His research interests lie in multi-agent reinforcement learning and its applications to the real-world problems.

**Ning Zhang** (Senior Member, IEEE) received the B.S. and Ph.D. degrees from Tsinghua University, Beijing, China, in 2007 and 2012, respectively.

He is currently an Associate Professor with Tsinghua University. His research interests include multiple energy systems integration, stochastic analysis, simulation of renewable energy, power system planning, and scheduling with renewable energy.

**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.

**Chongqing Kang** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Tsinghua University, Beijing, China, in 1997. He is currently a Professor with Tsinghua University.

His research interests include power system planning, power system operation, renewable energy, low carbon electricity technology, and load forecasting.