This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Two-Timescale Coordinated Voltage Regulation for High Renewable-Penetrated Active Distribution Networks Considering Hybrid Devices

Tingjun Zhang, Liang Yu, *Senior Member, IEEE*, Dong Yue, *Fellow, IEEE*, Chunxia Dou, *Senior Member, IEEE*, Xiangpeng Xie, *Member, IEEE*, and Gerhard P. Hancke, *Life Fellow, IEEE*

***Abstract*—The integration of large-scale distributed generators into active distribution networks (ADNs) will aggravate voltage fluctuations, which can affect the secure operation of power grids seriously. In this article, we investigate a cooperated voltage regulation problem of ADNs. Specifically, we first formulate a two-timescale voltage regulation problem considering the coordination of various hybrid devices while reducing the power loss of the whole ADNs. Given that the aforementioned problem is challenging to solve directly, we reformulate it as bilevel Markov games. Then, we propose a hierarchical multi-agent attention-based deep reinforcement learning algorithm to solve them. To be specific, the upper level Markov game is solved by a discrete multi-actor-attention-critic (MAAC) algorithm, and the lower level Markov game is solved by a continuous MAAC algorithm. In addition, the two-timescale coordination between upper level and lower level agents is implemented through the information exchange of rewards during the training process. Simulation results show that the proposed algorithm has good effectiveness, robustness, and scalability in voltage regulation.**

***Index Terms*—Active distribution networks (ADNs), bilevel Markov games, hierarchical multi-agent attention-based deep reinforcement learning (HMAADRL), hybrid devices, two-timescale voltage regulation.**

Manuscript received 7 April 2023; revised 29 June 2023; accepted 17 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62293500, Grant 62233010, Grant 62192751, and Grant 61972214, in part by the Basic Research Project of Leading Technology of Jiangsu Province under Grant BK20202011, in part by the Qinlan Project of Jiangsu Province (2022), in part by the Jiangsu Government Scholarship for Studying Abroad, and also in part by the 1311 Talent Project of Nanjing University of Posts and Telecommunications (NUPT). Paper no. TII-23-1227. *(Corresponding authors: Dong Yue; Liang Yu.)*

Tingjun Zhang is with the Institute of Advanced Technology for Carbon Neutrality, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: zhangtingjun_njupt@163.com).

Liang Yu is with the College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: liang.yu@njupt.edu.cn).

Dong Yue, Chunxia Dou, and Xiangpeng Xie are with the Institute of Advanced Technology for Carbon Neutrality, Nanjing University of Posts and Telecommunications, Nanjing 210023, China (e-mail: medongy@vip.163.com; cxdou@ysu.edu.cn; xiexp@njupt.edu.cn).

Gerhard P. Hancke is with the College of Automation and College of Artificial Intelligence, Nanjing University of Posts and Telecommunications, Nanjing 210023, China, and also with the School of Engineering, University of Pretoria, Pretoria 0002, South Africa (e-mail: g.hancke@ieee.org).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2023.3308348.

Digital Object Identifier 10.1109/TII.2023.3308348

I. Introduction

Vigorously developing renewable energy resources (RESs) [e.g., photovoltaics (PVs)] in active distribution networks (ADNs) is a crucial way to achieve carbon peaking and carbon neutrality goals [1]. However, the ADN operators will face several challenges due to the uncertainty and intermittency of RESs. For example, nodal voltages have a high risk of exceeding their upper voltage limits with the increase of PV plants in ADNs, which will endanger the safety of the whole power grid [2]. Therefore, it is imperative to study advanced voltage regulation approaches for modern ADNs with high-penetrated PVs.

A. Literature Review

In existing studies, several model-based approaches have been adopted for coordinated voltage regulation of ADNs. For example, Li et al. [3] designed a distributed approach for voltage control combining model predictive control and droop control method. The designed algorithm improved the voltage regulation performance through rolling optimization. Huang et al. [4], developed a different distributed approach for voltage control using consistent alternating direction multiplier algorithm to realize distributed reactive power control. To deal with uncertainties in ADNs, Xu et al. [5] proposed a multitimescale stochastic voltage control method using stochastic programming (SP). Different from [5], several voltage regulation algorithms were proposed based on robust optimization [6], [7], [8]. In addition, Jin et al. [9] proposed a multi-objective optimization problem for voltage regulation of ADNs with the consideration of global optimization, user preferences, and local control. Jha et al. [10] proposed a bilevel volt/Var optimization algorithm,

1551-3203 © 2023 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

where the upper level problem was formulated as a mixed integer linear programming, and the lower level problem was modeled as a nonlinear programming. Chowdhury and Kamalasadan [11], proposed a new second-order cone programming (SOCP) method for voltage regulation in ADNs. In addition, Zafar et al. [12] designed a multitimescale voltage control optimization algorithm to improve the safety of ADNs. Different from [11], the voltage regulation problem was modeled as a mixed-integer second-order cone program (MISOCP). Zheng et al. [13], proposed a dual-timescale cooperative voltage control problem, and the problem was solved using the column-and-constraint generation algorithm. Although the above model-based voltage regulation approaches achieved promising performance, they still have limitations. First, they need to know the exact model information and the prior knowledge of uncertain parameters, which may be challenging to obtain [14]. Second, some of the conventional model-based approaches (e.g., SP-based voltage regulation approach) have a heavy computational burden and their corresponding computation time may be unacceptable in practice [15].

To this end, several voltage regulation approaches based on deep reinforcement learning (DRL)/multi-agent DRL (MADRL) have been developed, which have been applied in numerous areas, e.g., smart grid [16], [17], smart buildings [18], [19], electric vehicle charging [20], [21], [22], and manufacturing systems [23]. For example, Wang et al. [24] proposed a deep deterministic policy gradient-based voltage control method by coordinating active and reactive power of electric vehicles. Wang et al. [25] proposed a DRL-based voltage control method by scheduling energy storage systems. Yang et al. [26] designed a multitimescale voltage control scheme for ADNs by combining the data-driven approach with model-based approach. However, the approach neglects the coordination between upper level and lower level devices. To overcome the above drawback, Sun and Qiu [27] proposed a two-stage DRL-based voltage regulation approach to mitigate the voltage violation. In the first stage, the optimization problem was formulated as a MISOCP. In the second stage, the multi-agent deep deterministic policy gradient (MADDPG) algorithm was used to solve the fast-timescale voltage control problem. Liu and Wu [28], proposed a bilevel DRL-based algorithm for voltage control. A multidiscrete soft actor–critic (SAC) algorithm was used to control slow-timescale discrete devices, and the SAC algorithm was adopted to learn a reliable voltage control policy in fast timescale. However, the proposed DRL-based voltage regulation algorithm adopted centralized control in both upper and lower layers. Different from [28], Cao et al. [29] proposed a different multitimescale voltage control method. The proposed method used the centralized SAC method to train upper level agents and used the multi-agent soft actor critic (MASAC) algorithm to train lower level agents for decentralized voltage control. Although the above multitimescale DRL-based voltage regulation methods have made several advances, they all adopted the single-agent centralized control method for discrete devices in slow timescale. When the number of discrete devices increases, the size of their discrete action space will increase exponentially, which will affect the efficiency of policy learning. Moreover, existing multitimescale DRL-based approaches neglect to coordinate more hybrid devices, such as battery energy storage systems (BESSs) and flexible loads (FLs), which will limit the voltage regulation potential of the system.

B. Motivation and Contribution

There are several challenges to achieve the aim of voltage regulation considering hybrid devices. First, it is difficult to obtain the accurate model of ADNs. Second, there are several uncertain parameters. Third, hybrid devices have different regulating timescales. To overcome these challenges, we investigate a two-timescale coordinated voltage regulation problem (i.e., regulate all bus voltages in a safe range and minimize the total power loss of whole ADNs) considering various controllable hybrid devices, such as on-load tap changers (OLTCs), capacitor banks (CBs), PV inverters, static Var compensators (SVCs), BESSs, and FLs. Moreover, we propose a decentralized voltage regulation algorithm in both fast timescale and slow timescale based on hierarchical multiagent attention-based deep reinforcement learning (HMAADRL).

The major contributions of this article are summarized as follows.

1) By taking discrete, continuous, multitimescale hybrid devices into consideration, we formulate a voltage optimization problem of ADNs. Due to the difficulty of solving such a complex decision-making problem directly, the optimization problem is further reformulated as bilevel Markov games.
2) A novel HMAADRL-based voltage regulation algorithm is proposed to solve the above bilevel Markov games. To be specific, a discrete multi-actor-attention-critic (DMAAC) algorithm is designed to control slow-timescale discrete devices. A continuous MAAC (CMAAC) algorithm is adopted to control fast-timescale continuous devices. The collaboration of fast-timescale devices and slow-timescale devices is implemented through the information exchange of reward during the training process.
3) Compared with model-based approaches, the proposed HMAADRL-based voltage regulation algorithm can achieve the approximate power loss without knowing precise model information and any prior knowledge of uncertain parameters. Moreover, compared with the algorithm in [29], the proposed algorithm achieves the lower power loss while ensuring the voltage safety of all buses.

The rest of this article is organized as follows. The voltage regulation problem of the ADN is first formulated in Section II. Moreover, the optimization problem is further formulated as bilevel Markov games. In Section III, we propose a HMAADRL-based algorithm to solve Markov games. In addition, in Section IV numerical results are analyzed and compared. Finally, Section V concludes this article.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

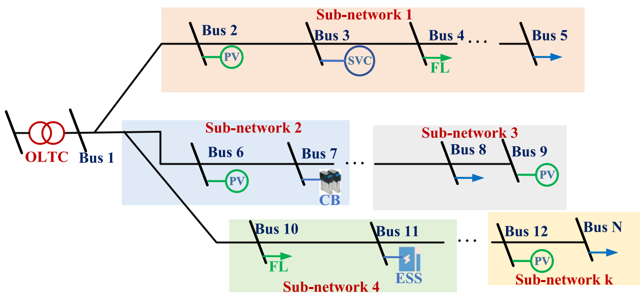


Fig. 1. Typical topology of the ADN.

II. Problem Formulation

In this part, the two-timescale voltage regulation problem is first formulated considering multiple hybrid devices. Then, the problem is reformulated as bilevel Markov games.

A. Voltage Regulation Problem Formulation

We study a typical radial ADN with N buses, as shown in Fig. 1. Hybrid devices, such as OLTCs, CBs, SVCs, PV inverters, BESSs, and FLs are connected to different buses. In addition, the ADN is divided into k subnetworks for the ease of operations. Detailed partitioning rules can be found in [30].

This article focuses on finding optimal cooperative voltage control strategies for hybrid devices without knowing the exact model of ADNs. Specifically, each day is separated into T time steps, and each time step consists of $\mathit{\Gamma}$ time intervals. In the slow timescale $t \in T$, OLTCs and CBs are scheduled cooperatively to minimize the voltage deviations. The switching number of these discrete devices is also optimized. In the fast timescale $\tau \in \mathit{\Gamma}$, smart PV inverters, SVCs, BESSs, and FLs are regulated for fast voltage fluctuations. In addition, the long-term power losses of whole ADNs are also minimized by coordinating two-timescale devices. Formally, we formulate an optimal voltage regulation problem as follows:

$$(\mathbf{P1})\ \min C_1 + \delta_1 C_2 + \delta_2 C_3$$

$$C_1 = \sum_{n=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{\mathit{\Gamma}} |\Delta V_{n,t,\tau}|$$

$$C_2 = \sum_{n=1}^{N}\sum_{m=1}^{N}\sum_{t=1}^{T}\sum_{\tau=1}^{\mathit{\Gamma}} P^L_{n,m,t,\tau}$$

$$C_3 = \sum_{n=1}^{N}\sum_{t=1}^{T} Z_{n,t} \tag{1}$$

s.t.

$$P^{\text{PV}}_{n,t,\tau} + P^{\text{BESS}}_{n,t,\tau} - (P^{\text{Load}}_{n,t,\tau} + \Delta P^{\text{FL}}_{n,t,\tau}) = V_{n,t,\tau} \sum_{m=1}^{N} V_{m,t,\tau}(G_{n,m}\cos\vartheta_{n,m,t,\tau} + B_{n,m}\sin\vartheta_{n,m,t,\tau}) \tag{2}$$

$$Q^{\text{PV}}_{n,t,\tau} + Q^{\text{SVC}}_{n,t,\tau} + Q^{\text{CB}}_{n,t} - Q^{\text{Load}}_{n,t,\tau} = V_{n,t,\tau} \sum_{m=1}^{N} V_{m,t,\tau}(G_{n,m}\sin\vartheta_{n,m,t,\tau} - B_{n,m}\cos\vartheta_{n,m,t,\tau}) \tag{3}$$

$$V_{\min} \le V_{n,t,\tau} \le V_{\max} \tag{4}$$

$$\Delta V_{n,t,\tau} = \begin{cases} V_{n,t,\tau} - V_{\max}, & \text{if } V_{n,t,\tau} > V_{\max} \\ V_{n,t,\tau} - V_{\min}, & \text{if } V_{n,t,\tau} < V_{\min} \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$P^L_{n,m,t,\tau} = G_{n,m}(V^2_{n,t,\tau} + V^2_{m,t,\tau} - 2V_{n,t,\tau}V_{m,t,\tau} \cos\vartheta_{n,m,t,\tau}) \quad \forall n,m \in \mathcal{N}_{\text{bus}} \tag{6}$$

$$\psi_{n,t} \in \{0,1,2,3,4,5\} \tag{7}$$

$$\chi_{n,t} \in \{-5,-4,-3,-2,-1,0,1,2,3,4,5\} \tag{8}$$

$$Q^{\text{CB}}_{n,t} = Q^{\text{CB}}_n \psi_{n,t} \tag{9}$$

$$V_{1,t+1,\tau} = V_{0,t,\tau} + \chi_{n,t} V_{\text{tap}} \tag{10}$$

$$Z_{n,t} = \begin{cases} |\chi_{n,t} - \chi_{n,t-1}|, & \text{if } n \in \mathcal{N}_{\text{OLTC}} \\ |\psi_{n,t} - \psi_{n,t-1}|, & \text{if } n \in \mathcal{N}_{\text{CB}} \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

$$(P^{\text{PV}}_{n,t,\tau})^2 + (Q^{\text{PV}}_{n,t,\tau})^2 \le (S^{\text{PV}}_n)^2 \tag{12}$$

$$-P^D_{n,\max} \le P^{\text{BESS}}_{n,t,\tau} \le P^C_{n,\max} \tag{13}$$

$$B_{n,t,\tau+1} = \begin{cases} B_{n,t,\tau} + \eta_C P^{\text{BESS}}_{n,t,\tau}, & P^{\text{BESS}}_{n,t,\tau} \ge 0 \\ B_{n,t,\tau} + \frac{P^{\text{BESS}}_{n,t,\tau}}{\eta_D}, & P^{\text{BESS}}_{n,t,\tau} \le 0 \end{cases} \tag{14}$$

$$B_{n,\min} \le B_{n,t,\tau} \le B_{n,\max} \tag{15}$$

$$P^{\text{FL}}_{n,t,\tau,\min} \le P^{\text{FL}}_{n,t,\tau} + \Delta P^{\text{FL}}_{n,t,\tau} \le P^{\text{FL}}_{n,t,\tau,\max} \tag{16}$$

$$Q^{\text{SVC}}_{n,\min} \le Q^{\text{SVC}}_{n,t,\tau} \le Q^{\text{SVC}}_{n,\max} \tag{17}$$

where (1) represents the objective function, which is the weighted sum of voltage deviations, power loss, and adjustments of discrete devices; $\Delta V_{n,t,\tau}$ indicates the voltage deviations of bus n that exceeds the safe range in time τ during time t; $P^L_{n,m,t,\tau}$ denotes the power loss through line (n,m), where $\mathcal{N}_{\text{bus}}$ represents bus indexes of the ADN; $Z_{n,t}$ represents the adjustments of OLTC and CBs, where $\mathcal{N}_{\text{OLTC}}$ and $\mathcal{N}_{\text{CB}}$ denote the set of bus indexes related to OLTCs and CBs, respectively; δ_1 and δ_2 denote the weighted coefficients used to balance voltage deviations, power losses, and adjustments of discrete devices; (2) and (3) represent the power flow equation constraints; $P^{\text{PV}}_{n,t,\tau}$, $P^{\text{BESS}}_{n,t,\tau}$, and $P^{\text{Load}}_{n,t,\tau}$ are the active power of PVs, BESSs, and loads linked to a bus n; $\Delta P^{\text{FL}}_{n,t,\tau}$ is the adjustment amount of FLs; $Q^{\text{PV}}_{n,t,\tau}$, $Q^{\text{SVC}}_{n,t,\tau}$, and $Q^{\text{CB}}_{n,t}$ represent the reactive power injection of PVs, SVCs, and CBs connected to n, respectively; $Q^{\text{Load}}_{n,t,\tau}$ is the reactive power demand of load connected to bus n; $G_{n,m}$ and $B_{n,m}$ are the real and imaginary part of admittance element between buses n and m, while $\vartheta_{n,m,t,\tau}$ indicates the voltage phase difference between buses n and m; (4) and (5) are the voltage constraints; (6) calculates the power loss [27]. Equations (7) and (8) denote the set of discrete action for OLTCs and CBs; (9) calculates the reactive power injection of CBs. Q^{CB}_n is the

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

capacity of each group of CB n; (10) computes the voltage, which is dependent on the position of OLTCs; V_{tap} represents the difference in voltage between two consecutive OLTC tap points; $\psi_{n,t}$ and $\chi_{n,t}$ are the positions of CBs and OLTCs in time step t, respectively; (12) constrains the active and reactive power range of PVs; (13)–(15) are the dynamic constraints of BESSs. $B_{n,t,\tau}$ represents the stored energy of BESS n in time τ during time t. $P^{C}_{n,\max}$ and $P^{D}_{n,\max}$ denote BESSs' maximum capacity for charging and discharging, respectively; (16) and (17) represent the active power constraint of FLs and reactive power constraint of SVCs, where $Q^{\text{SVC}}_{n,\min}$ and $Q^{\text{SVC}}_{n,\max}$ denote the minimum and maximum output of SVCs.

The following factors make $\mathbf{P1}$ challenging to solve. First of all, there are a lot of uncertain parameters, e.g., PV and load. Second, it may be challenging to obtain the explicit model information of the practical ADN. Third, hybrid devices have different regulating timescales. Fourth, there are operational limitations that are time-coupled in relation to OLTC, BESSs, and CBs. Finally, there are continuous and discrete decision variables. To overcome the above challenges, we intend to design a novel algorithm for $\mathbf{P1}$ based on HMAADRL without knowing accurate line parameters and prior knowledge of uncertainty parameters. To this end, we reformulate $\mathbf{P1}$ as bilevel Markov games.

B. Formulation of Bilevel Markov Games

In this article, the coordination of slow-timescale devices (i.e., OLTCs and CBs) are regarded as upper level Markov game, whereas the coordination of slow-timescale devices (i.e., SVCs, PV inverters, BESSs, and FLs) are regarded as lower level Markov game. Formally, a Markov Game with L agents usually includes a set of states $\mathcal{S}$, a set of actions $\mathcal{A}_1,\ldots\mathcal{A}_L$, a state transition function $\mathcal{F}$, and a reward function $\mathcal{R}_l(1 \leqslant l \leqslant L)$ [16], [31]. In this article, we assume X agents in upper level Markov game represent X controllers of OLTCs and CBs. Similarly, we assume I agents in lower level Markov game represent I controllers of subnetworks. The state transition function is unnecessary because the proposed algorithm is model free. Therefore, we focus on designing the state, action, and reward function related to solving $\mathbf{P1}$.

1) Upper-Level Markov Game:

1) *State:* The states $s^{u,\text{CB}}_{x,t}$ of agent x related to CB in time step t is designed as $s^{u,\text{CB}}_{x,t} = (P^{\text{PV}}_{x,t}, P^{\text{Load}}_{x,t}, Q^{\text{PV}}_{x,t-1}, Q^{\text{Load}}_{x,t}, V_{x,t}, \vartheta_{x,t}, \psi_{x,t-1})$, where $P^{\text{PV}}_{x,t}$ and $P^{\text{Load}}_{x,t}$ denote the active power injection of PV and load in its local subnetwork, respectively. Since OLTC can support regulate all bus voltages, its states in time step t is designed as $s^{u,\text{OLTC}}_{y,t} = (P^{\text{PV}}_{t}, P^{\text{Load}}_{t}, Q^{\text{PV}}_{t-1}, Q^{\text{Load}}_{t}, V_t, \vartheta_t, \chi_{y,t-1})$, where P^{PV}_{t} and P^{Load}_{t} are the active power injection of PV and load demand in all subnetworks, respectively.
2) *Action:* The action of agent x related to OLTC in time t is designed as $a^{u}_{x,t} = \chi_{x,t}$. The action of agent y related to CB in time t is designed as $a^{u}_{y,t} = \psi_{y,t}$.
3) *Reward:* OLTCs and CBs are responsible for regulating voltages within acceptable limits [i.e., 0.95–1.05 per unit (p.u.)] while minimizing the number of discrete device adjustments. Therefore, the reward of agent x in slow timescale is designed as

$$r_{x,t} = -(r_{x,t,1} + \beta_1 r_{x,t,2}) \tag{18}$$

where $r_{x,t,1} = \sum_{\tau=1}^{\Gamma} r^{1}_{i,t,\tau}$ denotes the penalty of voltages crossing the safe range at time t. $r^{1}_{i,t,\tau}$ denotes the penalty of voltages crossing the acceptable limits in time τ during time t, which is designed in the lower level Markov game. $r_{x,t,2} = Z_{x,t}$ denotes the adjustments of OLTCs and CBs. β_1 is the coefficient to balance voltage deviations and adjustments of OLTCs and CBs.

2) Lower Level Markov Game:

1) *State:* $s^{l}_{i,t,\tau}$ is designed as the state of lower level agent related to subnetwork i $(1 \le i \le I)$, which contains seven parts: $s^{l}_{i,t,\tau} = (P^{\text{PV}}_{i,t,\tau}, P^{\text{Load}}_{i,t,\tau}, Q^{\text{PV}}_{i,t,\tau-1}, Q^{\text{Load}}_{i,t,\tau}, V_{i,t,\tau}, \vartheta_{i,t,\tau}, B_{i,t,\tau})$, where $Q^{\text{PV}}_{i,t,\tau-1}$ represents the reactive power injection of PV in subnetwork i in time $\tau-1$ during time t.
2) *Action:* Lower level agent's actions are designed as $a^{l}_{i,t,\tau} = (Q^{\text{PV}}_{i,t,\tau}, Q^{\text{SVC}}_{i,t,\tau}, P^{\text{BESS}}_{i,t,\tau}, \Delta P^{\text{FL}}_{i,t,\tau})$, where $Q^{\text{PV}}_{i,t,\tau}$ and $Q^{\text{SVC}}_{i,t,\tau}$ represent the reactive power output of PVs and SVCs. $P^{\text{BESS}}_{i,t,\tau}$ denotes the charging or discharging active power of BESSs. $\Delta P^{\text{FL}}_{i,t,\tau}$ is the scheduling amount of FLs.
3) *Reward:* Since both upper and lower level agents are responsible for voltage regulation together, the penalty $r^{1}_{i,t,\tau}$ of bus voltages exceeding the safe range is regarded as a partial reward for both upper level and lower level agents, where $r^{1}_{i,t,\tau} = \sum_{n=1}^{N_i} \Delta V_{i,t,\tau}$. N_i represents the total number of buses in subnetwork i. In addition, the system power loss $P^{L}_{t,\tau} = \sum_{n=1}^{N}\sum_{m=1}^{N} P^{L}_{n,m,t,\tau}$ of the ADN in time τ during time t should be optimized at the same time. Moreover, since the frequent dispatch of FLs and excessive use of BESSs will increase the system cost, the dispatching of BESSs and FLs also needs to be optimized. Comprehensively consider four parts, the reward of lower level agent i can be computed by

$$r_{i,t,\tau} = -(r^{1}_{i,t,\tau} + \beta_2 P^{L}_{t,\tau} + \varepsilon_1 P^{\text{BESS}}_{i,t,\tau} + \varepsilon_2 \Delta P^{\text{FL}}_{i,t,\tau}) \tag{19}$$

where β_2, ε_1, and ε_2 denote the weighted coefficients to balance voltage deviations, power losses, and dispatched active power of BESSs and FLs.

III. HMAADRL-Based Voltage Regulation Algorithm

We propose a HMAADRL-based voltage regulation algorithm to solve the above bilevel Markov games. The proposed algorithm's framework is shown in Fig. 2, where three unique features different from existing DRL-based algorithms can be identified. First, the proposed algorithm's framework consists of two-level MADRL algorithms for slow-timescale and fast-timescale voltage regulation, respectively. Second, MAAC algorithm is used to train DRL multiple agents in each level. Since MAAC adopts SAC, attention mechanism, multitask learning,

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

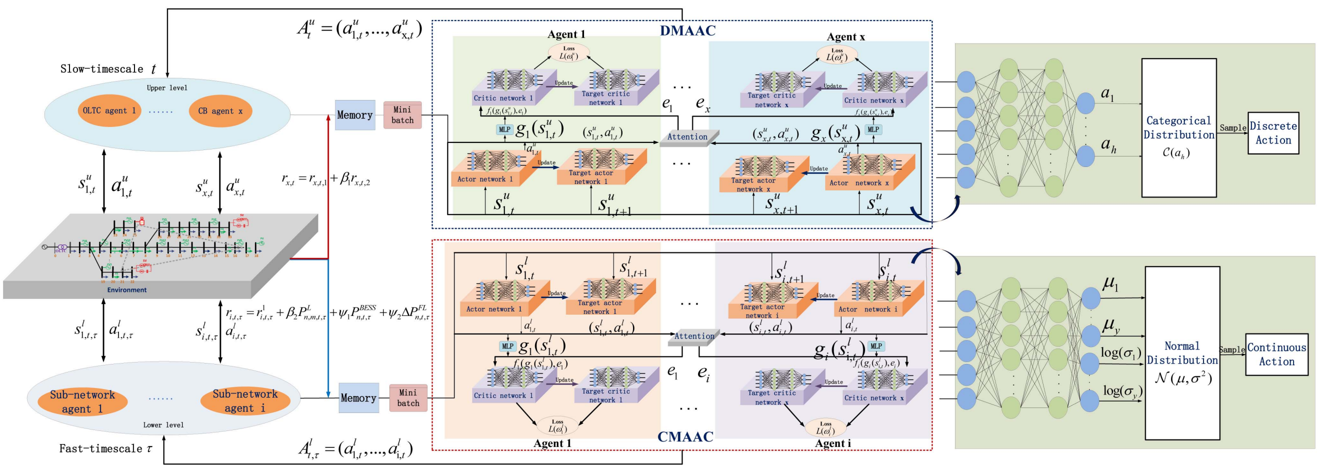


Fig. 2. Framework of HMAADRL-based voltage regulation algorithm.

and multi-agent advantage function, it can achieve better performance compared with many algorithms, such as MADDPG and MASAC, in existing works. Third, to accommodate discrete and continuous devices simultaneously, DMAAC and CMAAC are designed in two levels, respectively.

A. DMAAC-Based Voltage Regulation Algorithm

In the upper level DMAAC-based voltage regulation algorithm, each agent related discrete device consists of two actor networks and two critic networks. We define $Q_y(s_t, a_t)$ as the centralized action-value function to assess the actions of upper level agent y. Given a state s_t and an action a_t, $Q_y(s_t, a_t)$ is described as

$$Q_y(s_t, a_t) = \mathbb{E}\left\{\sum_{d}^{\infty} \gamma^d r_{y,t+d}(s_t, a_t)\right\} \tag{20}$$

where $\mathbb{E}$ is the expectation operator. $\gamma \in [0, 1]$ represents the discount factor utilized to determine the impact of the current policy on future long-term rewards. In addition, $Q_y(s, a; \omega_y)$ is defined to approximate action-value function, where ω_y denote the critic network parameters. Based on temporal-difference learning, the network parameters ω_y are updated by lowering the subsequent loss function

$$L(\omega_y) = \mathbb{E}_{(s,a,\overline{s},r)\sim M}\{(Q_y(s, a; \omega_y) - (r_y + \gamma \mathbb{E}_{\overline{a}\sim\pi(\overline{s})} Q_y(\overline{s}, \overline{a}; \overline{\omega}_y)))^2\} \tag{21}$$

where $Q_y(\overline{s}, \overline{a}; \overline{\omega}_y)$ denotes the target action-value function. Note that s, a, $\overline{s}$, r belong to M, where M represents the experience replay buffer that stores past experiences.

Similarly, we define $\pi_y(a|s; \theta_y)$ as the approximate actor policy function of $\pi_y(a|s)$, where θ_y are the actor networks' parameters. Then, θ_y can be updated by policy gradient method, which can be calculated by

$$\nabla_{\theta_y} J(\theta_y) = \mathbb{E}_{s,a\sim M}\{\nabla_{\theta_y} \log(\pi(a\,|s\,; \theta_y) Q_y(s, a; \omega_y))\}. \tag{22}$$

The core purpose of the policy gradient is to maximize the goal by moving in the direction of $\nabla_{\theta_y} J(\theta_y)$ by directly adjusting the strategy's parameters θ_y.

To facilitate the performance of training actor networks, SAC method and attention mechanism are adopted. Incorporating an entropy element into the policy gradient and learning a soft value function is the main goal of the SAC approach. Then, (22) can be rewritten as follows:

$$\nabla_{\theta_y} J(\theta_y) = \mathbb{E}_{s,a\sim M}\{\nabla_{\theta_y} \log(\pi(a\,|s\,; \theta_y) (-\mu \log(\pi(a\,|s\,; \theta_y)) + Q_y(s, a; \omega_y) - d(s))\} \tag{23}$$

where μ denotes the temperature parameter that is taken to equalize the weight between $\log(\pi(a|s; \theta_y)$ and $Q_y(s, a; \omega_y)$. $d(s)$ is the state-dependent baseline. In addition, the loss function $L(\omega_y)$ is accordingly reformulated as follows:

$$L(\omega_y) = \mathbb{E}_{(s,a,\overline{s},r)\sim M}\{(Q_y(s, a; \omega_y) - (r_y + \gamma \mathbb{E}_{\overline{a}\sim\overline{\pi}(\overline{s})}[Q_y(\overline{s}, \overline{a}; \overline{\omega}_y) - \mu \log(\overline{\pi}(\overline{a}\,|\overline{s}\,; \overline{\theta}_y))])^2\} \tag{24}$$

where $\overline{\pi}(\overline{a}|\overline{s}; \overline{\theta}_y)$ is the target policy function whose parameters of the target actor network are $\overline{\theta}_y$.

By introducing an attention mechanism, each agent chooses whatever information about other agents to focus on when calculating the action-value function $Q_y(s, a; \omega_y)$ [32]. $Q_y(s, a; \omega_y)$ is further described as

$$Q_y(s, a; \omega_y) = f_y(g_y(s_y), e_y) \tag{25}$$

where f_y denotes a two-layer multilayer perceptron (MLP); g_y represents a one-layer MLP; e_y represents the weighted contribution of other agents to agent y. e_y is designed as follows:

$$e_y = \sum_{y\neq z} \xi_z l(Y g_z(s_z, a_z)) \tag{26}$$

where Y denotes the nonlinear transformation matrix; l is the activation function; ξ_z is the attention weight that agent y pays

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

for agent z. ξ_z can be described as

$$\xi_z = \frac{\exp({f_y}^T(s_y, a_y){U_b}^T U_d f_z(s_z, a_z))}{\sum_{y \neq z} \exp({f_y}^T(s_y, a_y){U_b}^T U_d f_z(s_z, a_z))} \tag{27}$$

where U_b and U_d represent the transition matrices.

For the upper level devices, such as OLTCs and CBs, discrete control variables need to be designed. The approximate actor policy function $\pi(a|s;\theta)$ can be designed as a categorical distribution. We assume the actor network of upper level agent x has h-dimensional discrete actions. The "softmax" function is applied to the output layer, which can normalize the output value. Then, the categorical distribution $\mathcal{C}(a_h)$ of these discrete actions is established. In this sense, the discrete action a_x can be obtained by sampling from $\mathcal{C}(a_h)$ as follows:

$$a_x = \text{categorical\_sample}(\mathcal{C}(a_h)) \tag{28}$$

where categorical_sample$(\cdot)$ is used to select one expected discrete action.

B. CMAAC-Based Voltage Regulation Algorithm

Similar to the upper level DMAAC-based voltage regulation algorithm, each agent related to the continuous device in the lower level CMAAC-based voltage regulation algorithm also consists of two actor networks, one of which is the target actor network. Moreover, each lower level agent contains two critic networks, one of which is the target critic network. The parameter update rule of networks is similar to DMAAC algorithm, as shown in (20)–(27). While for the continuous action space, the approximate actor policy function is designed as a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$, where the mean μ and the standard deviation σ^2 can be calculated and optimized by the actor network. Then, the continuous action a_i can be obtained by sampling from $\mathcal{N}(\mu, \sigma^2)$ as follows:

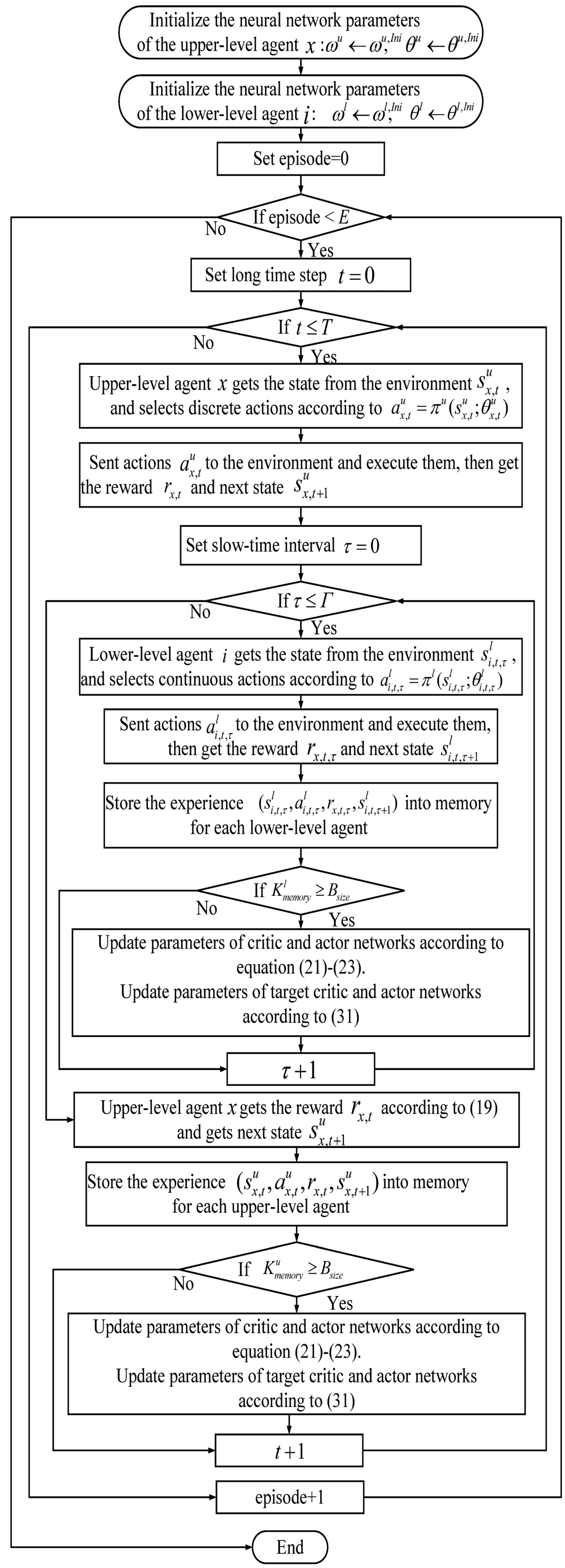
$$a_i = \tanh(\text{sample}(\mathcal{N}(\mu, \sigma^2))) \tag{29}$$

where $\tanh(\cdot)$ is the activation function.

C. Details of the Proposed Algorithm

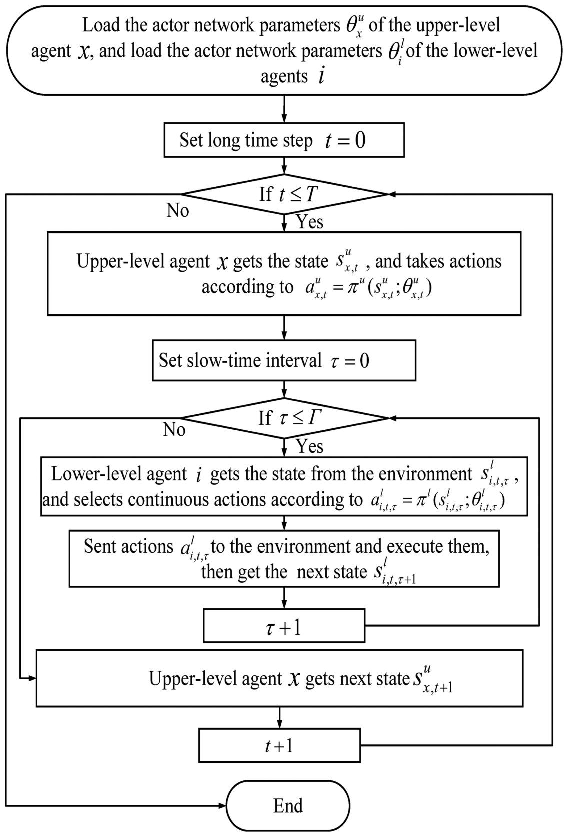
The proposed HMAADRL-based voltage regulation algorithm contains a centralized algorithm training process and a decentralized execution process. The collaborative training process is described in Algorithm 1. At the beginning of the proposed algorithm, the neural network parameters, i.e., ω^u, θ^u, ω^l, θ^l are first initialized. Then, these parameters are updated through E episodes of learning. In slow timescale, the upper level agents related to OLTC and CBs obtain the states of the ADN and make actions according to $a^u_{x,t} = \pi^u(s^u_{x,t}; \theta^u_{x,t})$ at each time step t. Then, the lower level agents get current states and take actions based on $a^l_{i,t,\tau} = \pi^l(s^l_{i,t,\tau}; \theta^l_{i,t,\tau})$. After all actions of lower level agents are carried out, an instant reward will be given for each agent, and the ADN environment moves to the next state of time interval τ. Next, the experience tuple $(s^l_{i,t,\tau}, a^l_{i,t,\tau}, r_{x,t,\tau}, s^l_{i,t,\tau+1})$ is further stored into the lower level experience memory K^l_{memory}. When the length of the buffer

Algorithm 1: Training Process of the Proposed Algorithm.



This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Algorithm 2: Execution Process of the Proposed Algorithm.



K_{memory}^l is greater than the length of the batch size B_{size}, the parameters of lower level neural networks are optimized and updated based on (21)–(23). At the same time, the lower level target networks' parameters are updated by

$$\overline{\omega}^l \leftarrow \zeta\omega^l + (1-\zeta)\overline{\omega}^l, \overline{\theta}^l \leftarrow \zeta\theta^l + (1-\zeta)\overline{\theta}^l \quad (30)$$

where $\overline{\omega}^l$ and $\overline{\theta}^l$ represent the parameters of lower level target networks, and ζ denotes the soft update coefficient.

When Γ time intervals are completed, the upper level agents get the reward according to (19), and the environment moves to the next state of time step t. Similar to lower level training process, the experience tuple $(s_{x,t}^u, a_{x,t}^u, r_{x,t}, s_{x,t+1}^u)$ is also stored into the upper level experience memory K_{memory}^u every one time step. Then, a mini-batch experience sampled from upper level memory is used to train the parameters of upper level neural networks when $K_{\text{memory}}^u \ge B_{\text{size}}$. In addition, the upper level target networks' parameters are updated by

$$\overline{\omega}^u \leftarrow \zeta\omega^u + (1-\zeta)\overline{\omega}^u, \overline{\theta}^u \leftarrow \zeta\theta^u + (1-\zeta)\overline{\theta}^u \quad (31)$$

where $\overline{\omega}^u$ and $\overline{\theta}^u$ represent the parameters of upper level target networks.

When the proposed algorithm has finished its training procedure, the critic networks are no longer used, and the parameters of actor networks will no longer be updated. The forward propagation of each actor network is only computed. Therefore, the complexity of the proposed algorithm depended on the

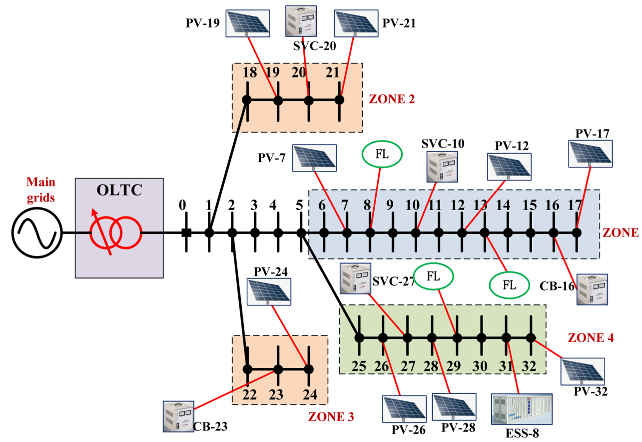


Fig. 3. IEEE 33-bus test feeder system.

forward propagation. Specifically, three types of computation (e.g., addition, multiplication, and activation) are engaged in the forward propagation process. We define U_{in}, U_{hid}, and U_{out} that represent the total number of neurons in the input, hidden, and output layer, respectively. Then, the number of addition, multiplication, and activation of the first neuron in the first hidden, is $U_{\text{in}}-1$, U_{in}, and 1, respectively. Therefore, there have been a total of $2U_{\text{in}}U_{\text{hid}}$ computations in the input layer. Similarly, the total number of computations in the second hidden layer is $2U_{\text{hid}}U_{\text{hid}}$. The total number of computations in the output layer is $2U_{\text{hid}}U_{\text{out}}$. Finally, we can compute the total complexity of the proposed algorithm by $2(U_{\text{in}}U_{\text{hid}} + U_{\text{hid}}U_{\text{hid}} + U_{\text{hid}}U_{\text{out}})$. The detailed execution procedure of the HMAADRL-based voltage regulation algorithm is introduced in Algorithm 2. To be specific, the upper level agents calculate actions according to $a_{x,t}^u = \pi^u(s_{x,t}^u;\theta_{x,t}^u)$ and execute them at time step t. Then, lower level agents get the environmental information based on the decisions of upper level agents and make corresponding actions based on $a_{i,t,\tau}^l = \pi^l(s_{i,t,\tau}^l;\theta_{i,t,\tau}^l)$. When the lower level agents continuously execute actions of all time slots in Γ, upper level agents make next actions at time step $t+1$. The algorithm repeats the aforementioned procedure until the testing phase is completed.

Remark: It is widely recognized that DRL-based techniques call for a sizable number of training samples. It is quite difficult to acquire such samples by directly interacting with the real ADN system due to the lengthy exploration time and high exploration expense [33]. A viable option is to create a simulation model of the real ADN using digital twin technology. Therefore, a digital twin model related to ADNs can be developed and used for DRL agents during the actual implementation process.

IV. Performance Analysis

A. Simulation Setup

In simulations, the IEEE 33-bus test feeder system embedded with 1 OLTC, 2 CBs, 9 PVs, 3 SVCs, 1 BESS, and 3 FLs, as shown in Fig. 3, is employed to evaluate the efficacy of the proposed algorithm. The desired security scope of voltage is set

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I
PARAMETERS OF DEVICES IN IEEE 33-BUS TEST FEEDER SYSTEM

Device	Capacity	Location
OLTC	$\pm 5 \times 1\%$	1
CB	5×200 Kvar	16, 23
BESS	1 MVA	31
SVC	600 KVar	10, 20, 27
PV	2.1 MVA	7, 12, 17, 19, 21, 24, 26, 28, 32
FL	$\pm 20\% \times P_n^{\mathrm{FL}}$	8, 13, 29

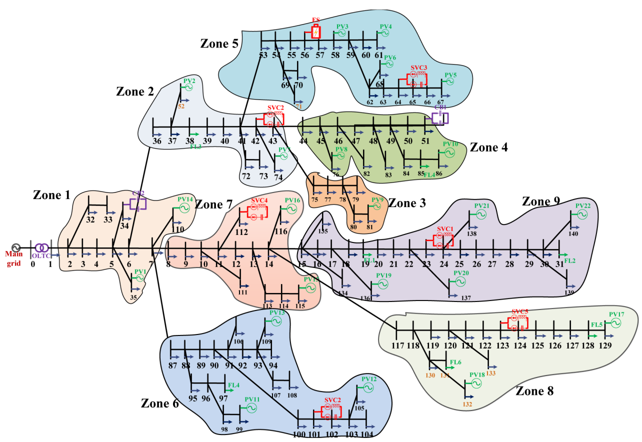


Fig. 4. IEEE 141-bus test feeder system.

TABLE II
PARAMETERS OF DEVICES IN IEEE 141-BUS TEST FEEDER SYSTEM

Device	Capacity	Location
OLTC	$\pm 5 \times 1\%$	1
CB	5×200 Kvar	34, 51
BESS	1 MVA	56
SVC	600 KVar	23, 42, 64, 101, 112, 123
PV	7.5 MVA	35, 42, 58, 61, 67, 68, 74, 76, 81, 86, 102, 105, 109, 110, 115, 116, 129, 132, 136, 137, 138, 140
FL	$\pm 20\% \times P_n^{\mathrm{FL}}$	19, 31, 38, 85, 97, 128, 131

as [0.95, 1.05 p.u.]. Detailed parameters and partition regions of the ADN can be found in [30]. The PVs and loads data in 2012–2015 are collected from Elia group[1] and Portuguese electricity consumption[2] for training, respectively. The data of five days in the summer of 2012 are used for testing. The capacity and location of several equipments are given in Table I. Moreover, the IEEE 141-bus test feeder system [30], as shown in Fig. 4, is utilized to assess the scalability of the proposed HMAADRL-based voltage control algorithm. The detailed parameters can be seen in Table II. The simulation experiment is implemented on a computer with a 3.50 GHz Intel Core i9-11900 K, a 3090 GPU, and 128 GB RAM.

[1][Online] Available: https://www.elia.be/en/grid-data/power-generation/solar-pv-power-generation-data.
[2][Online] Available: https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014.

TABLE III
PARAMETERS OF THE PROPOSED ALGORITHM

Parameter	Discrete MAAC	Continuous MAAC
Number of neurons in hidden layer about actor network	128	96
Number of neurons in hidden layer about critic network	128	96
Batch size (B_{size})	120	128
Discount factor (γ)	0.995	0.99
Memory size (M_{size})	2.4e4	5e3
Learning rate of actor network (α_a)	1e-4	8e-5
Learning rate of critic network (α_c)	1e-3	8e-4

In addition, for DMAAC of the proposed algorithm, all actor networks have similar network architecture. Specifically, one input layer, one hidden layer with Leaky ReLU activation functions, and one output layer with a softmax activation function make up the actor networks. In addition, the network architecture is the same for all critic networks. To be specific, one input layer, one hidden layer with leaky ReLU activation functions, and one output layer with a linear activation function make up the critic network of DMAAC. Similarly, for CMAAC, the network architecture is the same for all actor and critic networks. Specifically, one input layer, one hidden layer with linear activation functions, and one output layer with a ReLU activation function make up each actor network. Moreover, one input layer, one hidden layer with Leaky ReLU activation functions, and one output layer with a linear activation function make up the critic networks. The detailed parameters of the network are presented in Table III.

B. Benchmarks

Five baselines are designed to compare performance, and they are as follows.

1) *Baseline1 (B1)* is the basic scheme without any control of voltage.
2) *Baseline2 (B2)* only regulates the upper level devices, such as OLTCs and CBs, via the proposed algorithm.
3) *Baseline3 (B3)* uses the droop control strategy as described in IEEE Std-1547-2018 [34]. Specifically, the voltage regulation is realized by controlling the smart PV inverters' reactive power output using the droop control approach.
4) *Baseline4 (B4)* uses the SAC method to train upper level agents and uses the MASAC algorithm to train lower level agents for cooperative voltage control [29].
5) *Baseline5 (B5)* uses the DQN algorithm to train upper level agent, which is used to control discrete devices and uses the model-based SOCP method to control continuous devices [26].

C. Convergence Analysis

Three learning-based algorithms are trained, and the training processes of upper level and lower level algorithms are shown

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

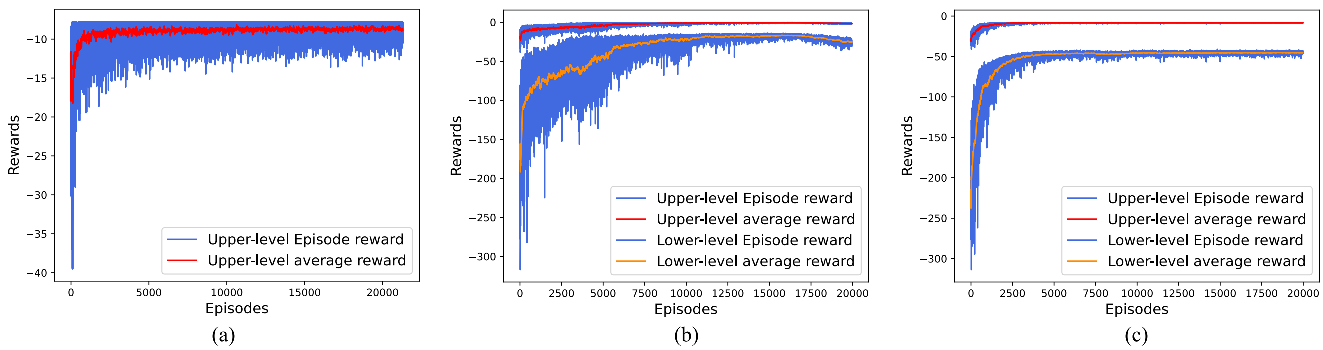


Fig. 5. Training processes of three learning-based algorithms. (a) B2. (b) B4. (c) Proposed.

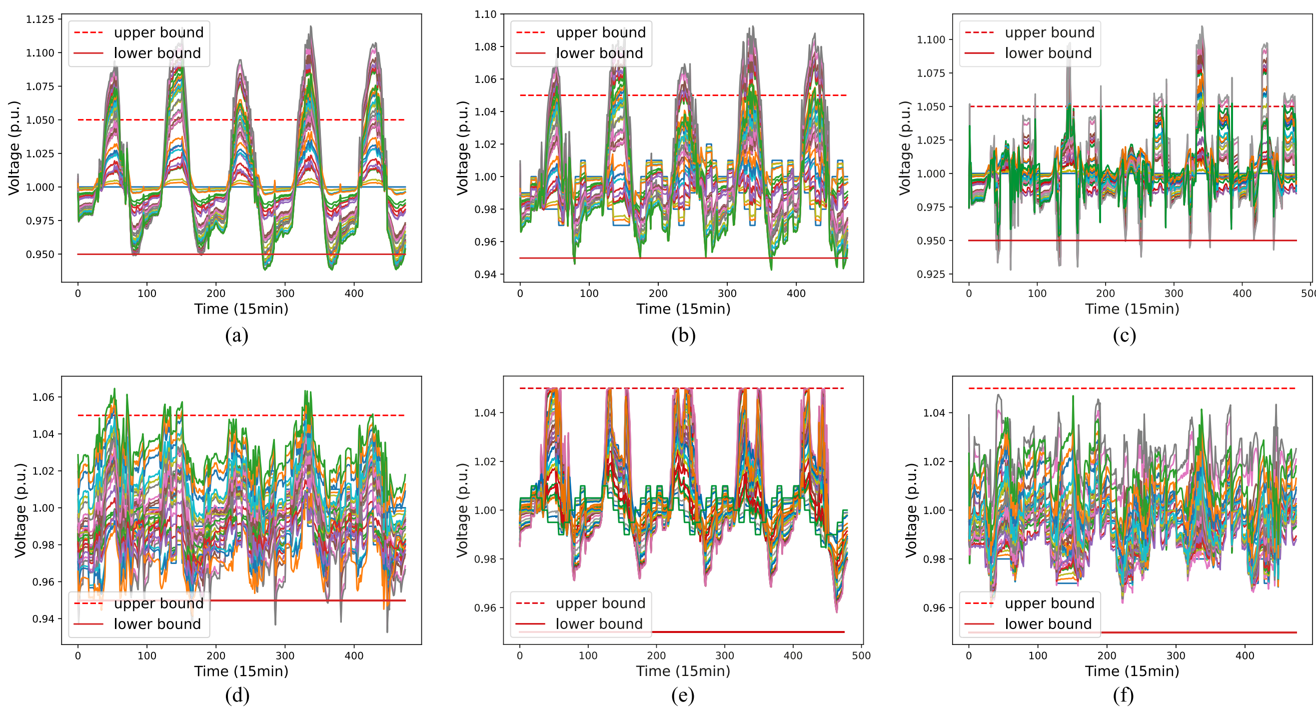


Fig. 6. Voltage profiles of 33 buses achieved by different baselines. (a) B1. (b) B2. (c) B3. (d) B4. (e) B5. (f) Proposed.

in Fig. 5. Since B2 controls the upper level discrete devices for voltage regulation, only the upper level reward curve is given. We can see that the reward curves of the proposed algorithm are more stable and have better convergence performance compared with B4 since the proposed algorithm adopted the attention mechanism in both upper and lower levels. In addition, the upper level rewards of the proposed algorithm have smaller fluctuations than B2, and the reason is that B2 only controls discrete devices that lack resources for coordinated voltage regulation.

D. Performance Analysis

Voltage regulation profiles and average voltage deviations under different baselines are given in Figs. 6 and 7(a). It can be observed that the proposed algorithm can regulate all bus voltages to the desired range compared with B1–B4. The reason is that B2 and B3 only regulate the upper level discrete devices or the lower level continuous devices. It is challenging for B2 and B3 to control the voltage to the safe range when high-penetrated PVs are connected to the ADN. Although B4 considers all resources for voltage regulation, its cooperative performance is unsatisfactory due to the absence of attention mechanism. In addition, from Fig. 7(b), the proposed algorithm can reduce the power loss by 27.05% and 7.59% compared with B3 and B4, respectively. It should be noted that due to the high voltage amplitudes of B1 and B2, their power losses have not been compared. Moreover, the control performance of model-based voltage regulation method B5, is given in Figs. 6(e), 7(a), and (b). We can see that both B5 and the proposed algorithm can regulate all bus voltage in the safe range. Furthermore, the relative power loss gap between the proposed algorithm, and B5 is less than 3%. However, B5 requires the accurate line parameters of the ADN

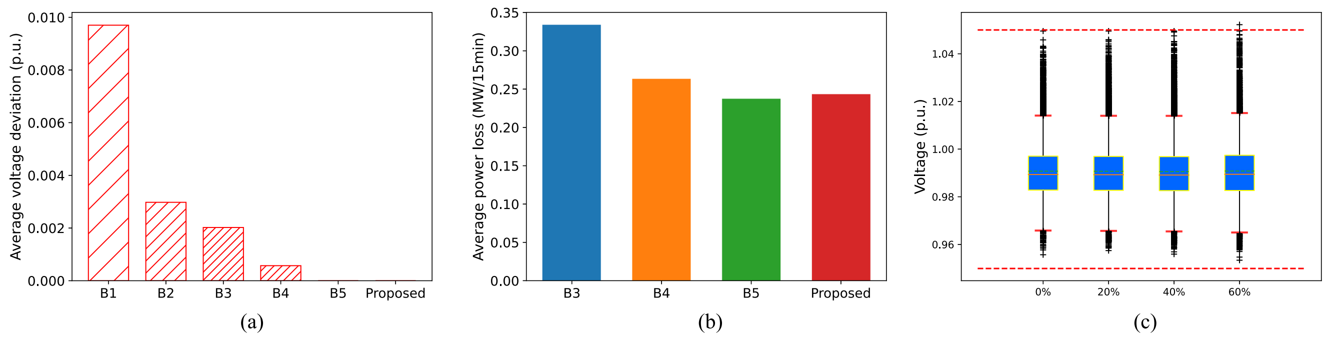


Fig. 7. Comparison results on the IEEE 33-bus test feeder system. (a) Average voltage deviation. (b) Average power loss. (c) Voltage distributions versus disturbances.

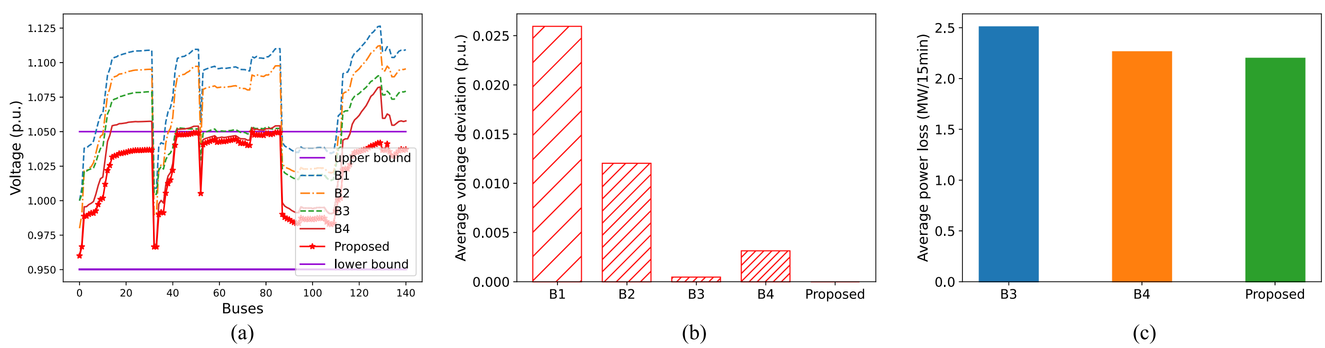


Fig. 8. Comparison results on the IEEE 141-bus test feeder system. (a) Voltage profiles (t=12:00 am). (b) Average voltage deviation. (c) Average power loss.

TABLE IV
COMPARISON RESULTS FOR DIFFERENT BASELINES

Method	AAN-OLTC (times/day)	AAN-CBs (times/day)	ARP-BESS (MW/15min)	ARP-FLs (MW/15min)	ACT (s)
B2	9.4	8.8	-	-	0.0056
B3	-	-	-	-	0.0062
B4	8.8	8.9	0.362	0.0586	0.0073
B5	11.6	9.0	0.237	0.256	7.2461
Proposed	**7.4**	**8.0**	**0.312**	**0.029**	**0.0067**

and needs to precisely predict the load and renewable energy generation information, whereas the proposed algorithm does not require above information.

In addition, the average adjustment number (AAN) of discrete devices, the average regulation power (ARP) of BESSs and FLs, and the average computational time (ACT) of each action during testing periods are given in Table IV. It is evident that the proposed algorithm has the lowest adjustments of discrete devices. Meanwhile, the proposed algorithm dispatches less active power of the BESS and FLs for voltage regulation compared with B4, which indicates that the proposed algorithm has a strong synergy ability. Moreover, the average computational time of the proposed algorithm is much lower than B5 and close to that of B3, which can meet practical engineering requirements.

E. Robustness Analysis

To evaluate the robustness of the proposed algorithm, we show the voltage regulation performance in Fig. 7(c), when 20%, 40%, and 60% of line parameter disturbances are injected, respectively. It can be seen that the proposed algorithm can still regulate the voltage to a safe range when the disturbance is increased to 40%. When 60% of line parameter disturbance is injected, voltage curves slightly cross the safe boundary. Therefore, the proposed algorithm is robust to line parameter uncertainties.

F. Scalability Analysis

To further verify the scalability of the proposed algorithm, voltage regulation performances of different baselines are shown in Fig. 8, where the IEEE 141-bus test feeder system with 22 PVs is considered. Since solving the voltage regulation optimization problem in IEEE 141-test feeder system is time-consuming under the model-based method B5, we just compare the voltage regulation performance of the proposed algorithm with B1–B4. We can see that several bus voltages cross the safe boundary when no control is performed. However, the proposed algorithm can regulate all bus voltages to a safe range and has less power loss compared with B3 and B4, demonstrating the effectiveness and scalability of the proposed algorithm.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

V. Conclusion

This article studied a multitimescale voltage optimization problem of ADNs considering the coordination of hybrid devices while minimizing the total power loss. Due to the solving challenges, the problem was reformulated as bilevel Markov games, and we proposed a HMAADRL-based voltage regulation algorithm to solve them. The proposed algorithm could achieve cooperative voltage regulation considering discrete, continuous, and multitimescale hybrid devices without knowing the exact model information of ADNs. Simulation results based on IEEE 33-bus test feeder system and IEEE 141-bus test feeder system showed the effectiveness, robustness, and scalability of the proposed algorithm.

References

[1] Z. Zhang, Y. Zhang, D. Yue, C. Dou, L. Ding, and D. Tan, "Voltage regulation with high penetration of low-carbon energy in distribution networks: A source–grid-load-collaboration-based perspective," *IEEE Trans. Ind. Informat.*, vol. 18, no. 6, pp. 3987–3999, Jun. 2022.

[2] J. Zhao et al., "Cloud-edge collaboration-based local voltage control for DGs with privacy preservation," *IEEE Trans. Ind. Informat.*, vol. 19, no. 1, pp. 98–108, Jan. 2023.

[3] P. Li et al., "Coordinated control method of voltage and reactive power for active distribution networks based on soft open point," *IEEE Trans. Sustain. Energy*, vol. 8, no. 4, pp. 1430–1442, Oct. 2017.

[4] S. Huang, Q. Wu, Y. Guo, X. Chen, B. Zhou, and C. Li, "Distributed voltage control based on ADMM for large-scale wind farm cluster connected to VSC-HVDC," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 584–594, Apr. 2020.

[5] Y. Xu, Z. Y. Dong, R. Zhang, and D. J. Hill, "Multi-timescale coordinated voltage/Var control of high renewable-penetrated distribution systems," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4398–4408, Nov. 2017.

[6] K. Christakou, M. Paolone, and A. Abur, "Voltage control in active distribution networks under uncertainty in the system model: A robust optimization approach," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 5631–5642, Nov. 2018.

[7] T. Soares, R. J. Bessa, P. Pinson, and H. Morais, "Active distribution grid management based on robust AC optimal power flow," *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6229–6241, Nov. 2018.

[8] P. Li, C. Zhang, Z. Wu, Y. Xu, M. Hu, and Z. Dong, "Distributed adaptive robust voltage/Var control with network partition in active distribution networks," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2245–2256, May 2020.

[9] D. Jin, H.-D. Chiang, and P. Li, "Two-timescale multi-objective coordinated volt/Var optimization for active distribution networks," *IEEE Trans. Power Syst.*, vol. 34, no. 6, pp. 4418–4428, Nov. 2019.

[10] R. R. Jha, A. Dubey, C.-C. Liu, and K. P. Schneider, "Bi-level volt-Var optimization to coordinate smart inverters with voltage control devices," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1801–1813, May 2019.

[11] M. M.-U.-T. Chowdhury and S. Kamalasadan, "A new second-order cone programming model for voltage control of power distribution system with inverter-based distributed generation," *IEEE Trans. Ind. Appl.*, vol. 57, no. 6, pp. 6559–6567, Nov./Dec. 2021.

[12] R. Zafar, J. Ravishankar, J. E. Fletcher, and H. R. Pota, "Multi-timescale voltage stability-constrained volt/Var optimization with battery storage system in distribution grids," *IEEE Trans. Sustain. Energy*, vol. 11, no. 2, pp. 868–878, Apr. 2020.

[13] W. Zheng, W. Wu, B. Zhang, and Y. Wang, "Robust reactive power optimisation and voltage control method for active distribution networks via dual time-scale coordination," *IET Gener., Transmiss. Distrib.*, vol. 11, no. 6, pp. 1461–1471, 2017.

[14] P. Li et al., "Deep reinforcement learning-based adaptive voltage control of active distribution networks with multi-terminal soft open point," *Int. J. Elect. Power Energy Syst.*, vol. 141, pp. 108138–108147, 2022.

[15] D. Cao et al., "Data-driven multi-agent deep reinforcement learning for distribution system decentralized voltage control with high penetration of PVs," *IEEE Trans. Smart Grid*, vol. 12, no. 5, pp. 4137–4150, Sep. 2021.

[16] T. Zhang, D. Yue, L. Yu, C. Dou, and X. Xie, "Joint energy and workload scheduling for fog-assisted multimicrogrid systems: A deep reinforcement learning approach," *IEEE Syst. J.*, vol. 17, no. 1, pp. 164–175, Mar. 2022.

[17] T. Zhang, L. Yu, D. Yue, C. Dou, X. Xie, and L. Chen, "Coordinated voltage regulation of high renewable-penetrated distribution networks: An evolutionary curriculum-based deep reinforcement learning approach," *Int. J. Elect. Power Energy Syst.*, vol. 149, 2023, Art. no. 108995.

[18] L. Yu et al., "Deep reinforcement learning for smart home energy management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020.

[19] L. Yu, Z. Xu, X. Guan, Q. Zhao, C. Dou, and D. Yue, "Joint optimization and learning approach for smart operation of hydrogen-based building energy systems," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 199–216, Jan. 2023.

[20] Y. Liang, Z. Ding, T. Zhao, and W.-J. Lee, "Real-time operation management for battery swapping-charging system via multi-agent deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 14, no. 1, pp. 559–571, Jan. 2023.

[21] Y. Lu, Y. Liang, Z. Ding, Q. Wu, T. Ding, and W.-J. Lee, "Deep reinforcement learning-based charging pricing for autonomous mobility-on-demand system," *IEEE Trans. Smart Grid*, vol. 13, no. 2, pp. 1412–1426, Mar. 2022.

[22] Y. Liang, Z. Ding, T. Ding, and W.-J. Lee, "Mobility-aware charging scheduling for shared on-demand electric vehicle fleet using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 1380–1393, Mar. 2021.

[23] R. Lu, Y.-C. Li, Y. Li, J. Jiang, and Y. Ding, "Multi-agent deep reinforcement learning based demand response for discrete manufacturing systems energy management," *Appl. Energy*, vol. 276, 2020, Art. no. 115473.

[24] Y. Wang, D. Qiu, G. Strbac, and Z. Gao, "Coordinated electric vehicle active and reactive power control for active distribution networks," *IEEE Trans. Ind. Informat.*, vol. 19, no. 2, pp. 1611–1622, Feb. 2022.

[25] S. Wang, L. Du, X. Fan, and Q. Huang, "Deep reinforcement scheduling of energy storage systems for real-time voltage regulation in unbalanced LV networks with high PV penetration," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 2342–2352, Oct. 2021.

[26] Q. Yang, G. Wang, A. Sadeghi, G. B. Giannakis, and J. Sun, "Two-timescale voltage control in distribution grids using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 11, no. 3, pp. 2313–2323, May 2020.

[27] X. Sun and J. Qiu, "Two-stage volt/Var control in active distribution networks with multi-agent deep reinforcement learning method," *IEEE Trans. Smart Grid*, vol. 12, no. 4, pp. 2903–2912, Jul. 2021.

[28] H. Liu, W. Wu, and Y. Wang, "Bi-level off-policy reinforcement learning for two-timescale volt/Var control in active distribution networks," *IEEE Trans.Power Syst.*, vol. 38, no. 1, pp. 385–395, Jan. 2023.

[29] D. Cao et al., "Deep reinforcement learning enabled physical-model-free two-timescale voltage control method for active distribution systems," *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 149–165, Jan. 2022.

[30] J. Wang, W. Xu, Y. Gu, W. Song, and T. Green, "Multi-agent reinforcement learning for active voltage control on power distribution networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, vol. 34, pp. 3271–3284.

[31] L. Yu et al., "Multi-agent deep reinforcement learning for HVAC control in commercial buildings," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 407–419, Jan. 2021.

[32] S. Iqbal and F. Sha, "Actor-attention-critic for multi-agent reinforcement learning," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 2961–2970.

[33] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, and X. Guan, "A review of deep reinforcement learning for smart building energy management," *IEEE Internet Things J.*, vol. 8, no. 15, pp. 12046–12063, Aug. 2021.

[34] D. G. Photovoltaics and E. Storage, *IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces*, IEEE Standard 1547–2018, 2018.

Tingjun Zhang received the M.S. degree in control theory and control engineering from the Guangxi University of Science and Technology, Liuzhou, China, in 2019, and the Ph.D. degree in information acquisition and control from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2023.

He is currently a Postdoctoral Fellow at Nanjing University of Posts and Telecommunications. His research interests include power system operation and voltage control, energy management and optimization, distributed optimization, and deep reinforcement learning.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Liang Yu (Senior Member, IEEE) received the B.E. degree in communication engineering and the M.E. degree in communications and information systems from Yangtze University, Jingzhou, China, in 2007 and 2010, respectively, and the Ph.D. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2014.

He was a Postdoctoral Fellow with Xi'an Jiaotong University, Xi'an, China. He is currently a Professor and Doctoral Supervisor with the College of Automation & College of Artificial Intelligence, the Nanjing University of Posts and Telecommunications, Nanjing, China. His current research interests include cyber-physical systems (e.g., smart grids, data centers, and smart buildings), cloud-fog computing, distributed optimization, and deep reinforcement learning.

Dong Yue (Fellow, IEEE) received the Ph.D. degree in control theory and control engineering from the South China University of Technology, Guangzhou, China, in 1995.

He is currently a Professor, the Dean of the Institute of Advanced Technology and College of Automation & Artificial Intelligence, and Director of Academic Committee of the University at Nanjing University of Posts and Telecommunication, Nanjing, China. He has authored and coauthored more than 400 papers in prestigious international journals. His current research interests include networked control, optimization, multiagent systems, and smart grid.

Dr. Yue was the recipient of 2022 IEEE Rudolf Chope Research & Development Award, Norbert Wiener Review Award by IEEE/CAA JOURNAL OF AUTOMATICA SINICA in 2020, and the Best Paper Award of IEEE SYSTEMS JOURNAL, in 2022. He is the Chair of IEEE IES Technical Committee on NCS and applications. He is the Coeditor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and an Associate Editor for IEEE INDUSTRIAL ELECTRONICS MAGAZINE, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, and *Journal of the Franklin Institute*.

Chunxia Dou (Senior Member, IEEE) received the B.S. and M.S. degrees in automation from the Northeast Heavy Machinery Institute, Qiqihar, China, in 1989 and 1994, respectively, and the Ph.D. degree in control theory and control engineering from the Institute of Electrical Engineering, Yanshan University, Qinhuangdao, China, in 2005.

In 2010, she joined the Department of Engineering, Peking University, Beijing, China, where she was a Postdoctoral Fellow for two years. Since 2015, she has been a Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. Her research interests include multiagent-based control, event-triggered hybrid control, distributed coordinated control, multimode switching control and their applications in power systems, microgrids, and smart grids.

Xiangpeng Xie (Member, IEEE) received the B.S. and Ph.D. degrees in engineering from Northeastern University, Shenyang, China, in 2004 and 2010, respectively.

From 2010 to 2014, he was a Senior Engineer with Metallurgical Corporation of China Ltd, Beijing, China. He is currently a Professor with the Institute of Advanced Technology, Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include fuzzy modeling and control synthesis, state estimation, optimization in process industries, and intelligent optimization algorithms.

Dr. Xie is an Associate Editor for the *International Journal of Fuzzy Systems* and *International Journal of Control, Automation, and Systems*.

Gerhard P. Hancke (Life Fellow, IEEE) received the B.Sc., B.Eng., and M.Eng. degrees in electronic engineering from the University of Stellenbosch, Stellenbosch, South Africa, in 1970, 1970, and 1973, respectively, and the Ph.D. degree in electronic engineering from the University of Pretoria, Pretoria, South Africa, in 1983.

He is currently a Professor with the Nanjing University of Posts and Telecommunications, Nanjing, China, and also with the University of Pretoria, Pretoria, South Africa. He is recognized internationally as a Pioneer and the Leading Scholar of industrial wireless sensor networks research. He initiated and coedited the first Special Section on Industrial Wireless Sensor Networks of IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, in 2009, and IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, in 2013.

Dr. Hancke has been an Associate Editor and the Guest Editor for IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEEACCESS, and previously, the IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS. He is currently the Editor-in-Chief of IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS.