Inverter PQ Control With Trajectory Tracking Capability for Microgrids Based on Physics-Informed Reinforcement Learning

Buxin She, *Graduate Student Member, IEEE*, Fangxing Li, *Fellow, IEEE*, Hantao Cui, *Senior Member, IEEE*, Hang Shuai, *Member, IEEE*, Oroghene Oboreh-Snapps, *Graduate Student Member, IEEE*, Rui Bo, *Senior Member, IEEE*, Nattapat Praisuwanna, *Graduate Student Member, IEEE*, Jingxin Wang, *Member, IEEE*, and Leon M. Tolbert, *Fellow, IEEE*

***Abstract*—The increasing penetration of inverter-based resources (IBRs) calls for an advanced active and reactive power (PQ) control strategy in microgrids. To enhance the controllability and flexibility of the IBRs, this paper proposes an adaptive PQ control method with trajectory tracking capability, combining model-based analysis, physics-informed reinforcement learning (RL), and power hardware-in-the-loop (HIL) experiments. First, model-based analysis proves that there exists an adaptive proportional-integral controller with time-varying gains that can ensure any exponential PQ output trajectory of IBRs. These gains consist of a constant factor and an exponentially decaying factor, which are then obtained using a model-free deep RL approach known as the twin delayed deeper deterministic policy gradient. With the model-based derivation, the learning space of the RL agent is narrowed down from a function space to a real space, which reduces the training complexity significantly. Finally, the proposed method is verified through numerical simulation in MATLAB-Simulink and power HIL experiments in the CURENT center. With the physics-informed learning method, exponential response time constants can be freely assigned to IBRs, and they can follow any predefined trajectory without complicated gain tuning.**

***Index Terms*—Microgrids, inverter PQ control, inverter-based resources, physics-informed reinforcement learning, trajectory tracking, power hardware-in-the-loop experiment.**

NOMENCLATURE

Parameters

k_p	Proportional gain of PI controllers.
k_{p0}, k_{p1}	Constant coefficients of time-varying proportional gain.
k_i	Integral gain of PI controllers.
k_{i0}, k_{i1}	Constant coefficients of time-varying integral gain.
$i_{\mathrm{dref}}/i_{\mathrm{qref}}$	d-axis/q-axis reference current.
K_{pwm}	Pulse width modulation gain.
L_f	Filter inductance.
$P_{\mathrm{ref}}/Q_{\mathrm{ref}}$	Active/Reactive power reference.
$P_{\mathrm{trj}}/Q_{\mathrm{trj}}$	Active/Reactive power trajectory.
r	Agent reward.
$r_{\mathrm{P}}/r_{\mathrm{Q}}$	Active/Reactive power loop reward.
T_s	Sampling time delay.
u_{ref}	Input reference signal.
τ	Time constant of the exponentially decaying signal.
$\mathcal{E}$	Exploration noise.
σ	Variance of Gaussian distribution.
η	Soft-update weights.
γ	Punishment factor of reward function.

Variables

$\boldsymbol{a}$	Action vector.
$\boldsymbol{B}$	Replay buffer.
d	Soft-update frequency of critic-network.
D	Degree of a polynomial.
$e_{\mathrm{a}}, e_{\mathrm{b}}, e_{\mathrm{c}}$	Three-phase terminal voltage of inverter.
$i_{\mathrm{ga}}, i_{\mathrm{gb}}, i_{\mathrm{gc}}$	Three-phase current injection to grid network.
$P_{\mathrm{mes}}/Q_{\mathrm{mes}}$	Measured active/reactive power.
$\boldsymbol{Q}_{\theta c1}/\boldsymbol{Q}_{\theta c2}$	Dueling critic network parameterized by θ_{c1} and θ_{c2}.
$\boldsymbol{Q}_{\pi_\theta}$	Critic-network integrating policy π_θ.
Q-value	Performance index used in RL.
s	Laplace operator.
$\boldsymbol{S}$	State vector.
t	Time.
m	Epoch number during the training.
M	Maximum number of training epochs.
$u_{\mathrm{a}}, u_{\mathrm{b}}, u_{\mathrm{c}}$	Three-phase terminal voltage of filter.
w	Angel frequency.
y	Estimated Q-value by critic-network.
θ	Power angle.

Manuscript received 3 November 2022; revised 28 March 2023; accepted 5 May 2023. Date of publication 17 May 2023; date of current version 26 December 2023. This work was supported in part by the U.S. DOD ESTCP Program under Grant EW20-5331. Paper no. TSG-01649-2022. *(Corresponding author: Fangxing Li.)*

Buxin She, Fangxing Li, Hang Shuai, Nattapat Praisuwanna, Jingxin Wang, and Leon M. Tolbert are with the Department of EECS, The University of Tennessee, Knoxville, TN 37996 USA (e-mail: fli6@utk.edu).

Hantao Cui is with the School of ECE, Oklahoma State University, Stillwater, OK 74078 USA.

Oroghene Oboreh-Snapps and Rui Bo are with the Department of ECE, Missouri University of Science and Technology, Rolla, MO 65401 USA.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSG.2023.3277330.

Digital Object Identifier 10.1109/TSG.2023.3277330

$\boldsymbol{\theta}_a/\boldsymbol{\theta}_c$	Vector of actor-/critic-network weights and bias.
π_θ	Policy-network (actor-network).

Functions

$\mathbf{E}$	Excepted value function.
$E(s)$	Frequency-domain error function.
$n(s)$	Numerator transfer function of a general system.
$m(s)$	Denominator transfer function of a general system.
$G_{\text{Fixed}}(s)$	Transfer function of a fixed gain PI controller.
$G_{\text{PI}}(s)$	General transfer function of a PI controller.
$G_{\text{sys}}(s)$	Transfer function of a general system.
$G_{\text{Varying}}(s)$	Transfer function of an adaptive gain PI controller.
f	Action mapping function.
∇J	Deterministic policy gradient function.
$K_p(s)$	Frequency-domain proportional gain function.
$K_i(s)$	Frequency-domain integral gain function.
$Y(s)$	Frequency-domain output function.

I. Introduction

A MICROGRID is defined as an integrated energy system consisting of interconnected loads and distributed energy sources with a clear boundary [1], [2], which can operate in both grid-connected and islanded modes. Due to their capability to accommodate a variety of clean energy sources, microgrids play a significant role in environmental and energy strategies [3], including enhancing power system resiliency to withstand extreme weather [4], achieving zero carbon emissions [5], and improving national energy security [6].

One main difference between a microgrid and a conventional bulk power system is that a microgrid is composed of many inverter-based resources (IBRs) [7], which reshape the DC power generated by distributed energy resources (DERs), such as photovoltaic (PV) panels, wind turbines, battery energy storage systems (BESS), and so on [8]. The high penetration of IBRs makes microgrid control complicated. A typical hierarchical control structure for microgrids has three levels [9]: primary control, secondary control, and tertiary control. Each control level has specific tasks, and they coordinate to maintain microgrid stability and achieve economic benefits by controlling the output of each synchronous generator and IBR [10].

Whether a microgrid operates in grid-connected or islanded mode, active and reactive power (PQ) control is a basic control mode for IBRs [11]. The controllers at the secondary and tertiary levels generate PQ reference values and supplementary signals for the primary controllers [12]. In PQ control, the inverter is controlled as a current source [13] and the three-phase rotating voltage and current are converted to direct and quadrature DC variables through Park transformation. Then, these DC quantities can be regulated by proportional-integral (PI) controllers in the outer PQ regulation loop and the inner current regulation loop [14]. This double loop structure with PI controllers has been used extensively in industry and academia [15].

To enhance the flexibility and controllability of inverters so as to provide better ancillary services to microgrids, the existing literature developed several gain tuning methods for inverter PQ controllers, including the trial-and-error method, model-based method, heuristic method, and artificial intelligence (AI) based method [16]. The straightforward trial-and-error approach has typically been used in the field of industry. However, such case-by-case tuning was a time-consuming job for utility engineers. Hence, [17] implemented differential evolution metaheuristic algorithms to update PQ controller gains automatically. Reference [18] obtained the optimal fixed gains based on the controller bandwidth and the phase margin of the single-phase inverter-based system. Although [17], [18] found a proper fixed-gain, the PQ output of the inverter cannot be adjusted after different disturbances.

To make the IBRs more controllable, some adaptive strategies have been proposed to update PQ controller gains in real time. Reference [19] proposed a robust load frequency control strategy using a fuzzy logic based adaptive PI controller. In [20], a fuzzy-adaptive strategy was adopted to compensate for the dead time in the three-phase grid-connected inverter. Although the fuzzy logic controller has good performance in real-time gain scheduling, its membership function still needs an elaborate case-by-case design based on a system model. Then, [21] developed a novel control strategy for grid-connected PV systems based on adaptive controllers. The controller gains are continuously updated based on the gradient of tracking error. Reference [22] designed an adaption law for inverter control based on the Lyapunov function. In [23], an adaptive controller was designed for a three-phase constant voltage constant frequency inverter with an output filter, using adaptive gain scheduling control and feedback control. References [24], [25] discretized the control time window and scheduled gains according to real-time error. The above adaptive control theory-based methods have two disadvantages: 1) the shape of the inverter response cannot be freely designed and accurately controlled, which degrades the inverter's flexibility and controllability; 2) some adaptive parameters still require case-by-case detailed design, which is not only time-consuming but is vulnerable to parameter or model uncertainties.

Existing adaptive microgrid PQ controllers are not truly controllable because the PQ output of the inverter cannot accurately track the predefined trajectories, and thus cannot respond to the changing grid-side demand. Therefore, this paper proposes an adaptive microgrid PQ controller with trajectory tracking capability. To design such a PQ controller, the first question to answer is *whether there exists a PQ controller or not that can track a predefined exponential response trajectory?*. If it exists, then how can we find the controller gains without performing time-consuming gain tuning, and the controller itself can accommodate parameter or model uncertainties? The trajectory tracking concept was first proposed for voltage control [24], [26], and this paper extended it to PQ control using an upgraded inverter model.

Specifically, this paper develops an adaptive PQ controller with time-varying gains to track the predefined trajectory, using a hybrid model-based and physics-informed reinforcement learning (RL) method.

The integration of physics knowledge is an effective way to ensure efficient and safe learning of RL [1]. For example, [27] proposed a physical-aware, safe multi-agent RL method for the power management of distributed generators and energy storage systems in microgrids, where the gradient information is applied for constraint satisfaction during the training process. Reference [28] narrowed down the learning space and avoided baseline violations of network physical constraints. Hence, this paper first performs model-based analysis to prove the existence of an adaptive PI controller with time-varying gains to guarantee the predefined trajectory. The time-varying-gain is a function of time with a constant factor and an exponentially decaying factor. Then, a model-free deep RL algorithm known as the twin delayed deeper deterministic (TD3) policy gradient [29] is implemented to determine the time-varying PI gains, which is suitable for continuous control and proven to be more efficient than Deep Deterministic Policy Gradient (DDPG) algorithm by AI research [30] and engineering control research [31], [32]. With the guidelines provided by the model-based derivation, the RL agent just needs to find the real constant coefficients instead of the time-domain gain function. Hence, the learning space is narrowed down from a function space to a real space. This function reduction provides a new perspective on integrating physics knowledge into RL. It is more effective than value reduction and thus reduces the training complexity significantly. This hybrid method can be applied to other systems that employ PI controllers and wish to have a desired system response.

In addition to the numerical simulation, this paper further verifies the proposed controller through power hardware-in-the-loop (HIL) experiments. A hardware test-bed (HTB) platform has been developed by the Center for Ultrawide Area Resilient Electric Transmission Networks (CURENT) at the University of Tennessee to emulate power systems by programming IBRs to behave like power system components [33], [34]. Then, the contributions of this manuscript are summarized as follows:

- A mathematically rigorous proof of the existence of an adaptive PI controller that can track a predefined exponential response trajectory in a general feedback system.
- Derivation of the formulas for inverter PQ control to enable trajectory tracking capability and validation through both numerical simulation and power HIL experiments.
- Combination of the model-based analysis and the physics-informed RL approach to speed up the learning and solve the problem of model or parameter unavailability and uncertainty.

The remaining sections of this paper are arranged as follows: Section II derives the formula of an adaptive PI controller that can track an exponential trajectory in a generic system and then implements it in inverter PQ control. In Section III, a physics-informed deep RL implementation is proposed to learn the coefficients of the PI gains formula, since inverter models may not be readily available or accurate. Section IV verifies the proposed hybrid control algorithm in a modified Banshee microgrid through numerical simulation and power HIL experiments. Finally, Section V gives discussions and conclusions.

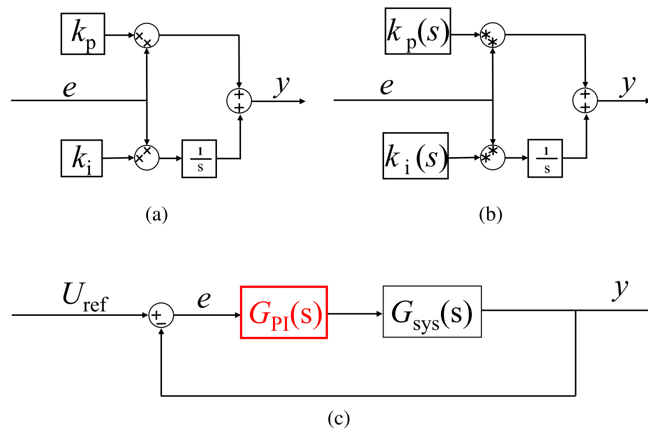


Fig. 1. (a) Diagram of the fixed-gain PI controller; (b) Diagram of the adaptive PI controller with time-varying gains; (c) PI controller in a feedback system.

II. Model-Based Theory and Analysis

This section first derives the adaptive PI controller with time-varying gains that can track an exponential response trajectory with a specific time constant in a general feedback system. Subsequently, the controller is applied to inverter PQ control.

A. Adaptive PI Controller

1) Transfer Function: The conventional PI controller uses fixed gains, and its transfer function is obtained in (1).

$$G_{\text{Fixed}}(s) = \frac{Y(s)}{E(s)} = k_p + \frac{k_i}{s} \tag{1}$$

As shown in Fig. 1(a)-(b), the diagram of the PI controller changes if time-varying gains are used. Because the multiplication operation in the time domain corresponds to the convolution operation in the frequency domain, the transfer function of the time-varying PI controller is derived as follows.

$$Y(s) = K_p(s) * E(s) + \frac{K_i(s) * E(s)}{s} \tag{2}$$

$$G_{\text{Varying}}(s) = \frac{1}{E(s)}\left[K_p(s) * E(s) + \frac{K_i(s) * E(s)}{s}\right] \tag{3}$$

where '*' is the convolution operator.

In fact, if $K_p(s)$ and $K_i(s)$ in (3) are constant, (3) is identical to (1), which means the time-varying gain PI controller is simplified to the conventional fixed-gain PI controller.

2) Adaptive PI Controller in a Feedback System: When a general PI controller is implemented in a feedback system, such as the one shown in Fig. 1(c), its transfer function is represented by the input error $E(s)$, system transfer function

$G_{sys}(s)$, and system output $Y(s)$.

$$G_{\mathrm{PI}}(s) = \frac{Y(s)}{E(s)G_{\mathrm{sys}}(s)} \tag{4}$$

Combining (3) and (4), one obtains (5). Then, it is possible to derive $k_p(t)$ and $k_i(t)$ in the time domain when $G_{sys}(s)$ and $Y(s)$ are known or predefined.

$$K_p(s) * E(s) + K_i(s) * \frac{E(s)}{s} = \frac{Y(s)}{G_{\mathrm{sys}}(s)} \tag{5}$$

B. Analytical Formulation of Adaptive Gains

1) Design of Ideal Smooth Trajectory: Assume the controller input U_{ref} is a step signal and the error e is a decaying exponential signal with a time constant τ in the control diagram shown in Fig. 1(b), then the output y is an ideal smooth trajectory. Their expressions in the time domain and frequency domain are shown in (6) and (7), respectively.

$$u_{ref} = \begin{cases} 1 & t \geq 0 \\ 0 & t < 0 \end{cases} \quad e(t) = e^{-t/\tau}, \quad y(t) = 1 - e^{-t/\tau} \tag{6}$$

$$U_{\mathrm{ref}} = \frac{1}{s}, \quad E(s) = \frac{1}{s + 1/\tau}, \quad Y(s) = \frac{1}{s} - \frac{1}{s + 1/\tau} \tag{7}$$

2) Derivation of Adaptive Gains: Assume $G_{sys}(s) = n(s)/m(s)$ and plug (6)-(7) in (5). Then,

$$K_p(s) * \frac{1}{s + 1/\tau} + K_i(s) * \frac{1}{s(s + 1/\tau)} = \frac{1}{\tau s(s + 1/\tau)} \cdot \frac{m(s)}{n(s)} \tag{8}$$

Next, perform an inverse Laplace transformation for both the left and right sides of (8). The left side is

$$\mathcal{L}^{-1}[\text{ left side }] = \tau k_i(t) + \big[k_p(t) - \tau k_i(t)\big]e^{-t/\tau} \tag{9}$$

The system transfer function $G_{sys}(s)$ is found on the right side and determines whether or not the left side = right side has a time domain solution. Let D represent the degree of a polynomial. $G_{sys}(s)$ can be categorized into three types based on the numerator and denominator degrees, which results in three different solutions.

Condition 1 : $D[n(s)] = 0$ and $D[m(s)] \leq 2$. The system transfer function does not have zero points and thus will not bring a new pole to the right side. Then,

$$\begin{aligned} \mathcal{L}^{-1}[\text{ right side }] &= \mathcal{L}^{-1}\left[\frac{1}{\tau s(s + 1/\tau)} \cdot \frac{m(s)}{n(s)}\right] \quad (10) \\ &= \mathcal{L}^{-1}\left[\frac{l_1}{s + 1/\tau} + \frac{l_2}{s}\right] \\ &= l_1 \cdot e^{-t/\tau} + l_2 \\ &= \mathcal{L}^{-1}[\text{ left side }] \\ k_p(t) = l_1 + l_2, &\quad k_i(t) = l_2/\tau \quad (11) \end{aligned}$$

where l_1 and l_2 are constants. In *Condition 1*, the adaptive PI controller changes to a conventional PI controller with fixed gains.

Condition 2 : $D[n(s)] \neq 0$ and $D[m(s)] - D[n(s)] \leq 2$.

$$\begin{aligned} \mathcal{L}^{-1}[\text{ right side }] &= \mathcal{L}^{-1}\left[\frac{1}{\tau s(s + 1/\tau)} \cdot \frac{m(s)}{n(s)}\right] \\ &= \mathcal{L}^{-1}\left[\frac{l_1}{s + 1/\tau} + \frac{l_2(s)}{s \cdot n(s)}\right] \\ &= l_1 \cdot e^{-t/\tau} + \mathcal{L}^{-1}\left[\frac{l_2(s)}{s \cdot n(s)}\right] \\ &= \mathcal{L}^{-1}[\text{left side}] \end{aligned} \tag{12}$$

$$\begin{cases} k_p(t) = l_1 + \mathcal{L}^{-1}\left[\frac{l_2(s)}{s \cdot n(s)}\right] \\ k_i(t) = \mathcal{L}^{-1}\left[\frac{l_2(s)}{s \cdot n(s)}\right]/\tau \end{cases} \tag{13}$$

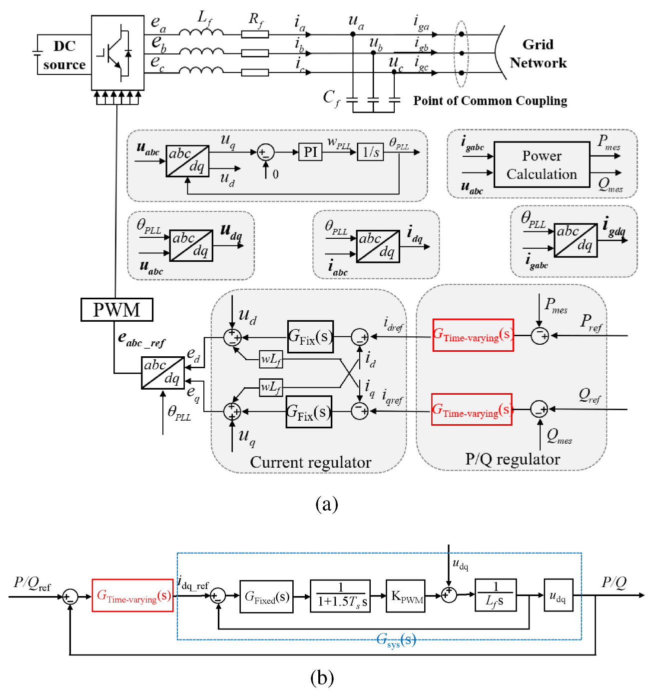


Fig. 2. (a) Control diagram of inverter-based PQ control; (b) Decoupled PQ control block-diagram.

where l_1 is constant and $l_2(s)$ is an s function obtained through fractional decomposition. In *Condition 2*, $k_p(t)$ and $k_i(t)$ are time-varying gains.

Condition 3 : $D[m(s)] - D[n(s)] \geq 3$. The right side is irreversible because the numerator has a higher degree than the denominator. In *Condition 3*, there is no solution for $k_p(t)$ and $k_i(t)$ in the time domain.

C. Inverter PQ Control With Trajectory Tracking Capability

Fig. 2(a) shows the complete diagram of this inverter-based PQ control. The decoupled control diagram is shown in Fig. 2(b) using the feedforward decoupling method. Here, the adaptive PI controller is only implemented in the PQ regulator because the bandwidth of the inner current regulator is wider than that of the power regulator, and the output of the power regulator determines the shape of the final PQ response. Furthermore, this paper assumes the phase-lock-loop (PLL) is tuned properly and has wider bandwidths than the PQ regulation loop [35], [36], [37]. Note, although the model-based derivation does not encompass the dynamics of the PLL, it is modeled in great detail in both the numerical simulator and HIL experiment. Implementing a data-driven approach in Section III can eliminate the modeling errors.

Based on Fig. 2, $G_{sys}(s)$ is written in (14).

$$G_{\text{sys}}(s) = \frac{K_{pwm}\left(k_{p2}s + k_{i2}\right)}{wL_f s^2(1+1.5T_s s) + K_{pwm}\left(k_{p2}s + k_{i2}\right)} \tag{14}$$

where k_{p2} and k_{i2} are fixed-gains of the current regulator in Fig. 2(a). Assume $G_{sys}(s) = n(s)/m(s)$, then $D[n(s)] = 1$, $D[m(s)] = 3$, and $D[m(s)] - D[n(s)] = 2$, satisfying *Condition 2*. Through fractional decomposition, the $k_p(t)$ and $k_i(t)$ of the power regulator are shown in (15).

$$\begin{cases} k_p(t) = k_{p0} + k_{p1}e^{-t/\tau'} \\ k_i(t) = k_{i0} + k_{i1}e^{-t/\tau'} \end{cases} \tag{15}$$

where

$$\begin{cases} k_{p0} = \frac{L_f(1-1.5T_s/\tau)}{\tau K_{pwm}(k_{i2}/k_{p2}-1/\tau)} \\ k_{p1} = \frac{L_f}{\tau K_{pwm}}\left(1.5T_s + \frac{1.5T_s/\tau-1}{k_{i2}/k_{p2}-1/\tau}\right) \\ k_{i0} = 0, k_{i1} = k_{p1}/\tau \\ \tau' = k_{p2}/k_{i2} \end{cases} \tag{16}$$

The adaptive gains that can help track a predefined PQ trajectory consist of a constant factor and an exponentially decaying factor. The four constant coefficients k_{p0}, k_{p1}, k_{i0}, and k_{i0} as well as the decaying time constant τ' are determined by PWM gain K_{pwm}, sampling delay T_s, filter reactance L_f, trajectory time constant τ, and fixed current regulator PI gains k_{p2} and k_{i2}.

D. Importance and Challenges of Analytical Formulation

The analytical formulation in the previous subsections illustrates that there exists an adaptive controller that can perfectly track a predefined trajectory following an exponential decay. The derivations, on the other hand, give a theoretical foundation for controller design, while the previous works [26] only determined how to track a given trajectory. Although this model-based mathematical proof is rigorous, it may not be suitable for direct implementation in a real-world system for the following reasons:

- It is difficult to model each component of the inverter in detail.
- The microgrid parameters are not always accessible; even if accessible, they are not necessarily accurate.
- Model-based suggestions also require further manual adjustments in the real application. The more simplified model needs more tuning effort.

With these challenges as motivations, a data-driven approach is proposed to implement the adaptive PQ control in the next section.

III. Physics-Informed Learning and Power HIL Demonstration

This section implements the adaptive PQ controller in a physics-informed data-driven way and demonstrates it through power HIL experiments.

A. Motivation for Deep Reinforcement Learning

To address the challenges discussed in the above Section II-D, a deep RL approach is implemented with the following considerations.

- RL is a goal-oriented machine learning algorithm outputting sequences of actions. It does not require a large number of labeled datasets like supervised learning.
- RL is adaptable because the uncertainties of the model and parameters are offset by the interactive training between the agent and environment.
- Since RL is an intelligent algorithm, it releases microgrid operators from time-consuming manual tuning.

Although the RL agent can directly replace the PI controller and output control signals, its training complexity will increase exponentially as the discretized control interval increases. Also, the breakdown of the closed-loop dynamics makes it difficult to guarantee security and stability [1]. To address this issue, this paper further narrows down the learning space based on the physical knowledge derived in Section II. Because the derived coefficient has a typical range based on engineering practice, this physics-informed implementation can enhance the safety of the data-driven method and avoid damaging the hardware device.

B. Physics-Informed Reinforcement Learning

1) Overview of the Proposed Framework: Fig. 3 shows the general framework of the proposed method, where the offline training is performed in numerical simulators (MATLAB-Simulink and Python) and training results are further validated through online power HIL experiments. Python receives the buffer data generated by Simulink to update the parameters of the actor- and critic-networks. In reverse, Simulink receives the actions in Python to regenerate the buffer data. This process repeats until the rewards converge. Then, the well-trained parameterized policy is implemented online in the CURENT HTB for further demonstration.

2) Physics-Informed TD3 Implementation: This subsection exemplifies the physics-informed design of a single inverter. Note that the proposed method is scalable because the inverter PQ control is local and independent of the knowledge of other generators.

(i). Control Agent: Each inverter holds two control agents for the active power (P) loop and the reactive power (Q) loop, each with an actor and a critic parameterized by fully connected neural networks (NNs). The actor outputs control gains, and the critic estimates the Q-value of the state-action pair. In TD3, actor and critic are further extended into twin NNs to prevent the overestimation of the Q-value.

(ii). State Set: The P loop state vector is defined as $\boldsymbol{S^P} = [\Delta P_{ref}, P_t, \tau, T_s]$, while that for Q loop agent is $\boldsymbol{S^Q} = [\Delta Q_{\text{ref}}, Q_t, \tau, T_s]$.

(iii). Action Set: The action vector is defined as $\boldsymbol{a} = [k_{\text{p0}}, k_{\text{p1}}, k_{\text{i0}}, k_{\text{i1}}, \tau']$, which means the agent outputs four coefficients and a time constant. Then, real-time k_p and k_i can be further calculated based on the model-based derivation in (15). The decaying exploration noise was added and implemented in the TD3 training. Without the model-based

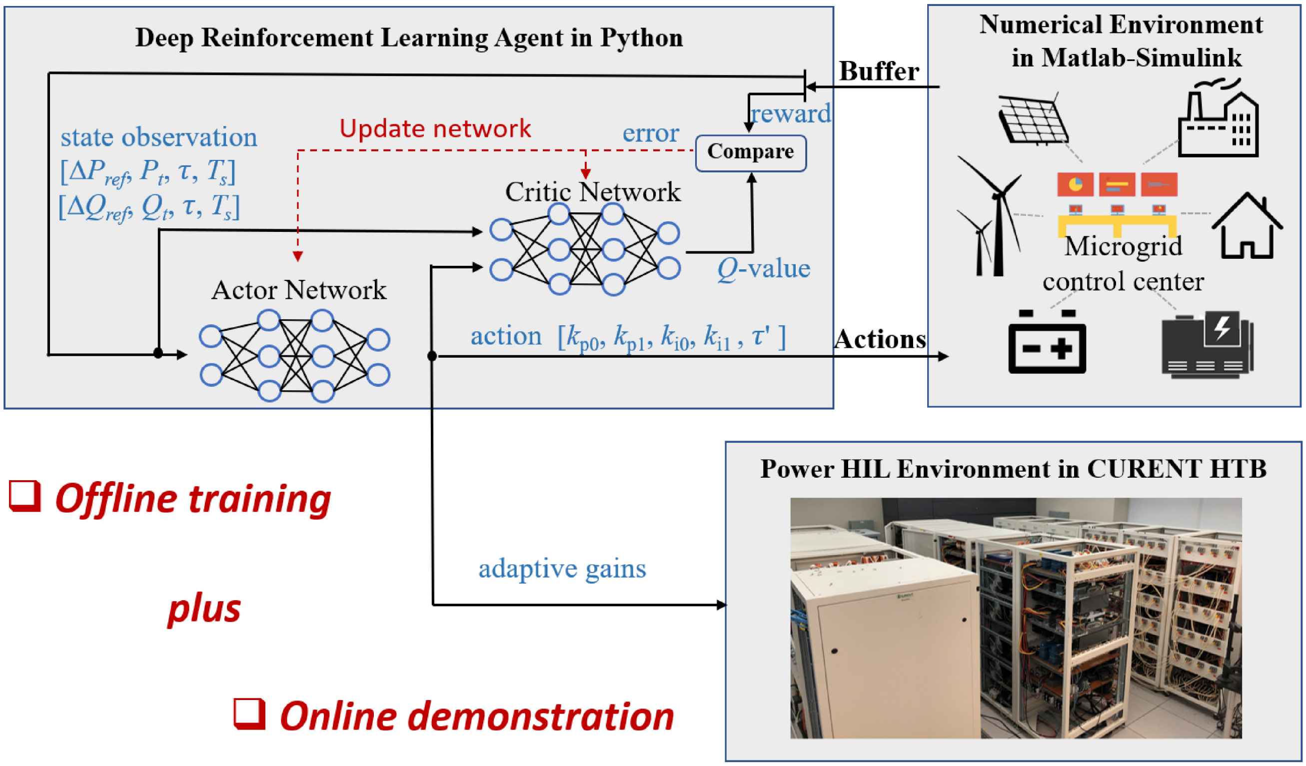


Fig. 3. Diagram of physics-informed learning in the numerical simulator and power HIL demonstration in HTB.

conclusion, the agent must output real-time gains following the simulation step size throughout the whole episode in Simulink, which may result in millions of distinct actions in each episode.

(iv). Control Policy: The control policy is denoted as π_θ, a deterministic function with parameter θ. It is actually the parameterized actor outputting the deterministic actions $\boldsymbol{a} = \pi_\theta(\boldsymbol{S})$.

(v). Design of reward function: The training reward is designed as an integral part of the error between the real-time PQ output and the designed trajectory. To better differentiate the features of the real-time trajectory, e.g., initial oscillation, overshooting, and steady-state errors, the reward function has a punishment factor γ. The final reward function for active power and reactive power regulation is shown in (17).

$$r = \frac{1}{2}(r_P + r_Q) \tag{17}$$

where

$$r_P = -\int \gamma(t) \cdot \left[\mathrm{P}_{trj}(t) - P(t)\right] dt \tag{18}$$

$$r_Q = -\int \gamma(t) \cdot \left[Q_{trj}(t) - Q(t)\right] dt \tag{19}$$

(vi). Policy update: After defining the required set, the agent keeps interacting with the environments and updating the policy through temporal-difference learning. The complete training process is detailed in *Algorithm 1*. Physics-informed TD3 implements three training techniques to prevent the overestimation of the Q-value in DDPG [38], [39] and use physics information to reduce training complexity.

- Twin critic networks: two critic-networks estimate the state-action value at the same time, and the smaller one is chosen as the estimated Q-value, as shown in line 12-17, *Algorithm 1*.
- Delayed update of target and policy: the updated frequency of the critic network is higher than that of the actor-network, as shown in line 13, *Algorithm 1*.
- Target policy smoothing: random noise is added to $\pi_{\theta a}(\boldsymbol{S})$ when the actor-network outputs the future actions, as shown in line 10, *Algorithm 1*.
- Physics-informed action mapping: The output of the actor-network is transformed to real-time gains based on (15), as shown in line 7, *Algorithm 1*. This physics integration also induced an additional chain-rule-based gradient ∇f_a, as illustrated in line 14, *Algorithm 1*.

C. Power HIL Demonstration

1) Power HIL Environment: As shown in Fig. 3, a power HIL experiment was conducted to further demonstrate the proposed control method. The HIL environment was emulated through CURENT HTB, which uses identical commercial-grade power electronics inverters to emulate real microgrids. Each inverter in the HTB is programmed digitally with built-in digital signal processors (DSPs) to behave as various devices, including sources, loads, energy storage, and solar PV [33], [34]. HTB has various interfaces with a real-time simulator, so the parameterized policy was programmed through RSCAD to control the hardware devices. The detailed configuration of a test microgrid was illustrated in Section IV-D.

TABLE I
KEY PARAMETERS OF TD3 TRAINING

Item	Parameters with model-based derivation	Parameters without model-based derivation
Punishment factor	$\gamma(t) = 20t + 1$	$\gamma(t) = 20t + 1$
Actor-network structure	$[4] \times [256] \times [256] \times [5]$	$[4] \times [256] \times [256] \times [2]$
Critic-network structure	$\begin{bmatrix} 4 \\ 5 \end{bmatrix} \times \begin{bmatrix} 32 \\ 32 \end{bmatrix} \times [256] \times [256] \times [1]$	$\begin{bmatrix} 4 \\ 2 \end{bmatrix} \times \begin{bmatrix} 32 \\ 32 \end{bmatrix} \times [256] \times [256] \times [1]$
Actor-network update frequency	20	20
Actor-network learning rate	0.0005	0.0005
Critic-network learning rate	0.001	0.001
Optimizer	Adam	Adam
Simulation step size in Simulink	5×10^{-5}	5×10^{-5}

Algorithm 1 Physics-Informed TD3 Training

Select $T, N, \boldsymbol{b}, \boldsymbol{\sigma}, \boldsymbol{\eta}, \alpha$
Initialize $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_c$; Initialize physics function f based on (15)
Initialize replay buffer $\boldsymbol{B}$
for $t \leftarrow$ to T **do**
  $\boldsymbol{S} \leftarrow \boldsymbol{S}'$, [Update state]
  $\boldsymbol{a} = \pi_\theta(S) + \varepsilon$, where $\varepsilon \sim N(0, \sigma)$ [Select action]
  $k_p, k_i \leftarrow f(\mathbf{a})$ [Physics-informed action mapping]
  $\boldsymbol{B} \leftarrow$ Append $(\boldsymbol{S}, \boldsymbol{a}, r, \boldsymbol{S}')$ [Store transition tuple]
  $\boldsymbol{B}_M \leftarrow \boldsymbol{B}'_M$ [Sample mini-batch tuples]
  $\boldsymbol{a}' = \pi_{\theta_a}(\boldsymbol{S}) + \boldsymbol{\varepsilon}'$, where $\mathcal{E}' = \text{clip}(\boldsymbol{\varepsilon}, -\boldsymbol{b}, \boldsymbol{b})$
  $y \leftarrow r + \alpha \min\left(\boldsymbol{Q}_{\theta c1}(\boldsymbol{S}', \tilde{\mathbf{a}}), \boldsymbol{Q}_{\theta c2}(\boldsymbol{S}', \tilde{\mathbf{a}})\right)$
  $\boldsymbol{\theta}_c \leftarrow \text{argmin}_{\theta_c} \mathbf{E} \sum \left[y - Q_\theta(S, a)\right]^2$[Update critics]
  if t mode d **then**
    $\nabla J(\theta) = E \nabla_a Q_{\pi\theta}(s, a)|_{a=\pi_\theta(s)} \nabla_\theta \pi_\theta(s) \nabla f_a$ [deterministic policy gradient]
    $\boldsymbol{\theta}_a \leftarrow \boldsymbol{\eta}\boldsymbol{\theta}_a + (1 - \boldsymbol{\eta})\boldsymbol{\theta}'_a$ [Soft update for target actor networks]
    $\boldsymbol{\theta}_c \leftarrow \boldsymbol{\eta}\boldsymbol{\theta}_c + (1 - \boldsymbol{\eta})\boldsymbol{\theta}'_c$ [Soft update for target critic networks]
  end if
end for
Output well-trained parameterized policy π_{θ_a}

2) Safe HIL Demonstration: Because the RL agent may output bad actions and damage the hardware devices during the exploration, the offline learning results are demonstrated in the power HIL environment after the reward curve converges in the numerical simulator. The converged reward curve means the policy is well-parameterized. Further, the HIL environment is equipped with some protection devices and built-in protection algorithms to guarantee safe data-driven implementation. Once a voltage or current exceeds the predefined threshold, switches will be turned off to isolate the endangered device or shut down the whole system.

IV. Case Study

A. Test System: Modified Banshee Microgrid

Fig. 4 shows the single-line diagram of the test microgrid to demonstrate the proposed adaptive PQ controller. The test microgrid is modified from the Banshee distribution system [40], [41] by keeping feeder 1 and adding renewable energy and energy storage devices. A 500-kW BESS on Bus 102 and a 2,500-kW PV device on Bus 105 diversify the power sources. The BESS supplements the diesel generator for power output. When combined with the BESS and diesel generator, the PV device can achieve all-day self-sustaining operation in grid-forming modes.

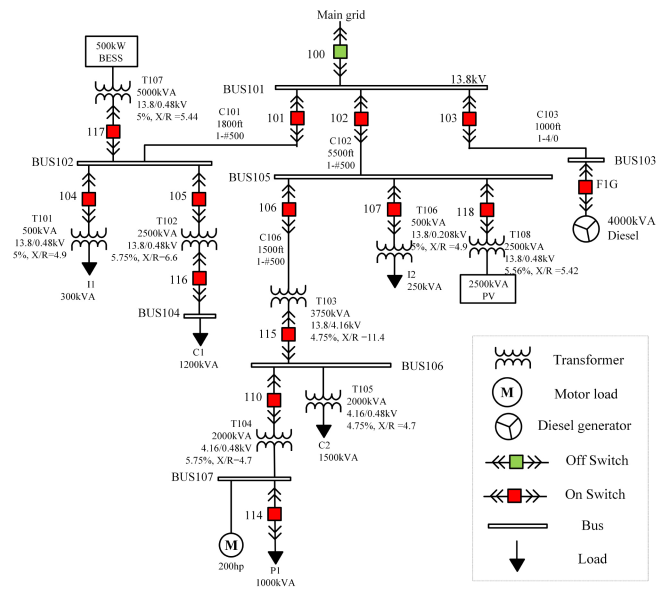


Fig. 4. Single-line diagram of modified Banshee microgrid [40].

B. Training in Numerical Simulators

1) Basic Settings: The proposed adaptive PI controller is implemented in the BESS connected to Bus 102. Through normalization, the active (P) and reactive (Q) loops can share one PI controller in the training process. After training, two PI controllers are then applied in the P and Q loops separately to enable asynchronous control of active and reactive power.

2) Distinct Settings With and Without the Physics Information: To show the advantages of the physics-informed learning approach, the RL agent is trained with and without model derivations. The key training parameters are listed in Table I. The main difference between the methods with and without physics information lies in the state and action sets, which further result in differences in the design of actor-network and critic-network, and the final search space.

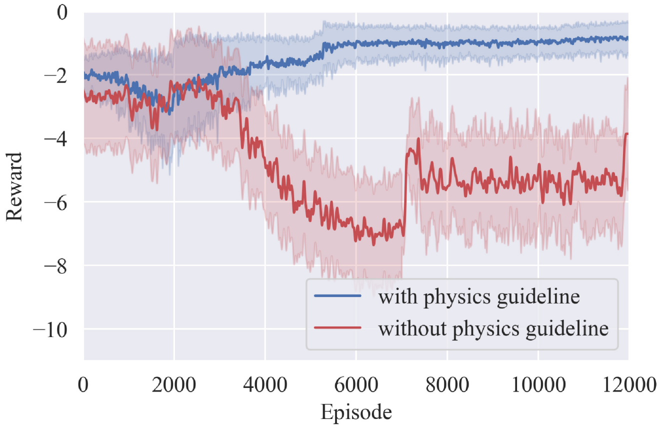


Fig. 5. Reward curve with and without physics guideline.

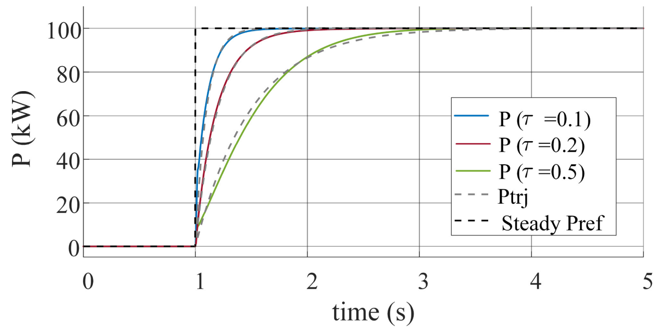


Fig. 6. Inverter response with the proposed method under scheduling single-loop single-step reference change.

With the physics information, the state vector is defined as $S_t^P = [\Delta P_{ref}, P_t, \tau, T_s]$ and the action vector as $\boldsymbol{a} = [k_{\mathrm{p}0}, k_{\mathrm{p}1}, k_{\mathrm{i}0}, k_{\mathrm{i}1}, \tau']$. Hence, the critic network has $[S_t^P, \boldsymbol{a}]$ (4+5 elements) as input and Q-value (1 element) as output, and the actor-network has S_t^P (4 elements) as input and $\boldsymbol{a}$ (5 elements) as output. The actor-network outputs four constant coefficients and a time constant that are independent of the time at the beginning of each episode, which is then transformed to real-time k_p and k_t based on the model-based derivation in (15).

Without the physics information, the RL agent shares the same state vector but has a different action vector as $\boldsymbol{a} = [k_{\mathrm{p}}, k_{\mathrm{i}}]$. Hence, the actor network has S_t^P (4 elements) as input and $\boldsymbol{a}$ (2 elements) as output. Note that $\boldsymbol{a}$ is dependent of time without knowing (15). Then, the actor-network must output real-time gains following the simulation step size throughout the whole episode in Simulink, which may result in millions of distinct actions in a single episode.

3) Training Results: All simulations have been performed in MATLAB version R2020a, Python version 3.7, and Tensor flow version 2.1 with a PC Intel Core i7-8665U CPU at 2.10 GHz and 16 GB RAM. Fig. 5 shows the TD3 training results. The training time with and without physics guidelines for a single episode is around 5.45 s and 10.67 s, respectively. In Fig. 5, the average reward curve based on model analysis converges after training for 6,000 episodes, while the reward curve without model analysis cannot converge. The physics guidelines provided by model-based derivation greatly reduce the training complexity and, therefore, facilitate convergence and reduce training time.

C. Validation in Numerical Simulators

This subsection verifies the model-based derivation and physics-informed TD3 training results in a numerical simulator. Because the modified Banshee microgrid is more vulnerable to disturbances in islanded mode (switch 100 is off in Fig. 4), this subsection mainly shows the islanded test results. Three scenarios, i.e., scheduling reference change, generation loss, and grounded faults, are demonstrated with distinct assigned time constants. The conventional fixed-gain PI controller and the adaptive controller in [22] work as the benchmark to show the superior performance of the physics-informed RL approach. Considering that one group of fixed gains can only track a specific predefined trajectory, in this paper these gains are manually updated when assigned a new trajectory time constant to improve the performance of fixed-gain controllers and make a fair comparison.

1) Scenario 1 (Scheduling Reference Change): In Scenario 1, the reference values of the P-loop and Q-loop are scheduled to change, which includes two sub-conditions as follows.

(i). *Scheduling single-loop single-step reference change:* P_{ref} changes from 0 to 100 kW at 1 s. Fig. 6 depicts the active power of the inverter under scheduling single-loop single-step changes when applying the proposed method. The gray and black dashed lines represent the predefined trajectories and steady-state reference, respectively.

In Fig. 6, the active power response almost coincides with the trajectories when the RL agent adaptively tunes the PI controllers. This verifies the effectiveness of the model-based derivation in Section II and the training results in Section IV-B. Fig. 7 further shows the trajectory tracking errors when applying different controllers. In Figs. 7(a)-(c), the physics-informed approach has minimum tracking error for any trajectory time constant. Particularly, the fixed-gain controller and the adaptive controller in [22] have obvious tracking errors with $\tau = 0.5$, as detailed in Fig. 7(c). This demonstrates the superiority of the proposed physics-informed data-driven approach.

(ii) *Scheduling double-loop cascaded-step reference change:* P_{ref} and Q_{ref} continuously change at 0 s, 2 s, and 4 s. To verify the robustness of the offline training, P-loop and Q-loop trajectories are assigned distinct time constants of 0.1 and 0.2, respectively. Fig. 8 shows the inverter response with the proposed method under the scheduling of a double-loop cascaded-step reference change.

In Fig. 8, the actual PQ response follows the predefined trajectories exactly, which means that the active power and reactive power can be controlled separately and simultaneously with the proposed method. As visualized in Fig. 9, the trajectory tracking errors are close to zero with the physics-informed TD3 controller, whereas for the fixed gain controller and the adaptive control in [22], there are noticeable trajectory errors, especially at the beginning of the reference change.

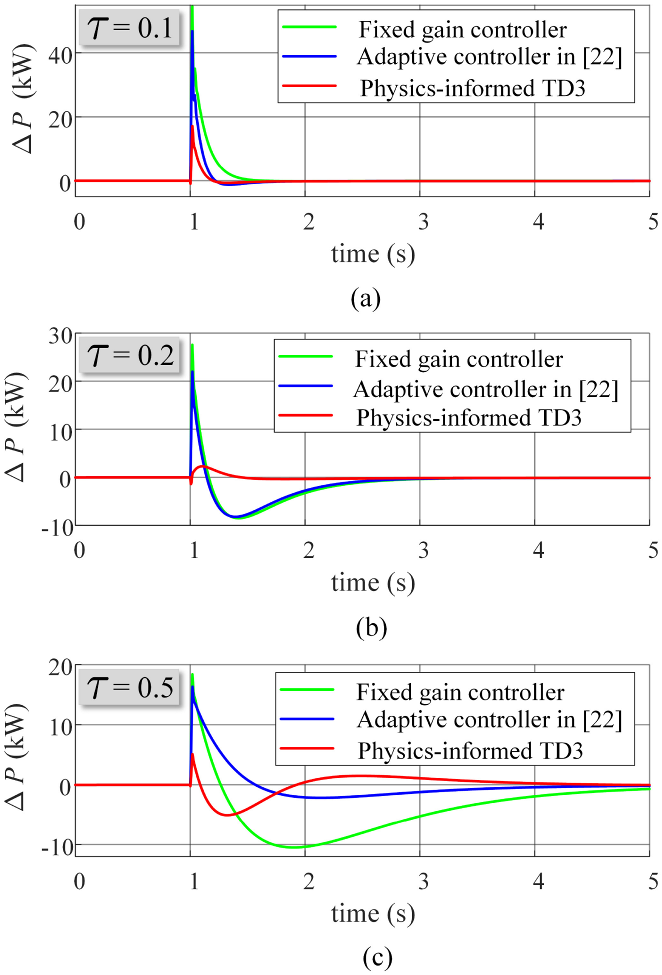


Fig. 7. Comparison of trajectory tracking errors with different controllers for single-loop single-step reference change: (a) $\tau = 0.1$; (b) $\tau = 0.2$; (c) $\tau = 0.5$.

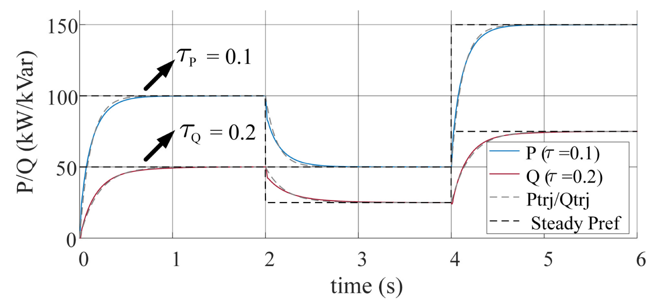


Fig. 8. Inverter response with the proposed method under scheduling double-loop cascaded-step reference change.

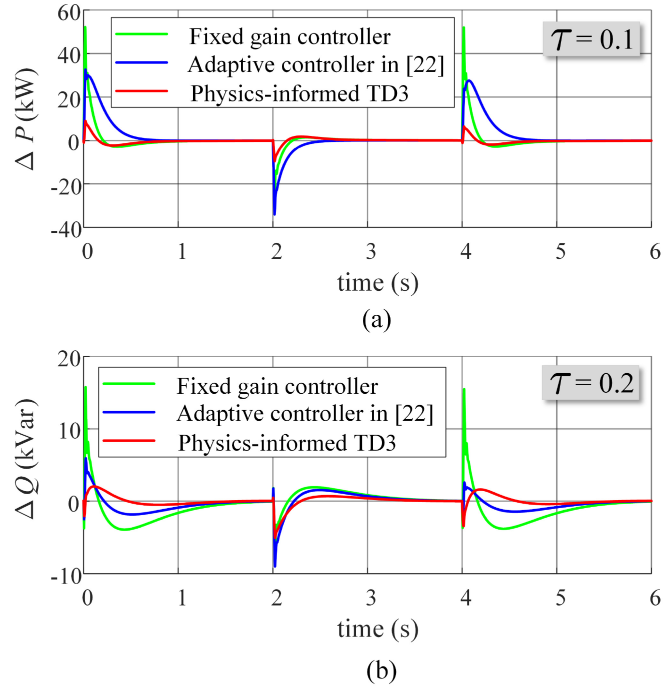


Fig. 9. Comparison of trajectory tracking errors with different controllers for double-loop cascaded-step reference change: (a) active power; (b) reactive power.

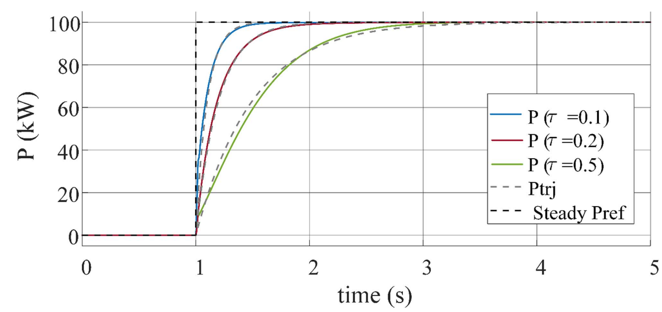


Fig. 10. Inverter response with the proposed method under generation loss.

2) Scenario 2 (Generation Loss): The uncertainty of renewable energy resources may result in the loss of generation from time to time. In Scenario 2, it is assumed that the PV panel loses 100 kW of generation at 1 s. The P_{ref} of the BESS increases from 0 to 100 kW to compensate for the generation loss. Similar to Scenario 1, three different time constants are assigned to the active power trajectory.

Fig. 10 shows that the inverter output can closely follow the predefined trajectories when using the proposed controller, as demonstrated by the active power response. In addition, Fig. 11 illustrates the trajectory tracking errors for three different controllers, where the fixed-gain controller and the adaptive method in [22] have large tracking errors. Despite continuously updating k_p and k_i online based on the pre-configured adaptation law in [22], the tracking errors are still larger than those of the proposed approach.

3) Scenario 3 (Grounded Fault): Assume a three-phase grounded fault occurs in transformer T106 at 5 s. Then, switch 107 is turned off to isolate the faults after 4 cycles (0.067 s). Instantly after the fault clearance, the BESS adjusts its output to follow the load change and thus mitigates the voltage and frequency deviations dynamically. Originally, $P_{ref} = 325$ kW and $Q_{ref} = 159$ kVar. P_{ref} and Q_{ref} are then changed to 100 kW and 50 kVar at 5.067 s to offset the 250 kVA shed load connected to T106.

Fig. 12 shows the inverter response, where the inverter output has obvious dips during the fault. With the proposed method, the active and reactive power can simultaneously

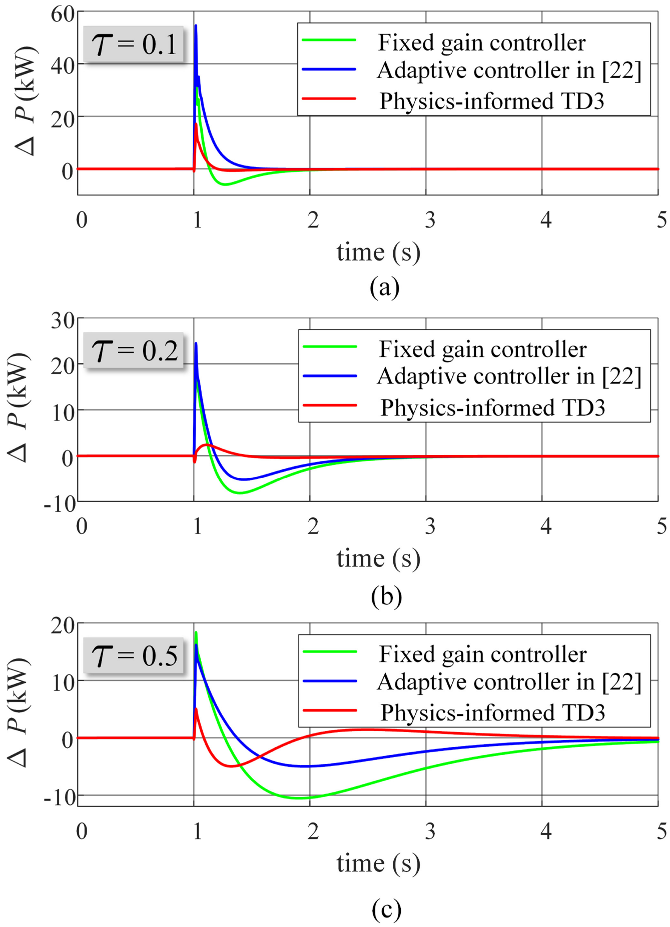


Fig. 11. Comparison of trajectory tracking errors with different controllers for generation loss: (a) $\tau = 0.1$; (b) $\tau = 0.2$; (c) $\tau = 0.5$.

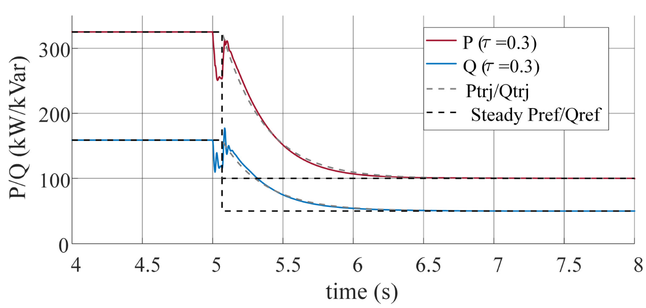


Fig. 12. Inverter response with the proposed method under grounded fault.

track the predefined trajectories right after the fault clearance. Fig. 13 further shows the trajectory tracking errors with different controllers. During the fault, all three controllers have evident tracking errors. Applying the proposed physics-informed approach quickly mitigates the tracking error instantly after the fault clearance. However, the fixed-gain controller has a considerable tracking delay (large negative tracking error), whereas the adaptive controller in [22] shows a significant overshoot (large positive tracking error) right after the fault clearance. These results confirm the superiority of the proposed physics-informed data-driven approach.

In general, the proposed PQ controller can accurately follow any predefined trajectory after a disturbance. It has smaller

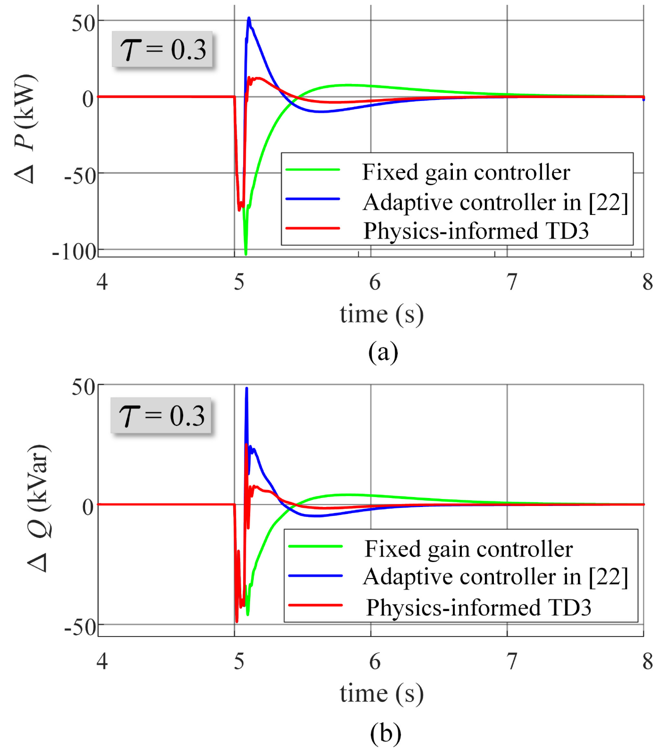


Fig. 13. Comparison of trajectory tracking errors with different controllers for grounded fault: (a) active power; (b) reactive power.

trajectory tracking errors and better adaptability than the existing methods. The numerical simulation demonstrates that the model-based derivation is valid, the TD3 agent is well-trained, and the exponentially decaying time constant can be freely assigned to the PQ response trajectory.

D. Validation in CURENT HTB

To further validate the proposed adaptive PQ controller, a power HIL experiment is performed in CURENT HTB [34] to validate the well-trained parameterized policy π_{θ_a}.

1) Configuration of a Modified Banshee Microgrid in CURENT HTB: There are generally six types of grid elements in the modified Banshee microgrid, i.e., the connection line, transformer, impedance-type load, motor load, inverter-based generation, and diesel generation. The impedance of the transformer is integrated into the connection line, which is emulated by real inductance. Except for the connection line and grid-following inverter, the rest of the elements are emulated by voltage-controlled current source inverters.

Four cabinets are involved in the configuration of the modified Banshee microgrid. The first cabinet contains the inductors and switches to fix the grid topology; the second cabinet contains several inverters to emulate the constant impedance loads; the third cabinet contains four inverters, two of which are to emulate synchronous generator and motor loads; and the last cabinet contains two inverters and the interface with RTDS. Fig. 14 shows how the HTB is controlled, where Fig. 14(a) is the diagram of the communication structure and 14(b) is the control panel when testing the adaptive PI controller. The video on the CURENT YouTube channel shows the detailed

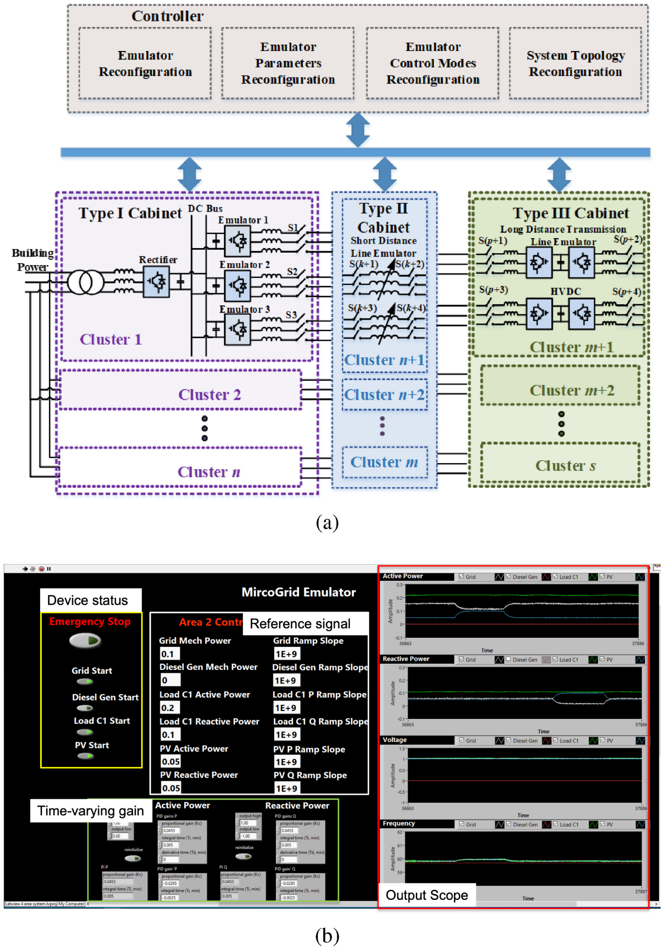


Fig. 14. Diagram of HTB: (a) communication structure; (b) control panel.

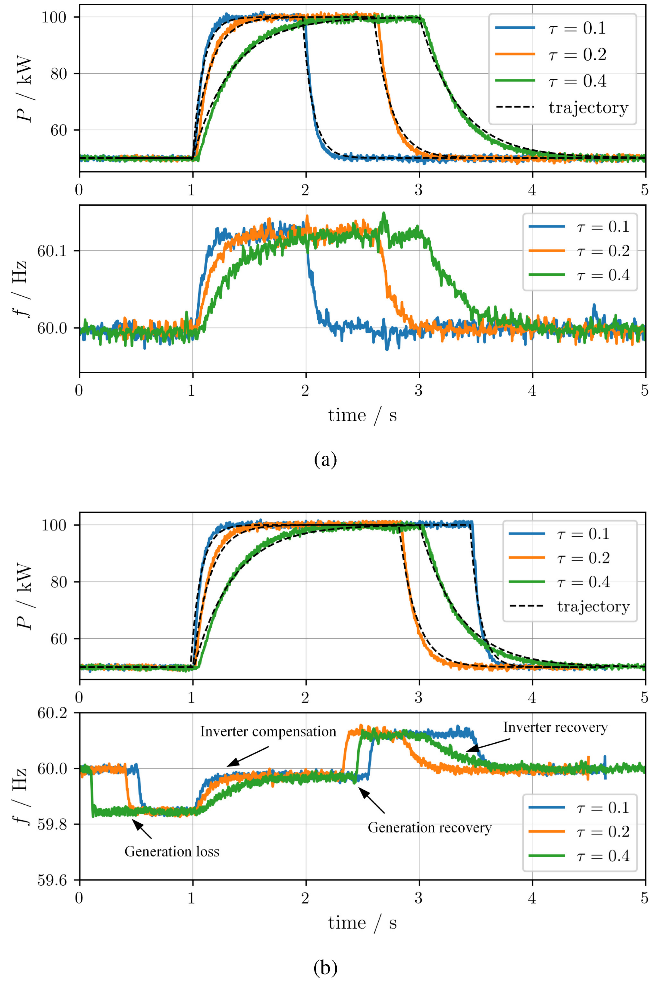


Fig. 15. Power HIL test results: (a) power reference change; (b) generation loss.

configuration [42]. A more detailed introduction to measurement, control, and communication architecture can be found in [33, Sec. III].

2) Controller Validation: Scheduling power reference change and generation loss are validated through the HTB. To test the generalization of the proposed method, a new time constant of $\tau = 0.4$ is assigned to the trained RL agent to output the adaptive gains for inverters in addition to the time constants of $\tau = 0.1$ and $\tau = 0.2$ used in numerical validation. Further, the initial P_{ref} is modified from 0 to 50 kW. The secondary controller was not implemented in HTB to better observe the impacts of assigned time constants on system dynamics. Hence, the steady state frequency may deviate from the nominal value after intentional power injection and generation loss.

The response of inverters employing the adaptive PQ controller to track the predefined trajectory is shown in Fig. 15. As shown in Fig. 15(a) and 15(b), the actual PQ response follows the predefined trajectories exactly, demonstrating the model-based analysis and the physics-informed learning in Sections II and III.

The frequency curves further indicate how the proposed method could help the grid work better. Specifically, faster power injection is beneficial to the microgrid when it needs support; while in normal conditions, the inverters are expected to have slower intentional power injection because the injected power breaks down the load-generation balance.

In the scenario of generation loss shown in Fig. 15(b), the blue curve RL agent was assigned a smaller time constant ($\tau = 0.1$), so it has faster active power injection after generation loss around 1 s. Then, the blue curve frequency recovers faster than the other two curves. In the scenario of the intentional reference change shown in Fig. 15(a), the power reference was changed from 50 kW to 100 kW. Then, the green curve frequency ($\tau = 0.4$) deviates slower from the nominal value, which brings milder dynamic disturbances. This shows the potential capability of the proposed method to support the main grids by customizing the trajectory time constant. The specific utilization of the proposed method for improved grid performance needs additional design, which is not covered in this paper.

V. Discussions and Conclusion

This paper proposes an adaptive microgrid PQ control method with guaranteed trajectory, combining model-based analytical proof, physics-informed learning, and power HIL experiments. The model-based analysis shows the existence of an adaptive controller that can perfectly track a predefined trajectory with exponential decay. This provides critical guidance to the RL implementation that is highly necessary, since direct controller substitution may bring about exponentially increased training complexity. The proposed PQ inverter control is essentially an optimized local control based on a PI controller. Using the PI controller as the basic implementation, an inverter-based PQ controller trained in a numerical simulator can ensure a good initial performance for plug-and-play use in microgrids. When it is applied in a new environment like a new distribution feeder or microgrid, incremental training using operational data can further optimize the performance to achieve the plug-and-play capability. In addition, improperly tuned PLL could have a negative impact on the performance of grid-following inverters [43], [44]. Further research is also necessary to help us understand the underlying mechanisms.

This paper tests the physics-informed training results through power HIL hardware experiments, which is beneficial for implementing the advanced model-free technique in real microgrids. The conclusions are summarized as follows.

1) The system transfer functions are categorized into three conditions, determining whether there exists a time-varying-gain adaptive PI controller that can track an exponentially traceable curve. In *Condition 1*, fixed-gains work; in *Condition 2*, time-varying gains are required; in *Condition 3*, no adaptive PI controller works.

2) The proposed controller outperforms the conventional fixed-gain and adaptive PI controllers. Without manual re-tuning, it can accurately track the predefined trajectory with any assigned time constant.

3) The microgrid inverter-based PQ control system meets Condition 2. After implementing the proposed adaptive PI controller, the active and reactive power output of inverters can track a predefined exponential trajectory. The trajectory time constant that benefits microgrid frequency and voltage could be customized in the application.

4) The model-based analysis provides guidelines for deep RL training, which relieves the training pressure and saves training time. In turn, the implementation of physics-informed deep RL solves the problem of unavailability and uncertainty in the model-based method.

5) The proposed control method allows inverter active and reactive outputs to follow a predefined exponential trajectory without the need for manual gain-tuning. The interaction between the RL agent and the training environment compensates for the uncertainties caused by model simplification, parameter distortion, and state variation.

In practice, higher-level controllers generate PQ references for inverter-level controllers. Coordination between higher- and lower-level controllers becomes feasible and significant owing to the controllability of the PQ output trajectory. In future work, specific coordination strategies will be developed and the impacts of the PLL on the proposed method will be investigated.

REFERENCES

[1] B. She, F. Li, H. Cui, J. Zhang, and R. Bo, “Fusion of microgrid control with model-free reinforcement learning: Review and vision,” *IEEE Trans. Smart Grid*, early access, Nov. 15, 2022, doi: 10.1109/TSG.2022.3222323.

[2] L. Huang, H. Xin, Z. Wang, L. Zhang, K. Wu, and J. Hu, “Transient stability analysis and control design of droop-controlled voltage source converters considering current limitation,” *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 578–591, Jan. 2019.

[3] N. Nasser and M. Fazeli, “Buffered-microgrid structure for future power networks; a seamless microgrid control,” *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 131–140, Jan. 2021.

[4] J. Alipoor, Y. Miura, and T. Ise, “Stability assessment and optimization methods for microgrid with multiple VSG units,” *IEEE Trans. Smart Grid*, vol. 9, no. 2, pp. 1462–1471, Mar. 2018.

[5] B. Alghamdi and C. A. Cañizares, “Frequency regulation in isolated microgrids through optimal droop gain and voltage control,” *IEEE Trans. Smart Grid*, vol. 12, no. 2, pp. 988–998, Mar. 2021.

[6] A. Navas-Fonseca et al., “Distributed predictive secondary control for imbalance sharing in AC microgrids,” *IEEE Trans. Smart Grid*, vol. 13, no. 1, pp. 20–37, Jan. 2022.

[7] D. Li, Q. Zhu, S. Lin, and X. Y. Bian, “A self-adaptive inertia and damping combination control of VSG to support frequency stability,” *IEEE Trans. Energy Convers.*, vol. 32, no. 1, pp. 397–398, Mar. 2017.

[8] A. Kumar, A. Mohapatra, S. N. Singh, and R. K. Panda, “Space vector rotation-based controlled decaying current injection for islanding detection of inverter-interfaced DG,” *IEEE Trans. Smart Grid*, vol. 13, no. 6, pp. 4638–4650, Nov. 2022.

[9] M. H. Andishgar, E. Gholipour, and R.-A. Hooshmand, “An overview of control approaches of inverter-based microgrids in islanding mode of operation,” *Renew. Sustain. Energy Rev.*, vol. 80, pp. 1043–1060, Dec. 2017.

[10] A. Bidram and A. Davoudi, “Hierarchical structure of microgrids control system,” *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1963–1976, Dec. 2012.

[11] F. Zhang and L. Mu, “A fault detection method of microgrids with grid-connected inverter interfaced distributed generators based on the PQ control strategy,” *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 4816–4826, Sep. 2019.

[12] J. Xu et al., “Carrier-based modulated model predictive control strategy for three-phase two-level VSIs,” *IEEE Trans. Energy Convers.*, vol. 36, no. 3, pp. 1673–1687, Sep. 2021.

[13] F. Chishti, S. Murshid, and B. Singh, “Robust normalized mixed-norm adaptive control scheme for PQ improvement at PCC of a remotely located wind–solar PV-BES microgrid,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 3, pp. 1708–1721, Mar. 2020.

[14] A. Eisapour-Moarref, M. Kalantar, and M. Esmaili, “Power sharing in hybrid microgrids using a harmonic-based multidimensional droop,” *IEEE Trans. Ind. Informat.*, vol. 16, no. 1, pp. 109–119, Jan. 2020.

[15] T. Morstyn, B. Hredzak, and V. G. Agelidis, “Control strategies for microgrids with distributed energy storage systems: An overview,” *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3652–3666, Jul. 2018.

[16] S. Li, Y. Sun, M. Ramezani, and Y. Xiao, “Artificial neural networks for volt/VAR control of DER inverters at the grid edge,” *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5564–5573, Sep. 2019.

[17] B. L. G. Costa, V. D. Bacon, S. A. O. da Silva, and B. A. Angélico, “Tuning of a PI-MR controller based on differential evolution metaheuristic applied to the current control loop of a shunt-APF,” *IEEE Trans. Ind. Electron.*, vol. 64, no. 6, pp. 4751–4761, Jun. 2017.

[18] J. Jiao, J. Y. Hung, and R. M. Nelms, “Gain scheduling control strategy for a single-phase grid-connected inverter,” in *Proc. IEEE 26th Int. Symp. Ind. Electron. (ISIE)*, 2017, pp. 723–728.

[19] A. Annamraju and S. Nandiraju, “Load frequency control of an autonomous microgrid using robust fuzzy PI controller,” in *Proc. 8th Int. Conf. Power Syst. (ICPS)*, 2019, pp. 1–6.

[20] Z. Zhang, J. Li, J. Chen, W. Wang, A. Xu, and P. Yang, “Research on dead-time compensation of inverter based on fuzzy adaptive PI control,” in *Proc. Chin. Autom. Congr. (CAC)*, 2019, pp. 5664–5668.

[21] A. Azad and H. Shateri, “A novel control strategy for on-grid photovoltaic systems based on adaptive PI control,” in *Proc. Iran. Conf. Renew. Energy Distrib. Gener. (ICREDG)*, 2019, pp. 1–6.

[22] S. S. Khorramabadi and A. Bakhshai, “Critic-based self-tuning PI structure for active and reactive power control of VSCs in microgrid systems,” *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 92–103, Jan. 2015.

[23] S. A. Q. Mohammed, M. S. Rafaq, H. H. Choi, and J.-W. Jung, "A robust adaptive PI voltage controller to eliminate impact of disturbances and distorted model parameters for 3-phase CVCF inverters," *IEEE Trans. Ind. Informat.*, vol. 16, no. 4, pp. 2168–2176, Apr. 2020.

[24] H. Li, F. Li, Y. Xu, D. T. Rizy, and S. Adhikari, "Autonomous and adaptive voltage control using multiple distributed energy resources," *IEEE Trans. Power Syst.*, vol. 28, no. 2, pp. 718–730, May 2013.

[25] Y. Xu, F. Li, Z. Jin, and M. H. Variani, "Dynamic gain-tuning control (DGTC) approach for AGC with effects of wind power," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3339–3348, Sep. 2016.

[26] H. Li, F. Li, Y. Xu, D. T. Rizy, and J. D. Kueck, "Adaptive voltage control with distributed energy resources: Algorithm, theoretical analysis, simulation, and field test verification," *IEEE Trans. Power Syst.*, vol. 25, no. 3, pp. 1638–1647, Aug. 2010.

[27] 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.

[28] C. Samende, J. Cao, and Z. Fan, "Multi-agent deep deterministic policy gradient algorithm for peer-to-peer energy trading considering distribution network constraints," *Appl. Energy*, vol. 317, Jul. 2022, Art. no. 119123.

[29] T. Joshi, S. Makker, H. Kodamana, and H. Kandath, "Application of twin delayed deep deterministic policy gradient learning for the control of transesterification process," 2021, *arXiv:2102.13012*.

[30] S. Dankwa and W. Zheng, "Twin-delayed DDPG: A deep reinforcement learning technique to model a continuous movement of an intelligent robot agent," in *Proc. 3rd Int. Conf. Vis., Image Signal Process.*, 2019, pp. 1–5.

[31] O. E. Egbomwan, S. Liu, and H. Chaoui, "Twin delayed deep deterministic policy gradient (TD3) based virtual inertia control for inverter-interfacing DGs in microgrids," *IEEE Syst. J.*, early access, Dec. 1, 2022, doi: 10.1109/JSYST.2022.3222262.

[32] J. Zhou, S. Xue, Y. Xue, Y. Liao, J. Liu, and W. Zhao, "A novel energy management strategy of hybrid electric vehicle via an improved TD3 deep reinforcement learning," *Energy*, vol. 224, Jun. 2021, Art. no. 120118.

[33] S. Zhang, B. Liu, S. Zheng, Y. Ma, F. Wang, and L. M. Tolbert, "Development of a converter-based transmission line emulator with three-phase short-circuit fault emulation capability," *IEEE Trans. Power Electron.*, vol. 33, no. 12, pp. 10215–10228, Dec. 2018.

[34] L. M. Tolbert et al., "Reconfigurable real-time power grid emulator for systems with high penetration of renewables," *IEEE Open Access J. Power Energy*, vol. 7, pp. 489–500, 2020.

[35] U. Markovic, O. Stanojev, P. Aristidou, E. Vrettos, D. Callaway, and G. Hug, "Understanding small-signal stability of low-inertia systems," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 3997–4017, Sep. 2021.

[36] O. Oboreh-Snapps, R. Bo, B. She, F. F. Li, and H. Cui, "Improving virtual synchronous generator control in microgrids using fuzzy logic control," in *Proc. IEEE/IAS Ind. Commercial Power Syst. Asia (I CPS Asia)*, 2022, pp. 433–438.

[37] Z. Zhang, R. Schuerhuber, L. Fickert, K. Friedl, G. Chen, and Y. Zhang, "Domain of attraction's estimation for grid connected converters with phase-locked loop," *IEEE Trans. Power Syst.*, vol. 37, no. 2, pp. 1351–1362, Mar. 2022.

[38] T. P. Lillicrap et al., "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

[39] X. Kou et al., "Model-based and data-driven HVAC control strategies for residential demand response," *IEEE Open Access J. Power Energy*, vol. 8, pp. 186–197, 2021.

[40] R. Salcedo et al., "Banshee distribution network benchmark and prototyping platform for hardware-in-the-loop integration of microgrid and device controllers," *J. Eng.*, vol. 2019, no. 8, pp. 5365–5373, 2019. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/10.1049/joe.2018.5174

[41] B. She et al., "Decentralized and coordinated V-f control for islanded microgrids considering DER inadequacy and demand control," *IEEE Trans. Energy Convers.*, early access, Mar. 21, 2023, doi: 10.1109/TEC.2023.3258919.

[42] B. She, F. Li, N. Praisuwanna, J. Wang, and L. M. Tolbert. "ESTCP project—Banshee microgrid configuration on CURENT HTB." May 2022. [Online]. Available: https://www.youtube.com/watch?v=jH6mqsDo0hs&ab_channel=CURENTerc

[43] L. Fan et al., "Real-world 20-Hz IBR subsynchronous oscillations: Signatures and mechanism analysis," *IEEE Trans. Energy Convers.*, vol. 37, no. 4, pp. 2863–2873, Dec. 2022.

[44] Y. Cheng et al., "Real-world subsynchronous oscillation events in power grids with high penetrations of inverter-based resources," *IEEE Trans. Power Syst.*, vol. 38, no. 1, pp. 316–330, Jan. 2023.

Buxin She (Graduate Student Member, IEEE) received the B.S.E.E. and M.S.E.E. degrees from Tianjin University, Tianjin, China, in 2017 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville. His research interests include microgrid operation and control, machine learning in power systems, distribution system plan, and power grid resilience. He was an Outstanding Reviewer of MPCE and IEEE OAJPE. He is a student Guest Editor of IET-RPG.

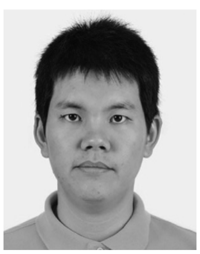
Fangxing Li (Fellow, IEEE) is also known as Fran Li. He received the B.S.E.E. and M.S.E.E. degrees from Southeast University, Nanjing, China, in 1994 and 1997, respectively, and the Ph.D. degree from Virginia Tech, Blacksburg, VA, USA, in 2001. He is currently the James W. McConnell Professor of Electrical Engineering with The University of Tennessee, Knoxville, TN, USA. He is also a Founding Member of CURENT, an NSF/DOE Engineering Research Center headquartered with UTK, and serves as the UTK Campus Director of CURENT. His research interests include resilience, artificial intelligence in power, demand response, distributed generation and microgrid, and electricity markets. From 2020 to 2021, he served as the Chair of IEEE PES Power System Operation, Planning, and Economics Committee. He has been serving as the Chair of IEEE WG on Machine Learning for Power Systems since 2019 and the Editor-in-Chief of IEEE OPEN ACCESS JOURNAL OF POWER AND ENERGY since 2020. He has received numerous awards and honors, including R&D 100 Award in 2020, the IEEE PES Technical Committee Prize Paper Award in 2019, five best or prize paper awards at international journals, and six best papers/posters at international conferences.

Hantao Cui (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Southeast University, China, in 2011 and 2013, respectively, and the Ph.D. degree in electrical engineering from The University of Tennessee, Knoxville in 2018. He is currently an Assistant Professor with the School of Electrical and Computer Engineering, Oklahoma State University. His research interests include power system modeling, simulation, and high-performance computing.

Hang Shuai (Member, IEEE) received the B.Eng. degree from the Wuhan Institute of Technology, Wuhan, China, in 2013, and the Ph.D. degree in electrical engineering from the Huazhong University of Science and Technology, Wuhan, in 2019. He was also a visiting student researcher with the University of Rhode Island (URI), Kingston, RI, USA, from 2018 to 2019, where he was a Postdoctoral Researcher from 2019 to 2020. He is currently a Research Assistant Professor of CURENT with the University of Tennessee Knoxville, Knoxville, TN, USA. His research interests include deep reinforcement learning, microgrid optimization, bulk power system resilience, and integrated energy system.

Oroghene Oboreh-Snapps (Graduate Student Member, IEEE) received the B.Eng. degree in electrical engineering from Madonna University, Nigeria, in 2017, the master's degree in electrical engineering from the Missouri University of Science and Technology, Rolla, USA, in 2019, where he is currently pursuing the Ph.D. degree. His research interest includes microgrid control, reinforcement learning, renewable energy and electric vehicle integration, and power system stability.

Rui Bo (Senior Member, IEEE) received the B.S.E.E. and M.S.E.E. degrees in electric power engineering from Southeast University, China, in 2000 and 2003, respectively, and the Ph.D. degree in electrical engineering from The University of Tennessee, Knoxville, in 2009. He is currently an Assistant Professor with the Electrical and Computer Engineering Department, Missouri University of Science and Technology (formerly the University of Missouri-Rolla). He worked as a Principal Engineer and a Project Manager with Midcontinent Independent System Operator from 2009 to 2017. His research interests include computation, optimization and economics in power system operation and planning, high performance computing, electricity market simulation, evaluation, and design.



Nattapat Praisuwanna (Graduate Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from the King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand, in 2009 and 2011, respectively. He is currently pursuing the Ph.D. degree in electrical engineering with The University of Tennessee. His research interests include power converters, renewable energy application, and renewable energy.

Jingxin Wang (Member, IEEE) received the Ph.D. degree in power electronics and electrical drives from Shanghai Jiao Tong University in 2011. He is a Research Associate with the Department of Electrical Engineering and Computer Science, The University of Tennessee, Knoxville. His research interests include high-performance motor control, modular multilevel converter, power flow control, and renewable energy.

Leon M. Tolbert (Fellow, IEEE) received the bachelor's, M.S., and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1989, 1991, and 1999, respectively. From 1991 to 1999, he was with Oak Ridge National Laboratory, Oak Ridge, TN, USA. In 1999, he was appointed as an Assistant Professor with the Department of Electrical and Computer Engineering, The University of Tennessee, Knoxville, TN, USA, where he is currently the Min H. Kao Professor with the Min H. Kao Department of Electrical Engineering and Computer Science. He is also a Founding Member of the NSF/DOE Engineering Research Center for Ultra-Wide-Area Resilient Electric Energy Transmission Networks. He is a Registered Professional Engineer with the State of Tennessee. He was a recipient of the 2001 IEEE Industry Applications Society Outstanding Young Member Award and Eight Prize Paper Awards from the IEEE Industry Applications Society and IEEE Power Electronics Society. He was an Associate Editor of the IEEE TRANSACTION ON POWER ELECTRONICS from 2007 to 2013. He was the Paper Review Chair for the Industry Power Converter Committee of the IEEE Industry Applications Society from 2014 to 2017.