Microgrid PQ Control with Guaranteed Trajectory: Model-Based Analysis, Physics-informed Learning and Hardware Experiment

Buxin She, *Student Member, IEEE,* Fangxing Li, *Fellow, IEEE,* Hantao Cui, *Senior Member, IEEE,* Hang Shuai, *Senior Member, IEEE,* Oroghene Oboreh-Snapps, *Student Member, IEEE,* Rui Bo, *Senior Member, IEEE,* Nattapat Praisuwanna, *Student Member, IEEE,* Jingxin Wang, *Member, IEEE,* 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 proposed an adaptive PQ control method with a guaranteed response trajectory, combining model-based analysis, physics-informed reinforcement learning, and power hardware-in-the-loop (HIL) experiment. First, model-based analysis proves that there exists an adaptive proportional-integral (PI) controller with time-varying gains that can ensure any exponential PQ output trajectory of IBRs. The gains consist of a constant factor and an exponentially decaying factor, which are then obtained using a model-free deep reinforcement learning (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*—microgrid PQ control, inverter-based resources, physics-informed reinforcement learning, trajectory tracking, power hardware-in-the-loop experiment**

I. Introduction

A microgrid is defined as an integrated energy system consisting of interconnected loads and distributed energy sources with a clear boundary [1], 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 [2], including enhancing power system resiliency to withstand extreme weather [3], achieving zero carbon emissions [4], and improving national energy security [5].

One main difference between a microgrid and a conventional bulk power system is that a microgrid is composed of many inverter-based resources (IBRs) [6], 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 [7]. The high penetration of IBRs makes microgrid control complicated. A typical hierarchical control structure for microgrids has three levels [8]: 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 IBRs [9]. Whether a microgrid operates in grid-connected or islanded mode, active and reactive power (PQ) control is a basic control mode for IBRs [10]. The controllers at the secondary and tertiary levels generate PQ reference values and supplementary signals for the primary controllers [11]. In PQ control, the inverter is controlled as a current source [12] 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 [13]. This double loop structure with PI controllers has been used extensively in industry and academia [14].

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 microgrid PQ controller, including the trial-and-error method, model-based method, heuristic method, and artificial intelligence (AI) based method [15]. 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, [16] implemented differential evolution metaheuristic algorithms to update PQ controller gains automatically. Ref. [17] obtained the optimal fixed gains based on the controller bandwidth and the phase margin of the single-phase inverter-based system. Although [16] - [17] found a proper fixed gain, the PQ output of inverter cannot be customized after different disturbances.

To make the inverter power more controllable, some adaptive strategies have been proposed to update PQ controller gains in real-time. Ref. [18] proposed a robust load frequency control strategy using a fuzzy logic based adaptive PI controller. In [19], 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 elaborate case by case design based on

B. She, F. Li, S. Hang, N. Praisuwanna, J. Wang, L.M. Tolbert are with the Dept. of EECS, The University of Tennessee, Knoxville, TN, USA.

H. Cui is with the School of ECE, Oklahoma State Univ., Stillwater, OK, USA.

O. Oboreh-Snapps and R. Bo are with Missouri Univ. of Science and Technology, Rolla, MO, USA.

a system model. Then, [20] 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. Ref. [21] designed an adaption law for inverter control based on the Lyapunov function. In [22], 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 [23] - [24] 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 customized grid-side demand. Therefore, this paper proposes an adaptive microgrid PQ controller with guaranteed trajectory. To design such a PQ controller, the first question to answer is '*whether there exists a PQ controller or not that can guarantee 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? For this, we developed an adaptive PQ controller with time-varying-gains to guarantee the output trajectory, using a hybrid model-based and physics-informed reinforcement learning approach.

First, we perform 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 reinforcement learning (RL) algorithm known as the twin delayed deeper deterministic (TD3) policy gradient [25] is implemented to determine the time-varying PI gains. 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, which reduces the training complexity significantly. In addition, the repeated interaction between the deep RL agent and the environment also solves the problems of parameter/model inaccuracy or unavailability. In addition to the numerical simulation, we further verified the proposed controller through power hardware-in-the-loop 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 [26] - [27]. This hybrid method may be applied to other systems that employ PI controllers and wish to have desired system response. Then, the contributions of this manuscript are as follows:

• A mathematically rigorous proof of the existence of an adaptive PI controller that can guarantee a predefined exponential response trajectory in a general feedback system.

• Derivation of the formulas for microgrid 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 reinforcement learning 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 guarantee an exponential response in a generic system and then implements it in an microgrid PQ control. In Section III, a physics-informed deep RL implementation is proposed to learn the coefficients of the PI gains formula, since microgrid 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 our conclusions.

II. Model-based Theory and Analysis

This section first derives the adaptive PI controller with time-varying gains that can guarantee an exponential response trajectory with a specific time constant in a general feedback system, and then applies it in microgrid PQ control.

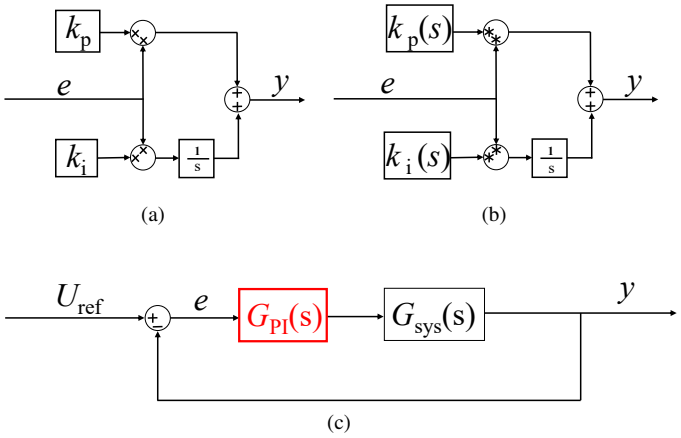


Fig. 1. (a) Diagram of fixed gain PI controller; (b) Diagram of adaptive PI controller with time-varying gains; (c) PI controller in a feedback system.

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 $Y(s)$.

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

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.

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

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.

$$u_{\text{ref}} = 1, \quad e(t) = e^{-t/\tau}, \quad y(t) = 1 - e^{-t/\tau} \tag{6}$$

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

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

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

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,

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

$$k_p(t) = l_1 + l_2, \quad k_i(t) = l_2/\tau \tag{11}$$

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 $D[m(s)] - D[n(s)] \leq 2$.

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

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

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. Microgrid PQ control with guaranteed trajectory

Fig. 2(a) shows the complete diagram of this inverter-based PQ control. The decoupled control diagram is displayed in Fig. 2(b) using the feedforward decoupling method [11]. Here, the adaptive PI controller is only implemented in the PQ regulator because the bandwidth of the inner current regulator is larger than that of the power regulator, and the output of the power regulator determines the shape of the final PQ response.

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

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

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).

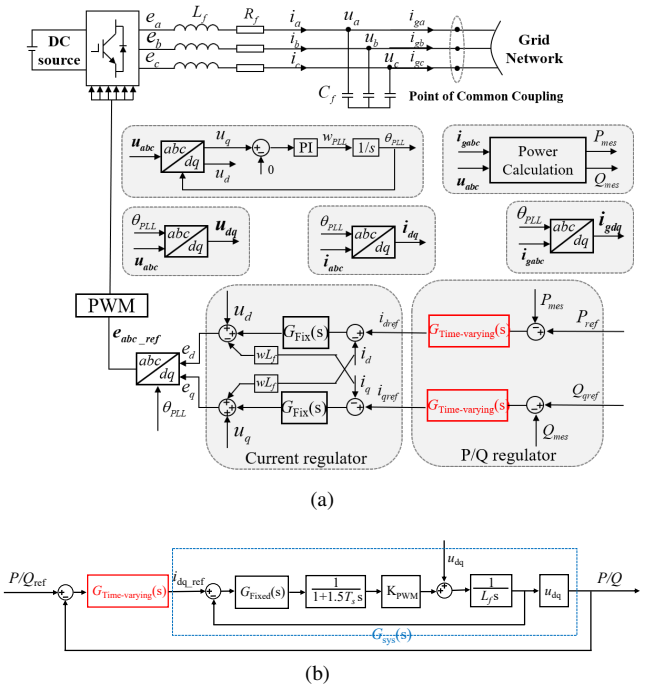


Fig. 2. (a) Control diagram of inverter-based PQ control ; (b) Decoupled PQ control block-diagram.

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

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 guarantee 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 [28] 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 our 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 data-driven way and demonstrates it through power HIL experiments.

A. Motivation for deep reinforcement learning

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

• RL is goal-oriented and can output sequences of actions. Also, 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 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 exponentialy as the decreztized control interval decreases. To address this issue, we further narrow down the learning space based on the physical knowledge derived in Section II.

B. Physics-informed reinforcement learning

1) Introduction to TD3: In this study, we chose a state-of-the-art deep deterministic RL method, the TD3 policy gradient. The TD3 policy gradient is an upgraded version of the deep deterministic policy gradient (DDPG) [29] that can handle continuous variables. Three techniques are developed to prevent the overestimation of the Q-value in DDPG as follows [30].

• 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.

• Delayed update of target and policy: the update frequency of the critic network is higher than that of the actor network.

• Target policy smoothing: random noise is added to a' when the critic network estimates $Q(s', a')$.

2) Physics-informed TD3 implementation: Fig. 3 shows the overall structure of the physics-informed TD3 algorithm, which has three main steps.

(i). Design of state, action, and actor/critic network: The derivation in Section II provides the physical guideline for RL agent learning. Three variables are chosen as the states, including trajectory time constant τ, PQ reference step change $\Delta P_{ref}/\Delta Q_{ref}$, and sampling interval T_s. The model-based derivation in (15)-(16) converts the learning space from function space to real space, resulting in four coefficients and the

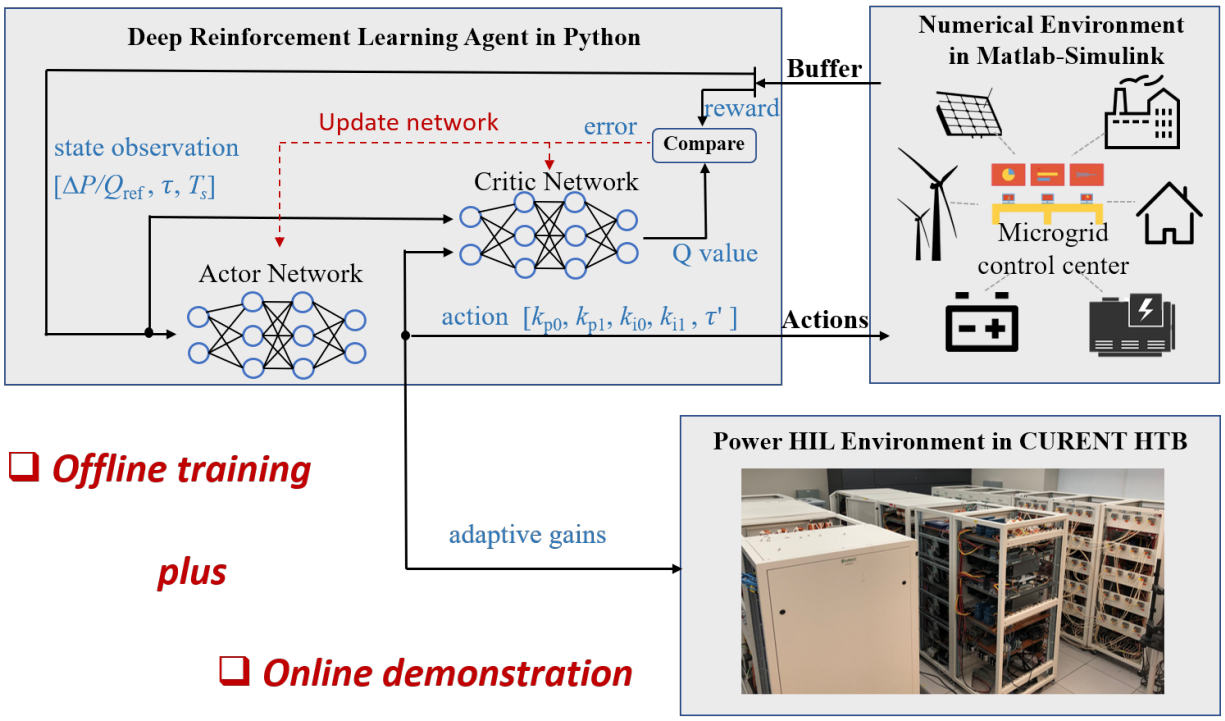


Fig. 3. Diagram of physics-guided learning in numerical simulator and power HIL demonstration in HTB.

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	$[3] \times [256] \times [256] \times [5]$	$[3] \times [256] \times [256] \times [2]$
Critic network structure	$\begin{bmatrix} 3 \\ 5 \end{bmatrix} \times \begin{bmatrix} 32 \\ 32 \end{bmatrix} \times [256] \times [256] \times [1]$	$\begin{bmatrix} 3 \\ 5 \end{bmatrix} \times \begin{bmatrix} 32 \\ 32 \end{bmatrix} \times [256] \times [256] \times [1]$
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}

decaying time constant being actions rather than real time control signals. In addition, actor and critic are fully connected neural networks.

(ii) 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 are shown in (17).

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

where

$$\begin{aligned} r_P &= -\int \gamma(t) \cdot [\mathrm{P}_{trj}(t) - P(t)]\, dt \\ r_Q &= -\int \gamma(t) \cdot [Q_{trj}(t) - Q(t)]\, dt \end{aligned} \tag{18}$$

(iii). Setup of the training environment An RL agent keeps interacting with the environment to update its policy based on environmental feedback. The environment can either be a real environment or a numerical simulator. This paper performs offline training in numerical simulators (Matlab-Simulink and Python) and validates the training results through online power HIL experiments. As shown in Fig. 3, Python receives the buffer data generated by Simulink to update 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. Knowing the format of gains in (15), the RL agent outputs four coefficients and a time constant at the beginning of each episode. Without the model-based conclusion, the deep RL agent must output real-time gains following the simulation step size throughout the whole episode in Simulink, which may result in millions of actions in a single episode. Hence, the physics-informed learning approach can greatly reduce the training complexity and save training time.

C. Power 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 numerical simulator. As shown in Fig. 3, power HIL environment is implemented in 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 [26] - [27].

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 [31] 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.

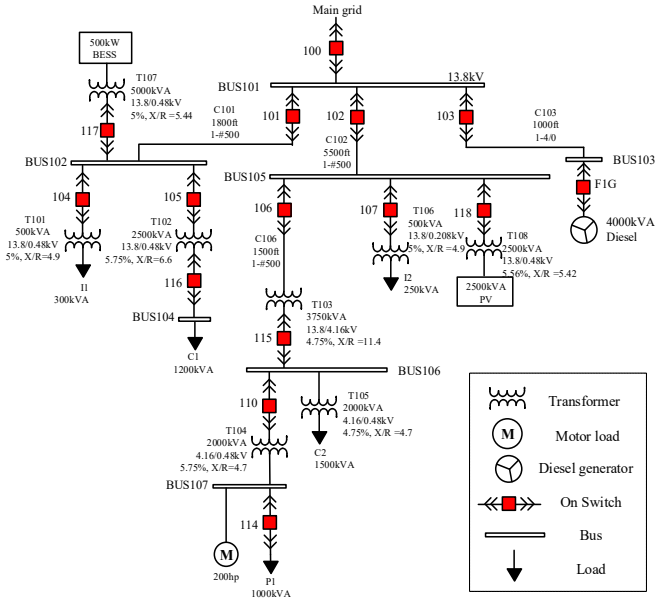


Fig. 4. Single-line diagram of modified Banshee microgrid [31].

B. Training in numerical simulators

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.

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. 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 guideline for a single episode is around 5.45s and 10.67s, 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 reduces training time.

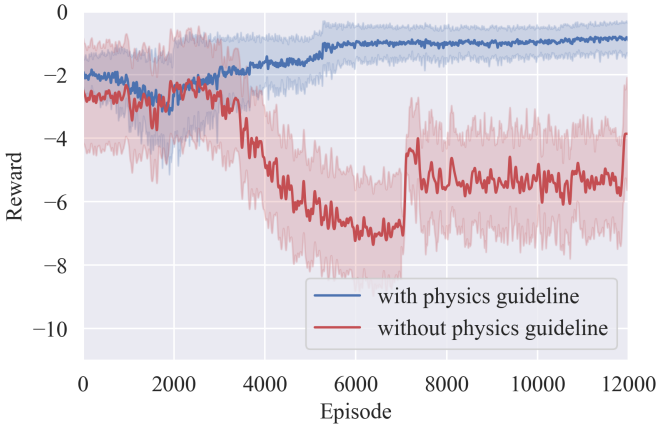


Fig. 5. Reward curve with and without physics guideline.

C. Validation in numerical simulators

This subsection aims to verify the model-based derivation and physics-informed TD3 training results in a numerical simulator. Because the modified Banshee microgrid is a typical weak grid [31] and more vulnerable to disturbances in islanded mode (switch 100 is off in Fig. 4), this subsection mainly shows the islanded test results.

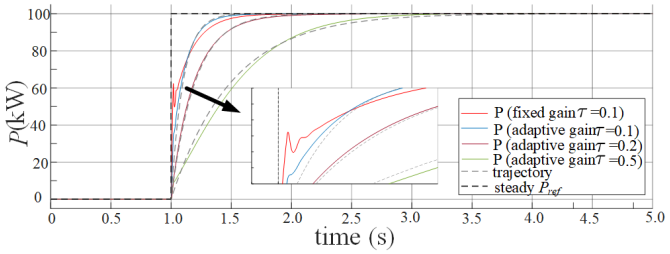
1) Scenario 1: Scheduling reference change: In Scenario 1, the reference values of the P loop and Q loop are scheduled to change, which includes two sub-conditions as follows. In each sub-condition, the trajectories are assigned different time constants.

(i). *Scheduling single-loop, single-step change:* P_{ref} changes from 0 to 100 kW at 1 s. Fig. 6 depicts the active power of IBRs under scheduling single-loop single-step changes, where the gray dashed lines represent the predefined trajectories. In Fig. 6(a), The active power response almost coincides with the trajectories when the PI controllers are adaptively tuned by the RL agent. This verifies the effectiveness of the model-based derivation in Section II and the training results in Subsection IV-B.

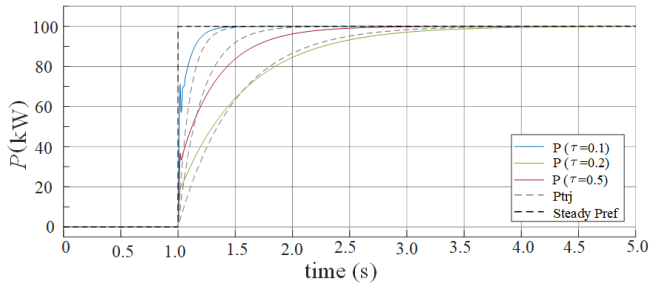
Moreover, the proposed adaptive PI controller is compared with the conventional fixed gain controller and the self-tuning method in [21]. In Fig. 6(a), the adaptive coefficients [k_{p0}, k_{p1}, k_{i0}, k_{i1}] set by the RL agent are [0.8, -0.5, 10.7, -5] when the trajectory time constant is 0.1. If the decaying coefficients k_{p1} and k_{i1} are removed, the time-varying gain controller changes to a conventional fixed gain controller with $k_p = 0.8$ and $k_i = 10.7$. As can be seen in Fig. 6(a), the adaptive gain curve (blue) has no minor oscillation at the beginning of the step change, and it follows the trajectory better than the fixed-gain

curve (red). In Fig. 6(b), although k_p and k_i are continuously updated online based on the pre-configured adaption law in [21], the active power response can hardly track the predefined trajectories.

(ii) *Scheduling double-loop cascaded-step 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 a time constant of 0.1 and 0.2, respectively. Fig. 7 shows the PQ response under the scheduling of a double-loop cascaded-step change. The actual PQ response follows the predefined trajectories exactly, which means that the active power and reactive power can be controlled separately and simultaneously.



(a)



(b)

Fig. 6. Microgrid response under scheduling single-loop single-step change reference change: (a) proposed adaptive controller; (b) adaptive controller in [21].

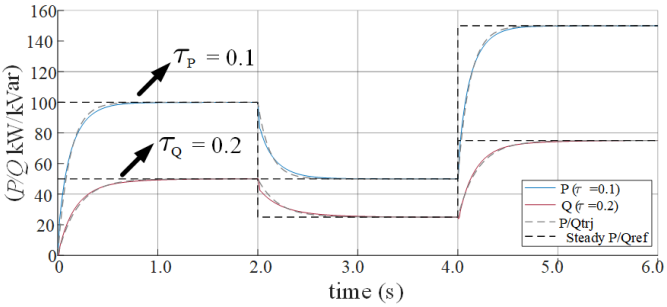
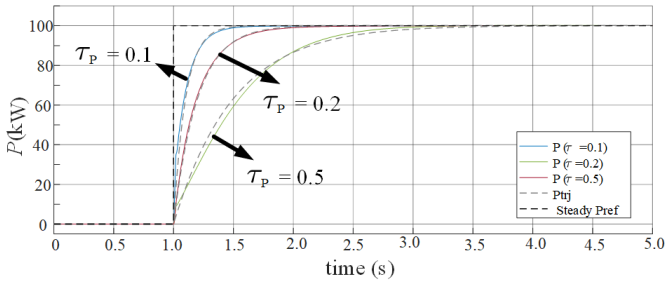
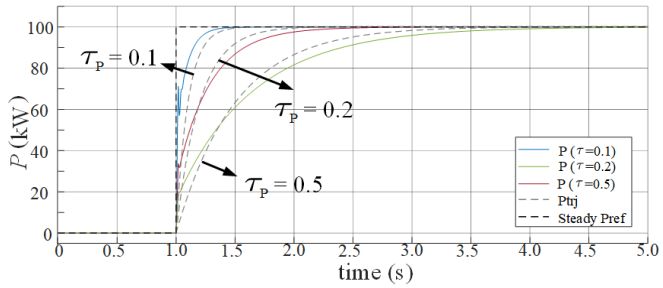


Fig. 7. Microgrid PQ response under scheduling double-loop cascaded-step change using proposed adaptive PI controller.

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. The active power response in Fig. 8(a) shows that BESS output can follow the predefined trajectories well when employing the proposed adaptive controllers. In addition, Fig. 8(b) shows the IBRs response using the adaptive method in [21], where the active power cannot track the dashed trajectories accurately.



(a)



(b)

Fig. 8. Microgrid response under generation loss: (a) proposed adaptive controller; (b) adaptive controller in [21].

In general, the proposed PQ controller can accurately follow any predefined trajectory after a disturbance. The numerical simulation demonstrates that the model-based derivation is valid, the TD3 agent is well-trained, and the exponential decaying time constant can be freely assigned to the PQ response trajectory.

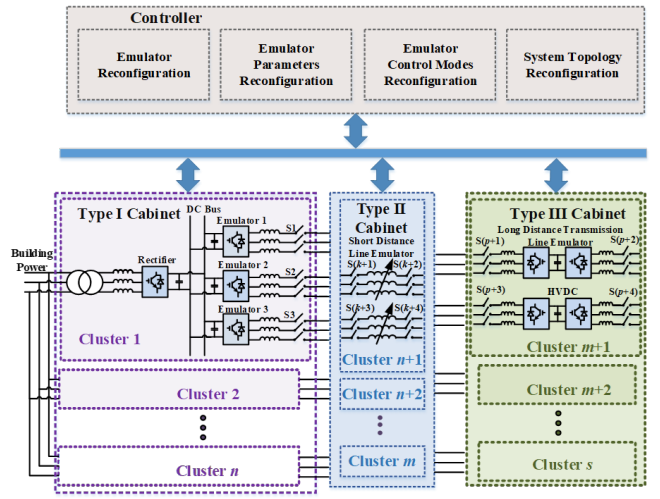
D. Validation in CURENT HTB

To further validate the proposed adaptive PQ controller, a power HIL experiment is performed in CURENT HTB [27].

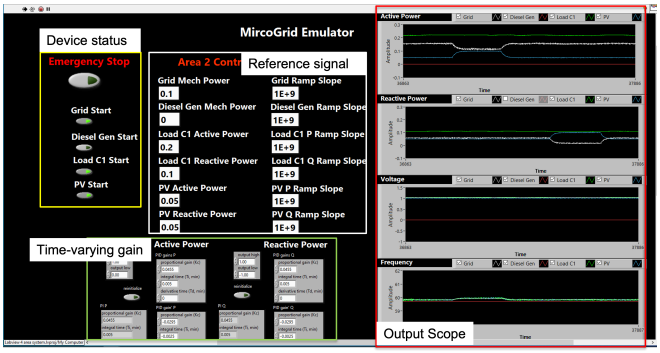
1) Configuration of HTB: Fig. 9 shows how HTB is controlled, where Fig. 9(a) is the diagram of communication structure and Fig. 9(b) is the control panel when testing the adaptive PI controller. For further introduction of setting up the modified Banshee microgrid in HTB, see the short video on Youtube [32].

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 IBRs 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 response of IBR employing the adaptive PQ controller with guaranteed trajectory is shown in Fig. 10. As shown in Fig.

10(a) and 10(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. Fig. 10 shows that the smaller time constants correspond to faster power response of IBRs and faster frequency response of microgrids, which could be customized in real applications.



(a)



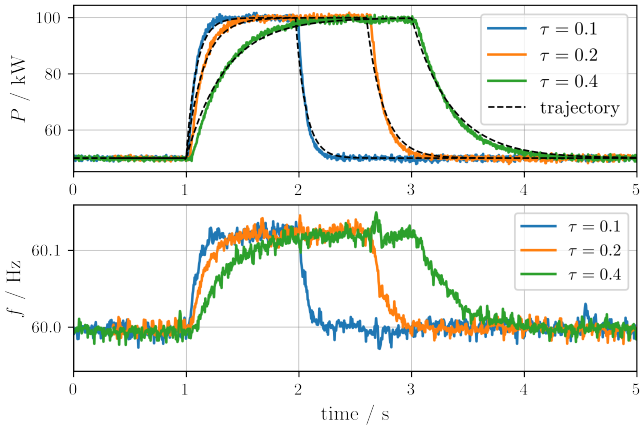
(b)

Fig. 9. Diagram of HTB: (a) communication structure (b) control panel.

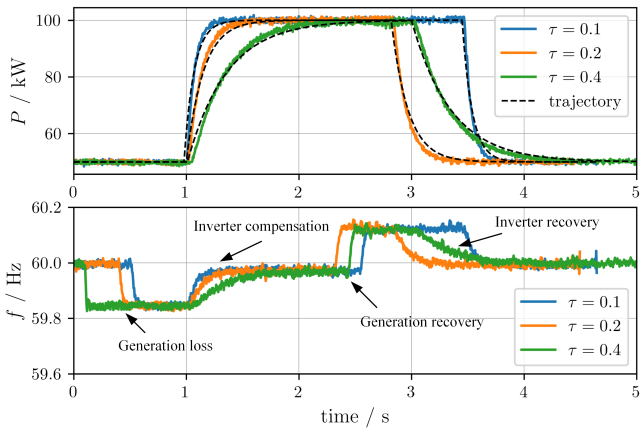
V. Conclusions

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 an 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. In addition, this paper tests RL training results through power HIL hardware experiments, which is beneficial for the implementation of 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 guarantee 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.



(a)



(b)

Fig. 10. Power HIL test results: (a) reference change; (b) generation loss.

2) The proposed controller outperforms the conventional fixed gain and adaptive PI controllers. 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 application.

4) The model-based analysis provides guidelines for the 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 are generating 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.

ACKNOWLEDGMENT

The authors would like to thank the financial support in part from the US DOD ESTCP program under the grant number EW20-5331 to complete this research work.

REFERENCES

[1] 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 Transactions on Smart Grid*, vol. 10, no. 1, pp. 578–591, 2017.

[2] N. Nasser and M. Fazeli, "Buffered-microgrid structure for future power networks; a seamless microgrid control," *IEEE Transactions on Smart Grid*, vol. 12, no. 1, pp. 131–140, 2020.

[3] J. Alipoor, Y. Miura, and T. Ise, "Stability assessment and optimization methods for microgrid with multiple vsg units," *IEEE Transactions on Smart Grid*, vol. 9, no. 2, pp. 1462–1471, 2016.

[4] B. Alghamdi and C. A. Cañizares, "Frequency regulation in isolated microgrids through optimal droop gain and voltage control," *IEEE Transactions on Smart Grid*, vol. 12, no. 2, pp. 988–998, 2020.

[5] A. Navas-Fonseca, C. Burgos-Mellado, J. S. Gómez, F. Donoso, L. Tarisciotti, D. Saez, R. Cardenas, and M. Sumner, "Distributed predictive secondary control for imbalance sharing in ac microgrids," *IEEE Transactions on Smart Grid*, vol. 13, no. 1, pp. 20–37, 2021.

[6] D. Li, Q. Zhu, S. Lin, and X. Bian, "A self-adaptive inertia and damping combination control of VSG to support frequency stability," *IEEE Transactions on Energy Conversion*, vol. 32, no. 1, pp. 397–398, 2016.

[7] 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 Transactions on Smart Grid*, 2022.

[8] M. H. Andishgar, E. Gholipour, and R.-a. Hooshmand, "An overview of control approaches of inverter-based microgrids in islanding mode of operation," *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1043–1060, 2017.

[9] B. She, F. Li, H. Cui, J. Wang, O. O. Snapps, and R. Bo, "Decentralized and coordinated vf control for islanded microgrids considering DER inadequacy and demand control," *arXiv preprint arXiv:2206.11407*, 2022.

[10] 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 Transactions on Smart Grid*, vol. 10, no. 5, pp. 4816–4826, 2018.

[11] J. Xu, T. B. Soeiro, F. Gao, L. Chen, H. Tang, P. Bauer, and T. Dragičević, "Carrier-based modulated model predictive control strategy for three-phase two-level VSIs," *IEEE Transactions on Energy Conversion*, vol. 36, no. 3, pp. 1673–1687, 2021.

[12] 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 Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1708–1721, 2019.

[13] A. Eisapour-Moarref, M. Kalantar, and M. Esmaili, "Power sharing in hybrid microgrids using a harmonic-based multidimensional droop," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 109–119, 2019.

[14] T. Morstyn, B. Hredzak, and V. G. Agelidis, "Control strategies for microgrids with distributed energy storage systems: An overview," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 3652–3666, 2016.

[15] S. Li, Y. Sun, M. Ramezani, and Y. Xiao, "Artificial neural networks for volt/var control of der inverters at the grid edge," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5564–5573, 2018.

[16] 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 transactions on industrial electronics*, vol. 64, no. 6, pp. 4751–4761, 2017.

[17] J. Jiao, J. Y. Hung, and R. M. Nelms, "Gain scheduling control strategy for a single-phase grid-connected inverter," in *2017 IEEE 26th International Symposium on Industrial Electronics (ISIE)*. IEEE, 2017, pp. 723–728.

[18] A. Annamraju and S. Nandiraju, "Load frequency control of an autonomous microgrid using robust fuzzy PI controller," in *2019 8th International Conference on Power Systems (ICPS)*. IEEE, 2019, pp. 1–6.

[19] 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 *2019 Chinese Automation Congress (CAC)*. IEEE, 2019, pp. 5664–5668.

[20] A. Azad and H. Shateri, "A novel control strategy for on-grid photovoltaic systems based on adaptive PI control," in *2019 Iranian Conference on Renewable Energy & Distributed Generation (ICREDG)*. IEEE, 2019, pp. 1–6.

[21] S. S. Khorramabadi and A. Bakhshai, "Critic-based self-tuning PI structure for active and reactive power control of VSCs in microgrid systems," *IEEE Transactions on smart grid*, vol. 6, no. 1, pp. 92–103, 2014.

[22] 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 Transactions on Industrial Informatics*, vol. 16, no. 4, pp. 2168–2176, 2019.

[23] H. Li, F. Li, Y. Xu, D. T. Rizy, and S. Adhikari, "Autonomous and adaptive voltage control using multiple distributed energy resources," *IEEE Transactions on Power Systems*, vol. 28, no. 2, pp. 718–730, 2012.

[24] Y. Xu, F. Li, Z. Jin, and M. H. Variani, "Dynamic gain-tuning control (DGTC) approach for AGC with effects of wind power," *IEEE Transactions on Power Systems*, vol. 31, no. 5, pp. 3339–3348, 2015.

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

[26] 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 Transactions on Power Electronics*, vol. 33, no. 12, pp. 10 215–10 228, 2018.

[27] L. M. Tolbert, F. Wang, K. Tomsovic, K. Sun, J. Wang, Y. Ma, and Y. Liu, "Reconfigurable real-time power grid emulator for systems with high penetration of renewables," *IEEE Open Access Journal of Power and Energy*, vol. 7, pp. 489–500, 2020.

[28] 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 Transactions on Power Systems*, vol. 25, no. 3, pp. 1638–1647, 2010.

[29] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," *arXiv preprint arXiv:1509.02971*, 2015.

[30] X. Kou, Y. Du, F. Li, H. Pulgar-Painemal, H. Zandi, J. Dong, and M. M. Olama, "Model-based and data-driven hvac control strategies for residential demand response," *IEEE Open Access Journal of Power and Energy*, vol. 8, pp. 186–197, 2021.

[31] R. Salcedo, E. Corbett, C. Smith, E. Limpaecher, R. Rekha, J. Nowocin, G. Lauss, E. Fonkwe, M. Almeida, P. Gartner *et al.*, "Banshee distribution network benchmark and prototyping platform for hardware-in-the-loop integration of microgrid and device controllers," *The Journal of Engineering*, vol. 2019, no. 8, pp. 5365–5373, 2019.

[32] B. She, F. Li, N. Praisuwanna, J. Wang, and L. M. Tolbert. (2022) ESTCP project - Banshee microgrid configuration on CURENT HTB (may 2022). [Online]. Available: https://www.youtube.com/watch?v=jH6mqsDo0hs