Data-driven Optimal Control Strategy for Virtual Synchronous Generator via Deep Reinforcement Learning Approach

Yushuai Li, *Member*, *IEEE*, Wei Gao, *Student Member*, *IEEE*, Weihang Yan, *Student Member*, *IEEE*, Shuo Huang, *Student Member*, *IEEE*, Rui Wang, Vahan Gevorgian, *Senior Member*, *IEEE*, and David Wenzhong Gao, *Fellow*, *IEEE*

***Abstract*—This paper aims at developing a data-driven optimal control strategy for virtual synchronous generator (VSG) in the scenario where no expert knowledge or requirement for system model is available. Firstly, the optimal and adaptive control problem for VSG is transformed into a reinforcement learning task. Specifically, the control variables, i.e., virtual inertia and damping factor, are defined as the actions. Meanwhile, the active power output, angular frequency and its derivative are considered as the observations. Moreover, the reward mechanism is designed based on three preset characteristic functions to quantify the control targets: ① maintaining the deviation of angular frequency within special limits; ② preserving well-damped oscillations for both the angular frequency and active power output; ③ obtaining slow frequency drop in the transient process. Next, to maximize the cumulative rewards, a decentralized deep policy gradient algorithm, which features model-free and faster convergence, is developed and employed to find the optimal control policy. With this effort, a data-driven adaptive VSG controller can be obtained. By using the proposed controller, the inverter-based distributed generator can adaptively adjust its control variables based on current observations to fulfill the expected targets in model-free fashion. Finally, simulation results validate the feasibility and effectiveness of the proposed approach.**

***Index Terms*—Adaptive control, virtual synchronous generator (VSG), reinforcement learning, deep learning.**

I. Introduction

THE increasing pressure from environment protection has made it urgent to conduct the research on accommodating high penetration level of renewable energy [1]-[4]. The renewable energy resources are converted to electricity which is then injected into the power system via power electronic inverters [5], [6]. Unlike the conventional synchronous generator (SG) with inherent rotating inertia, inverter-based distributed generator (IBDG) does not provide inertia support, which may make the system sensitive to network disturbances and even jeopardize system stability [7], [8]. To remedy the problem of system inertia, virtual synchronous generator (VSG), as a promising solution, has been proposed to control the grid-connected inverter to emulate the dynamic behavior of SGs [9], [10]. By designing the level of virtual inertia as well as damping, VSG can respond like the SG with slow frequency drop, which is beneficial for the frequency stability of power system [11], [12]. Therefore, the study of optimal control strategy of VSG becomes more significant to ensure high-quality power injection and maintain the safe operation of power system.

It is notable that the control operation of VSG is executed by software. As a result, the control parameters, i.e., virtual inertia and damping factor, can be set arbitrarily without physical limits. Up to now, a lot of control strategies for VSG have been presented to achieve the desired dynamic performance, which can be roughly classified into two categories, i.e., rule-based approach and optimization-based approach. The rule-based approach determines the control behavior by using the predefined operation rule. For instance, an adaptive-gain inertial control is proposed in [13], which focuses on improving the frequency nadir and guaranteeing the stable operation. By evaluating the change of rotor speed as well as the change of its differential, some adjustment strategies for a class of adaptive parameter(s) of VSG, i.e., inertia and/or damping factor, are proposed in [14]-[16]. Based on the preset operation table, the control parameters can be adaptively increased or decreased within a range of large and small parameters in different intervals, with final objective to achieve small over-shoot and short settling time. Based on small-signal modeling, a simple step-by-step parameter design strategy is presented in [17], which can take the double-line-frequency ripple into consideration. To achieve the tradeoff between active power and frequency regulations, a dual-adaptivity inertia control strategy is pro-

Manuscript received: May 17, 2020; accepted: October 26, 2020. Date of CrossCheck: October 26, 2020. Date of online publication: January 25, 2021.

This work was supported by the U.S. National Science Foundation (No. 1711951).

Y. Li, W. Gao, S. Huang, and D. W. Gao (corresponding author) are with the Department of Electrical and Computer Engineering, University of Denver, Denver, CO 80208, USA, and Y. Li is also with the School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, 110004, China (e-mail: yushuaili@ieee.org; wei.gao@du.edu; shuo.huang@du,edu; Wenzhong.Gao@du.edu).

R. Wang is with the School of Information Science and Engineering, Northeastern University, Shenyang, Liaoning, 110004, China (e-mail: 1610232@stu.neu.edu.cn).

W. Yan and V. Gevorgian are with the National Renewable Energy Laboratory, Golden, USA (e-mail: Weihang.Yan@nrel.gov; Vahan.Gevorgian@nrel.gov).

DOI: 10.35833/MPCE.2020.000267

posed in [18], which is based on a preset operation principle to get the range of adaptivity. Recently, [19] analyzes the transient stability of VSG and proposes a novel mode-adaptive power-angle control to enhance the transient stability effectively. By using this approach, the positive-feedback mode of power-angle control of the VSG can be adaptively switched to the negative-feedback one after large disturbances, which avoids the loss of synchronization. Although the rule-based approaches are easy for implementation, the predefined rules depend on expert knowledge such as how to choose large and small parameters in [14]-[16].

Recently, there is an increasing interest in investigating the parameter setting for VSG by using optimization-based approach, where the adjustment of parameters is driven by optimal solutions. For example, the stability of a microgrid with multi-VSGs is assessed based on the voltage angle deviations [20]. Therein, the particle swarm optimization is employed to tune the control parameters of each VSG in real time to achieve smooth transition after disturbances and limit the voltage angle deviations within a special range. The small-signal angular stability of a power system composed of the VSG subsystem and the other subsystems is investigated in [21], where the modal proximity-based approach is presented to guide the parameter design of the VSG. The concept of the linear-quadratic regulator-based control is proposed to find the optimal inertia constant for single VSG in [22], which is further extended to multi-VSGs in [23]. By using this approach, the trade-off between the critical frequency limits and control cost can be achieved. The aforementioned optimization-based approaches have made outstanding contributions to the design of control parameters for VSGs based on different requirements of power system stability. Nevertheless, these approaches are built upon small-signal modeling approach with linearization procedure and simplified mathematical model. Note that the system stability is affected by not only VSG but also other components, e.g., SG, line parameters, load conditions, etc. The interaction between the VSG and its working environment (the whole system) is ignored in the existing research [13]-[18], [20]-[23]. To address this issue, one way is to establish the dynamics of the whole system, analyze the interaction between the VSG and the power system, and then design the corresponding control strategy for VSG. However, it is a very difficult task to establish an exact model of the whole power system under complex interconnected structure. Although such system model can be built in some special cases, it is featured with high-order, nonlinear and strong coupling in general. As a result, it is also difficult for engineers to analyze the impact of VSG on the power system stability and design the corresponding optimal control strategy with a variety of uncertain system disturbances. In addition, different systems may have different structures and components. This also means that each VSG may work in different environments. As a result, the control strategy for VSG by using exact system model may not be universal. Based on those above-mentioned discussions, it is an open problem and challenge in the field of adaptive control for VSG to design a universally optimal control strategy for VSG, which can only use observed data without building the whole system model, i.e., the model-free fashion.

Thanks to the rapid evolvement of artificial intelligence technology, the reinforcement learning approaches enable to find the optimal control policy by only using data interaction between agent and unknown environment, which can be considered as a promising approach to deal with the aforementioned challenge. Up to now, a lot of reinforcement learning algorithms have been proposed [24]-[29]. Among them, the most popular approaches include the deep Q-network (DQN) algorithm [26] and deep policy gradient (DPG) algorithm [28], which are obtained by successfully combining the deep neural networks (DNN) with the typical Q-learning algorithm [24] and the policy gradient algorithm [25], respectively. Based on different application scenarios, DQN and DPG are suitable for solving different reinforcement learning tasks. Specifically, DQN can better handle the case with continuous observation spaces and discrete action spaces, while DPG fits well with both continuous observation and continuous action spaces. In this paper, we consider continuous state observations, e.g., the variations of active power output as well as angular frequency, to achieve continuous control operations for VSG. Thus, the concept of DPG is more suitable for our work.

Mainly with the aforementioned inspirations, the paper investigates the optimal and adaptive control problem for VSG in model-free scenario, where a decentralized deep policy gradient (DDPG) algorithm is developed and employed to solve this problem. The DDPG is obtained by using the decentralized stochastic gradient descent approach [30] to replace the stochastic gradient descent approach in classical DPG algorithm for improving the convergence speed. The major contributions of this paper are summarized as follows.

1) The optimal and adaptive control problem for VSG is formulated and transformed into a reinforcement learning task. Therein, the expected performance to achieve multiple control targets for angular frequency and active power regulations are simultaneously considered in the designed optimization target.

2) A data-driven optimal control policy is designed and embedded into the VSG controller based on the DDPG algorithm. It enables the IBDG to adaptively respond to system disturbances and obtain expected performance with the maximum long-term return in model-free fashion.

The remainder of this paper is organized as follows. Section II introduces VSG control, identifies its control variables as well as observation variables, and presents the unknown system dynamics. In Section III, multiple characteristic functions are defined to formulate the expected control targets. Subsequently, the optimal control problem is transformed into a reinforcement leaning task, which is further solved by introducing the DDPG algorithm. Several case studies are provided to verify the effectiveness of the proposed approach in Section IV. Finally, Section V concludes this paper.

II. System Model and Problem Formulation

A simplified diagram of power system is shown in the upper-right corner of Fig. 1, where IBDG as well as other components are integrated into the system. There are multiple configurations for power systems. Meanwhile, the IBDG does not know the system structure as well as the system model. The control diagram of the IBDG is shown in the upper-left corner of Fig. 1. Therein, the concept of VSG control is embedded into the active power control loop to improve the angular frequency stability. Meanwhile, the terminal voltage of IBDG is directly controlled through a proportional-integral (PI) controller to maintain the terminal voltage at the nominal value [31], [32]. The variables in Fig. 1 will be defined in the following text.

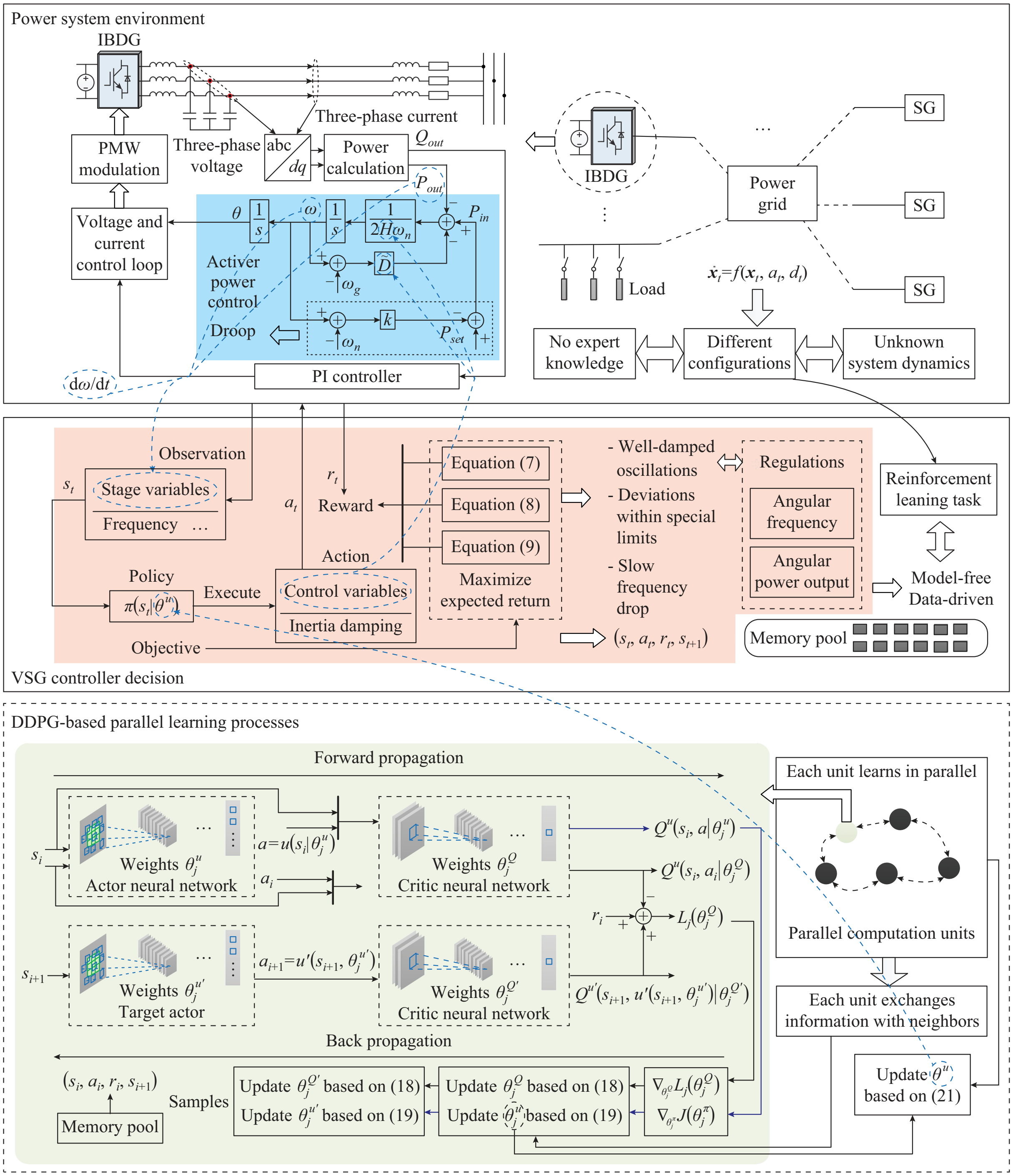

Fig. 1. Overall structure, control, decision and leaning process.

The emulated swing equation of the VSG controller is adopted as:

$$P_{in} - P_{out} = 2\tilde{H}\omega_n \frac{d\omega}{dt} + \tilde{D}(\omega - \omega_g) \tag{1}$$

where P_{in} is the emulated mechanical power; P_{out} is the output active power after low-pass filtering; ω_n is the nominal system angular frequency; ω is the virtual angular frequency of the corresponding IBDG; ω_g is the angular frequency measured by the phase-locked loop (PLL); and $\tilde{H}$ and $\tilde{D}$ are the virtual inertia and damping factor, respectively.

According to the system frequency deviation, the governor is implemented to adjust the input power command, i.e., P_{in}, which adopts the ω-P droop controller as follows:

$$P_{in}=P_{ref}-k(\omega-\omega_n) \tag{2}$$

where P_{ref} and k are the reference active power and droop coefficient, respectively. The choice of k is determined by standard approach [33], which reflects the change of P_{in} with respect to the angular frequency.

Unlike the droop coefficient, the choice of virtual inertia and damping factor is more flexible without special restrictions. Thus, we can adaptively adjust the two controllable parameters over time to obtain the expected performance. Note that increasing or decreasing the control parameters may result in different influences on the dynamic characteristics of the active power output and angular frequency in different system environments.

As a grid-forming converter control, the inertial control performance of a VSG depends on both the control parameter design and power system frequency response ω_g. Hence, in order to optimally design the VSG frequency control, the frequency response model of a complex power system should be considered. On one hand, accurate modeling of power system frequency response requires global information on governor data and generator inertial constants from multiple stations, which is difficult to obtain for local converter control design. On the other hand, the conventional power system frequency response model can no longer describe the frequency trajectory of a power system with high penetration level of renewable energy, which suffers more from deteriorated system frequency profile. Various energy sources including wind turbine generators, PV generators, and battery energy storage systems, have modified the electromechanical behavior of the original power system. Therefore, considering these two aspects, the data-driven control strategy is needed to be developed to optimally adjust VSG control design with the absence of power system model.

For each IBDG, there are two control parameters considered to be adjusted at time t, which is denoted by a_t:

$$a_t=\{\tilde{H}_t,\tilde{D}_t\} \tag{3}$$

To show the dynamic performance, each IBDG is equipped with a VSG controller to observe its real-time states of the output active power, angular frequency, and the derivative of angular frequency, i.e., $P_{out,t}$, ω_t, and $\mathrm{d}\omega_t/\mathrm{d}t$. The set of all observations at time t is defined as s_t:

$$s_t=\{P_{out,t},\omega_t,\mathrm{d}\omega_t/\mathrm{d}t\} \tag{4}$$

Note that the adaptive parameter adjustment a_t is based on the control policy $u(\cdot)$ to be designed and the observed system states s_t. In this paper, a deterministic control policy $u(\cdot)$ is defined as the following function, which maps s_t to a_t:

$$a_t=u(s_t) \tag{5}$$

The nonlinear state-space equation of the whole system in an implicit form can be written as:

$$\dot{\boldsymbol{x}}_t=f(\boldsymbol{x}_t,a_t,d_t) \tag{6}$$

where $\boldsymbol{x}_t$ is the vector of all the state variables, e.g., P_{out}, ω_t, output current and voltage of each IBDG, output frequency, active power of each SG, etc.; and d_t is the uncertain disturbance or variable such as the sudden change of active power reference and load demand, etc. Equation (6) provides a learning environment for the VSG controller. Note that (6) is unknown, which is hard to be modeled with explicit expression. In this paper, we do not need to know the explicit mathematical model of (6). Driven by data, the VSG controller interacts with the environment to obtain the optimal control policy $u(\cdot)$, which will be discussed in the next section in details.

III. Transformation and Solution

It is worth noting that the studied problem in this paper satisfies the Markov property [34]. It means that given the current state and action, the next state is independent of all the previous states. The deep reinforcement leaning algorithm can well tackle the Markov decision processes without relying on a model of the probability distributions underlying the state transitions, which fits well with our work. To get the data-driven adaptive VSG controller, we firstly transform the studied optimal control problem into a reinforcement leaning task. Then, the DDPG algorithm is employed to find the optimal control policy. The overall decision process for the data-driven VSG controller and the leaning diagram for the DDPG algorithm are shown in Fig. 1(b) and 1(c), respectively.

A. Formulation of Reinforcement Learning Task

For a reinforcement learning task, three key elements need to be defined, i.e., observation state, action, and reward. In this paper, the observation state and action correspond to s_t and a_t shown in (4) and (5), respectively. As shown in Fig. 1, the VSG controller interacts with the power system, i.e., learning environment, which is named as power system environment to avoid ambiguity. At each time t, the power system environment provides an observation of s_t to the VSG controller. The VSG controller performs an action from the action space based on policy $u(\cdot)$, and then observes the immediate reward $r(t)$ to update the value of the state-action pair. Next, the interactions of data-driven VSG controller and power system environment via exploration and improvement during the learning process lead the data-driven VSG controller to obtain the approximated optimal control policy. In this paper, we mainly focus on the regulations of angular frequency and active power output. The design of reward $r(t)$ is based on the immediate responses of ω_t, $P_{out,t}$, and $\mathrm{d}\omega_t/\mathrm{d}t$ after disturbances.

With regard to the frequency regulation, the occurrence of poorly damped oscillation is not designed. Define $\psi_\omega=|\omega_t-\omega_n|$ as the absolute value for angular frequency deviation and $\psi_\omega^{\max}$ as the preset upper bound of ψ_ω. There are two cases, i.e., $\psi_\omega\le\psi_\omega^{\max}$ and $\psi_\omega>\psi_\omega^{\max}$, that need to be considered separately. For the case $\psi_\omega\le\psi_\omega^{\max}$, although the frequency deviation is within the allowable limits, we expect the frequency deviation to be as small as possible and the corresponding settling time to be as short as possible. To achieve this goal, we can set a small penalty item for ψ_ω to assess the immediate frequency deviation. Moreover, the larger ψ_ω becomes, the bigger the penalty is. For another case $\psi_\omega > \psi_\omega^{\max}$, the system undergoes huge security risk. Thus, to reduce the occurrence of this situation, we should add a very big penalty once $\psi_\omega > \psi_\omega^{\max}$. Based on the aforementioned discussion, the characteristic function for the deviation of angular frequency is defined as:

$$C(\omega_t)=\begin{cases}\varrho_\omega \psi_\omega & \psi_\omega \le \psi_\omega^{\max}\\ \rho_\omega & \psi_\omega > \psi_\omega^{\max}\end{cases} \tag{7}$$

where ϱ_ω and ρ_ω are the small and big penalty coefficients, respectively.

Note that one major functionality of VSG control is to obtain slow electromechanical dynamics like the SG. In other word, a better transient process should contribute to the reduced rate of change of frequency (ROCOF). To this end, the characteristic function for the change rate of angular frequency is defined as:

$$C(\mathrm{d}\omega_t/\mathrm{d}t)=\varrho_{d\omega}|\mathrm{d}\omega_t/\mathrm{d}t| \tag{8}$$

where $\varrho_{d\omega}$ is a small penalty coefficient.

For the characteristic of active power output, it is also expected to obtain well-damped oscillation. Similar to the functionality of the first part of (7), the characteristic function for the deviation of active power output is defined as:

$$C(P_{out,t})=\varrho_P\psi_P \tag{9}$$

where ϱ_P is the corresponding penalty coefficient; and $\psi_P = |P_{out,t}-P_{ref}|$ is the absolute value for the deviation of active power output. Since P_{ref} may change greatly due to the intermittent renewable energy resources, e. g., wind and solar, it is not important to limit the upper bounder of ψ_P during the transient process. Moreover, the capacity of the inverter is selected so that the headroom is available for necessary inertial support.

According to the expected performance and the characteristic function defined above, the reward at time t is denoted by r_t as:

$$r_t=-b_\omega C(\omega_t)-b_{d\omega}C(\mathrm{d}\omega_t/\mathrm{d}t)-b_P C(P_{out,t}) \tag{10}$$

where $b_\omega>0$, $b_{d\omega}>0$, and $b_P>0$ are the weight coefficients. By choosing different weight coefficients, different output characteristics can be obtained.

Note that the dynamic performance of the active power and angular frequency regulation is measured by a relatively long time reward. For example, we consider a case where a sudden change in load happens at t_0, resulting in large frequency oscillation. The beginning to the end of the frequency oscillation corresponds to a time interval. Whether the dynamic performance gets better or not depends on cumulative penalties for long time response but not for one moment only. To this end, the return from state s_t is further defined as the cumulative future rewards R_t, whose mathematical expression is given by:

$$R_t=\sum_{k=t}^{T}\gamma^{k-t}r_k \tag{11}$$

where T is the total time; and γ is the discount factor.

Then, after making an observation s_t and executing an action a_t, the action value function under the control policy $u(\cdot)$ is the expected return defined as Q^u:

$$Q^u(s_t,a_t)=E[R_t|s_t,a_t,u(\cdot)] \tag{12}$$

where E denotes the expected value of R_t. Our objective becomes finding the optimal control policy $u^*(\cdot)$ that maximizes the expected return from the start of the disturbance.

B. DDPG Algorithm

As stated in Section II, both the system observation state and action are continuous. To account for this attribute, the concept of DPG algorithm based on actor-critic architecture is adopted and further extended in this paper. More importantly, we focus on adopting the decentralized stochastic gradient descent approach to replace the stochastic gradient descent approach in the learning process of traditional DPG algorithm, which is further referred to as DDPG algorithm. By using the DDPG algorithm, the global computation process can be divided into individual computation unit, resulting in faster convergence process. It is assumed that there are κ computation units. The information sharing among the computation units is described by a graph $G=(\mathcal{V},\mathcal{E},\mathcal{W})$, where $\mathcal{V}=\{j=1,2,\ldots,\kappa\}$ is the set of nodes representing the computational units; $\mathcal{V}\subset\mathcal{E}\times\mathcal{E}$ represents the available communication links; $\mathcal{W}=\{w_{j\tilde{j}}\}\in\mathbf{R}^{\kappa\times\kappa}$ is the associated adjacency matrix, and $\tilde{j}$ is the neighbor node of j. It is assumed that graph G is undirected and connected. To achieve experience replay, the experiences $e_t=(s_t,a_t,r_t,s_{t+1})$ at each time step t will be stored in a data set D, which is accessible to every computation unit.

The overall block diagram exhibiting the realization of the policy updating based on the distributed DDGD algorithm is presented in Fig. 1. The actor function is employed to estimate the policy, which maps the observation state of the current power system environment to a specific action deterministically. The critic function is employed to estimate the action value function, in which the output of the actor is fed as one of inputs of the critic. Two neural networks referred to as actor network and critic network are used to approximate the actor and critic functions with parameters θ^u and θ^Q, respectively. In this scenario, the control policy $u(s_t)$ parameterized by θ^u in the actor network is rewritten as $u(s_t|\theta^u)$. Meanwhile, the action value function $Q^u(s_t,a_t)$ parameterized by θ^Q in the critic network is represented by $Q^u(s_t,a_t|\theta^Q)$. Additionally, similar to [26], the separate target networks are used to stabilize the reinforcement leaning algorithm. The updating for the parameters in target networks slowly tracks the actor and critic networks, denoted as $\theta^{u'}$ and $\theta^{Q'}$, respectively. It has been widely verified that learning without target networks does not perform well in many reinforcement learning tasks. For the reinforcement learning task, the exploration in continuous action spaces is important and necessary. In this paper, we employ the exploration policy by adding a random Gaussian disturbance/noise $\delta_t=\mathcal{N}(0,\sigma_t^2 I)$ to the actor policy [35], where σ_t^2 is the variance, and $a_t=u(s_t|\theta^u)+\delta_t$. Note that the random noise is persistently exciting. To obtain effective learning, we often set a large noise during the early learning stages, since no reliable knowledge has been learned by the VSG agent. Thus, more explorations are needed. Later, the magnitude of the noise should be gradually reduced so that the VSG agent can effectively use the accumulated experience to select the action and obtain larger cumulative rewards. To capture this concept, the exponential damping is further employed for σ_t, whose mathematical expression is given by:

$$\sigma_t = \exp(-\Im t) \tag{13}$$

where $\Im$ is the decay rate.

Define θ_j^Q and θ_j^u as the estimated actor network parameters of the j^{th} computation unit, and $\theta_j^{Q'}$ and $\theta_j^{u'}$ as the corresponding parameters in target networks. The loss functions used to update the critic and actor network parameters are given by:

$$L(\theta^Q) = \frac{1}{\kappa}\sum_{j=1}^{\kappa} L_j(\theta_j^Q) \tag{14}$$

$$L_j(\theta_j^Q) = E_{(s_i, a_i, r_i, s_{i+1})\sim D}[(r_i + \gamma Q^{u'}(s_{i+1}, u'(s_{i+1}|\theta_j^{u'})\,|\theta_j^{Q'} - Q^u(s_i, a_i|\theta_j^Q)^2] \tag{15}$$

In this paper, multiple computation units cooperate to train θ^Q. At each step, to minimize (14), every computation unit samples random mini-batch of experiences (s_i, a_i, r_i, s_{i+1}) from the memory pool D to compute local stochastic gradient denoted by $\nabla_{\theta_j^Q} L_j(\theta_j^Q)$. θ_j^u is updated by applying the chain rule to maximize the expected return. Specifically, the mathematical expression of the action gradient using samples for approximating is given by:

$$\nabla_{\theta_j^u} J(\theta_j^u) \approx E_{s_i}[\nabla_a Q^u(s_i, a|\theta_j^Q)|_{a=u(s_i|\theta_j^u)} \nabla_{\theta_j^u} u(s_i|\theta_j^u)] \tag{16}$$

where J is the approximate value function. The parameters θ_j^Q and θ_j^u are further updated via local computation based on the information of its own and that of the neighbors:

$$\theta_j^Q \leftarrow \sum_{\tilde{j}=1}^{\kappa} w_{j\tilde{j}} \theta_{\tilde{j}}^Q - \zeta_Q \nabla_{\theta_j^Q} L_j(\theta_j^Q) \tag{17}$$

$$\theta_j^u \leftarrow \sum_{\tilde{j}=1}^{\kappa} w_{j\tilde{j}} \theta_{\tilde{j}}^u - \zeta_u \nabla_{\theta_j^u} J(\theta_j^u) \tag{18}$$

where ζ_Q and ζ_u are the learning rates. Finally, we can obtain θ^u and θ^Q by using the averaged value of θ_j^Q and θ_j^u for all $j \in \mathcal{V}$.

Based on the current action, the VSG controller will change its control parameters. Then, new transition (s_t, a_t, r_t, s_{t+1}) will be generated, which is used to update the parameters θ^Q and θ^u. Correspondingly, the control policy $u(s_t|\theta^u)$ is updated. After that, the one-step learning process is finished. The detailed learning process based on DDPG algorithm to find the optimal control strategy is presented in Algorithm 1. Note that the DDPG algorithm is employed to train the data-driven VSG controller offline. After that, the well-trained controller can be used in online applications.

$$\theta_j^{Q'} \leftarrow \tau\theta_j^Q + (1-\tau)\theta_j^{Q'} \tag{19}$$

$$\theta_j^{u'} \leftarrow \tau\theta_j^u + (1-\tau)\theta_j^{u'} \tag{20}$$

Algorithm 1: DDPG algorithm

```
Input: Adjacency matrix $\mathcal{W}$; learning rates $\zeta_Q$, $\zeta_u$; mini-batch size $C$; number of episodes $M$; probability $\varepsilon$; smoothing factor $\tau$
Output: Optimal control policy $u(s_t|\theta^u)$
Initialize: Randomly initialize weights $\theta_j^Q$ and $\theta_j^u$ for critic network and actor network $\forall j \in \mathcal{V}$; initialize weights $\theta_j^{Q'} \leftarrow \theta_j^Q$ and $\theta_j^{u'} \leftarrow \theta_j^u$ for target network $\forall j \in \mathcal{V}$; initialize replay buffer $D$
for $episode = 1, 2, \ldots, M$ do
    Initialize a random disturbance for control behavior exploration
    Receive initial initial observation state $s_1$
    for $t = 1, 2, \ldots, T$ do
        Select action $a_t = u(s_t|\theta^u) + \delta_t$ based on current policy and exploration noise $\delta_t$
        Calculate reward using (10)
        Observe the new state $s_{t+1}$
        Store transition $(s_t, a_t, r_t, s_{t+1})$ into $D$
        for $j = 1, 2, \ldots, \kappa$ do
            Randomly sample mini-batch of $M$ transitions $(s_i, a_i, r_i, s_{i+1})$ from $D$
            Calculate stochastic gradient $\nabla L_j(\theta_j^Q)$ by minimizing function (15)
            Calculate sampled policy gradient according to (16)
            Update the estimated critic network parameter $\theta_j^Q$ by the $j^{th}$ computation unit according to (17)
            Update the estimated actor network parameter $\theta_j^u$ by the $j^{th}$ computation unit according to (18)
            Update target networks using (19) and (20)
        end for
        Update critic network parameter using (21)
        Update actor network parameter using (22)
    end for
end for
```

$$\theta^Q = \frac{1}{\kappa}\sum_{j=1}^{\kappa} \theta_j^u \tag{21}$$

$$\theta^u = \frac{1}{\kappa}\sum_{j=1}^{\kappa} \theta_j^u \tag{22}$$

Remark: Compared with the DPG algorithm, the decentralized stochastic gradient descent approach is embedded into the DDPG algorithm. With this effort, the DDPG algorithm can simultaneously employ multiple computation units to train the neural network parameters as shown in (19)-(22), resulting in faster convergence speed than the traditional DPG algorithm. In this paper, the reinforcement learning task is designed for the VSG controller of individual IBDG. It also means that all those parallel computation units are cooperative to train one VSG controller as shown in Fig. 1. To reduce the training time, this paper employs the DDPG algorithm.

IV. Simulation Results

In this section, we focus on verifying the effectiveness and feasibility of the DDPG algorithm with simulations in a modified IEEE 14-bus test system [32]. The topology of the modified test system is shown in Fig. 2. It is composed of two synchronous generators, one 2.5 MW IBDG installed at bus 14, twelve loads, and one load disturbance. Therein, the load disturbance is located at bus 4 with green arrow for the sake of distinction. One of the initiatives of integrating VSG into the power system is to mitigate the deteriorated system frequency regulation resulting from high penetration level of renewable energy. Hence, the system frequency transients after load disturbances are considered to train the VSG controller. To simulate the disturbances, we let the load disturbance randomly change within interval [0.2, 1.4]MW. Meanwhile, the active power reference is randomly chosen within interval [0.5, 1.8]MW. We consider four computation units interconnected with each other to form a ring communication network. In order to maintain sufficient rotor-angle stability margin, as a grid-forming control approach, the line impedance should be considered for the design of VSG system. As a result, the imitated rotor angle of VSG should be sufficiently small at rated power, such that the proposed VSG is able to ride through certain system faults during operation [19]. The simulations are conducted in MATLAB/Simulink software. Next, the first case study focuses on training the actor and critic neural networks to obtain the optimal control policy. The performance evaluation of the well-trained VSG controller will be tested after load disturbance and active power change in the second and third case studies, respectively.

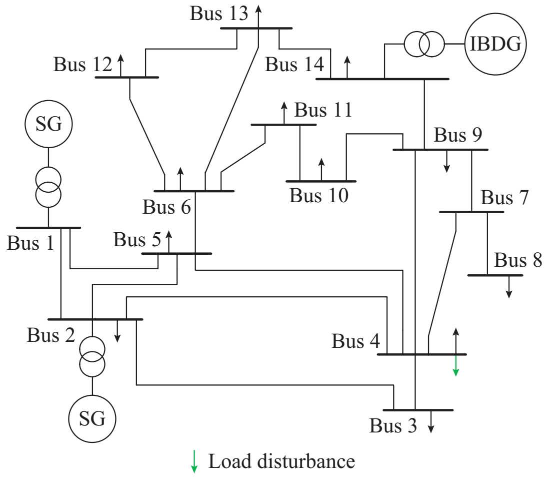

Fig. 2. Modified IEEE 14-bus test system.

A. Training Neural Networks and Comparison

In this case study, the adopted structures of the actor and critic neural networks are shown in Fig. 3. The critic network consists of the state path, action path, and common path. Therein, the observations and the actions are the inputs for state path and action path. The outputs of the state path and action path will be combined into one layer which are also the inputs of the common path. The output of the common path is the estimated action value function. For the actor network, the corresponding inputs and outputs are the observations and actions, respectively. The terms *ReLU* and tanh mentioned in Fig. 3 are the standard activation functions for neurons, which are widely used in the design of deep neutral network. Specifically, *ReLU* and tanh are the rectified linear unit function and hyperbolic tangent function, respectively, whose explicit formulations are given by:

$$\begin{cases} ReLU(x) = \max(0, x) \\ \tanh(x) = \dfrac{\mathrm{e}^{x} - \mathrm{e}^{-x}}{\mathrm{e}^{x} + \mathrm{e}^{-x}} \end{cases} \tag{23}$$

Moreover, the fully connected layer multiplies the input by a weight matrix and then adds a bias vector. The scaling layer is used for scaling the input variables. The rest of simulation parameters are listed in Table I. The DDPG algorithm is trained over $M = 350$ episodes by using 31 h 23 min and 32 s. The cumulative reward for each episode, named as episode reward, is shown in Fig. 4(a). It can be observed that there is no obvious improvement for the episode reward from the episode numbers 250 to 350. This implies that the DDPG becomes stable. Thus, the training can be stopped after 350 episodes. Meanwhile, the parameters for the actor and critic neural networks are saved, and then the optimal control policy is obtained. In addition, during the learning process, there are no requirements for any expert experience or the whole system model. We can obtain the optimal control policy based on numerous explorations and improvements driven by observation data only. Finally, the optimal control policy is embedded into the VSG controller resulting in well-trained VSG controller.

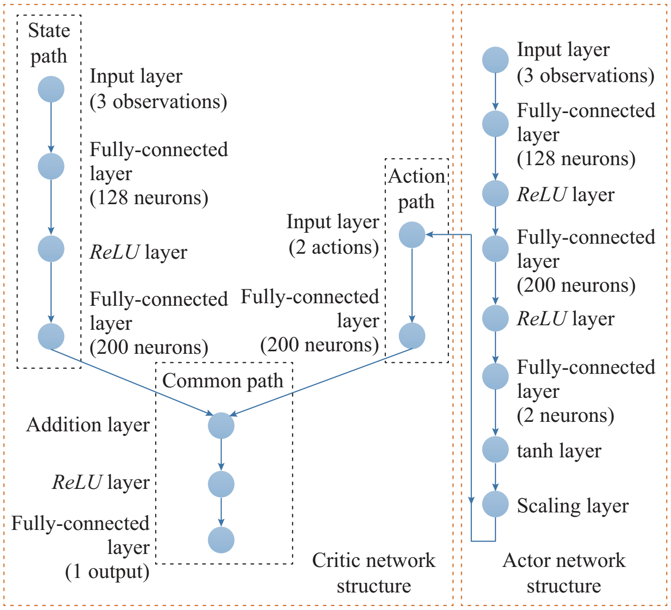

Fig. 3. Structures of critic and actor networks.

TABLE I
TRAINING PARAMETERS

Parameter	Value	Parameter	Value
ϱ_{ω}	10	γ	0.9
$\varrho_{d\omega}$	2	ζ_{Q}	0.001
ϱ_{P}	2	ζ_{u}	0.0005
ρ_{ω}	1000	τ	0.001
b_{ω}	1/3	$\psi_{\omega}^{\max}$	$2\pi \times 0.8$ Hz
$b_{d\omega}$	1/3	$\mathfrak{I}$	0.001
b_{P}	1/3		

Next, the traditional DPG algorithm is employed to solve the same problem, which can be seen as a special case of the DDPG algorithm with one computation unit, i. e., $\kappa = 1$. Meanwhile, the decentralized stochastic gradient descent approach is changed into the stochastic gradient descent approach during back propagation. With the same neural network structures and parameters, the episode reward obtained by using the DPG algorithm is shown Fig. 4(b). The total training time is 68 h 17 min and 25 s, which is longer than that using DDPG algorithm. In addition, it can been observed from Fig. 3(a) and 3(b) that the DDPG algorithm requires fewer episodes than the DPG algorithm to achieve the similar episode reward. These results exhibit the faster convergence feature of the DDPG algorithm. This is because the DDPG is able to use multiple computation units simultaneously to accelerate the training process.

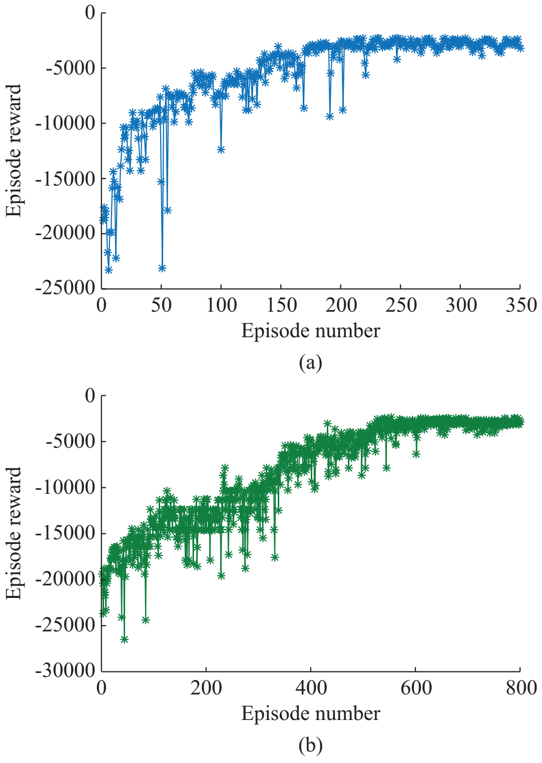

Fig. 4. Cumulative reward for each episode. (a) DDPG algorithm. (b) DPG algorithm.

B. Load Disturbance

In this case study, we aim at verifying the effectiveness of the well-trained VSG controller under load disturbance. At $t=20$ s, a 0.7 MW load disturbance is added in the test system. The simulation results are shown in Figs. 5 and 6. It can be observed that the IBDG can respond to the load disturbance adaptively and automatically. Specifically, the maximum angular frequency deviation is $2\pi\times 0.42$ Hz, which is within the preset upper bound of $\psi_{\omega}^{\max}=2\pi\times 0.8$ Hz. Meanwhile, the frequency changes relatively slow, i.e., with small ROCOF. As a result, the IBDG possesses slow frequency drop, which meets the major functionality of VSG control. Moreover, the oscillation of active power output is also well damped. Note that by implementing the well-trained VSG controller, the tradeoff between the frequency response and active power output can be achieved and maintained as desired, which fulfills the expected performance discussed in Section III-A. This is because the design of immediate reward provides the penalty for bad performance. Then, driven by the stimulation of long-term return, satisfactory results can be obtained. In addition, the secondary frequency control is not included and the frequency deviation at system steady state relates to predefined droop parameters of individual generation unit. Based on the above-mentioned discussions, it can be concluded that the well-trained VSG controller possesses good adaptability and performs well after load disturbance.

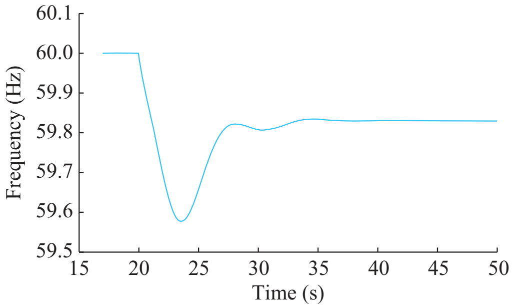

Fig. 5. Frequency response after load disturbance.

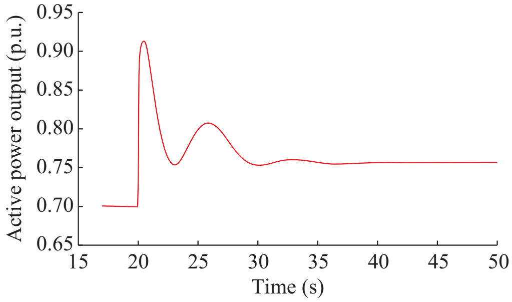

Fig. 6. Active power output of IBDG after load disturbance.

C. Change of Power Reference

In this case study, the focus is on testing the effectiveness of the well-trained VSG controller after the change of active power reference. At $t=20$ s, there is a step change for active power reference from 0.7 p.u. to 0.5 p.u.. The simulation results for the frequency response and active power output of the IBDG are shown in Figs. 7 and 8, respectively.

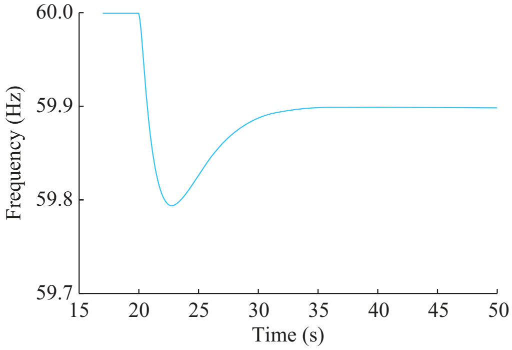

Fig. 7. Frequency response after change of power reference.

As observed, both the frequency and active power output gradually converge to a new stable equilibrium with well-damped oscillations, and the system ROCOF is mitigated. Thus, the expected performance targets are fulfilled. This implies that the well-trained VSG controller exhibits better adaptability and works well after the change of power reference.

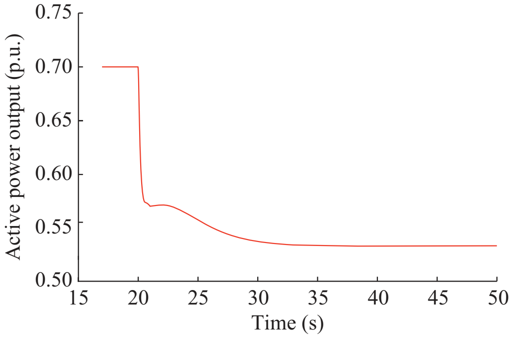

Fig. 8. Active power output of IBDG after change of power reference.

D. Performance Test in a New Test System

In this case study, the performance of the well-trained VSG controller obtained from the first case study is further tested in a new IEEE 14-bus test system, which is different from that used in offline training. Specifically, the SG at bus 2 is replaced with an IBDG and the IBDG at bus 14 is disconnected. Referring to the structure of IEEE 14-bus test system, three synchronous condensers are commissioned at bus 3, bus 8, and bus 6, respectively. By replacing the system SG with IBDG and integrating synchronous condensers, the equivalent inertial constant and frequency response model of the system are inevitably changed. At time $t=20$ s, a 0.4 MW load disturbance at bus 4 is added in the test system. The comparative system frequency responses and active power outputs with different converter controls after load disturbance are shown in Figs. 9 and 10, respectively.

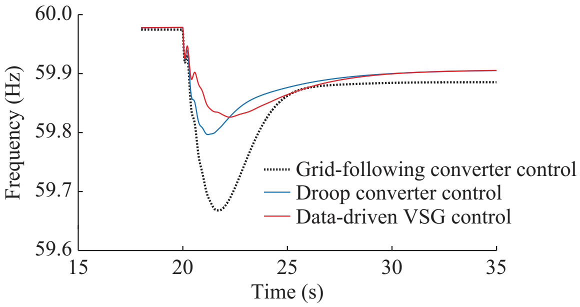

Fig. 9. Frequency responses with different converter controls after load disturbance.

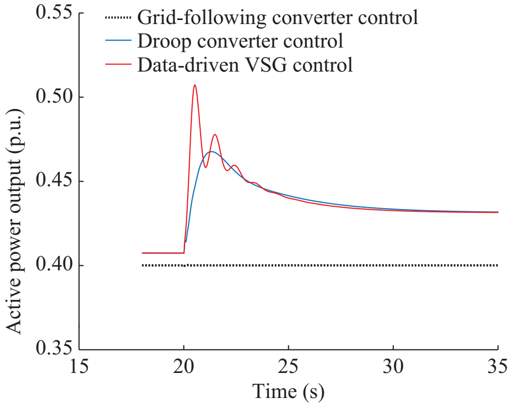

Fig. 10. Active power responses with different converter controls after load disturbance.

Typically, the grid-following converter control approach does not participate in power system frequency regulation, where it simply follows the system frequency through PLL. Both droop converter control and VSG are able to participate in the power system frequency regulation and enhance the system small-signal stability due to their grid-forming nature. Furthermore, the proposed data-driven VSG control is able to better arrest the ROCOF of power system and provide necessary inertial control. Meanwhile, the oscillation of active power output is also well damped. Note that it is impossible for the data-driven VSG controller to be trained in all transient scenarios.

Next, we further test the performance of the proposed VSG controller after fault transient. The system dispatching scenario is the same as that presented in Figs. 9 and 10. At $t=20$ s of the simulation time, a fault is introduced at the transmission line that connects bus 2 and bus 3. The fault lasts for 10 cycles and trips the transmission line. The simulation results with different converter controls are shown in Fig. 11. Note that the well-trained VSG controller is not trained in the fault transient scenario or in the new test system. Thus, the optimality of the convergence results cannot be guaranteed. However, taking advantage of the introduced virtual inertia from VSG, the power system stability can be enhanced, where the frequency deviation of the power system integrated with VSG is less than the other two cases in Fig. 11, and the synchronism [19] is better preserved.

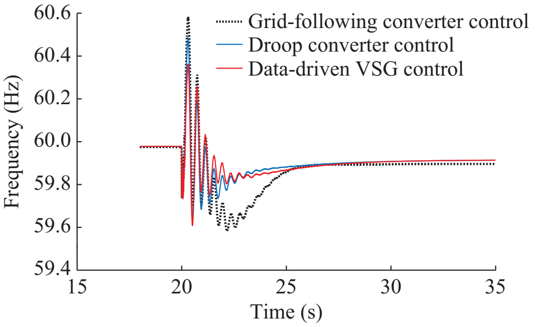

Fig. 11. Frequency responses with different converter controls after fault transient.

The simulation results show that the VSG controller also works well in the new test system. However, the better performance cannot always be ensured in any kind of new systems, since it is not trained in the new environment. In practical application, the VSG controller requires re-training if used in different systems.

V. Conclusion

This paper investigates the adaptive and optimal control problem for VSG. To achieve the expected control performance target for frequency regulation and active power regulation, multiple characteristic functions are defined and further used to form the immediate reward. With this effort, the optimal control problem is finally formulated as a reinforcement learning task. To handle this task, the DDPG algorithm is employed to learn the optimal control policy with the objective of maximum long-term return. The implementation of the DDPG algorithm does not need any expert knowledge and does not rely on the system model. Thus, we can obtain the optimal control policy in a model-free fashion, which is the major advantage compared with the existing optimal control approaches used in VSG. In the future, the voltage stability and further application of the DDPG algorithm will be considered.

References

[1] H. Zhang, Y. Li, D. W. Gao *et al*., "Distributed optimal energy management for energy internet," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3081-3097, Dec. 2017.

[2] J. Zhou, Y. Xu, and H. Sun, "Distributed power management for networked AC/DC microgrids with unbalanced microgrids," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1655-1667, Mar. 2020.

[3] Y. Li, H. Zhang, X. Liang *et al*., "Event-triggered based distributed cooperative energy management for multienergy systems," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 14, pp. 2008-2022, Apr. 2019.

[4] Y. Li, D. W. Gao, W. Gao *et al*., "Double-mode energy management for multi-energy system via distributed dynamic event-triggered Newton-Raphson algorithm," *IEEE Transactions on Smart Grid*, vol. 11, no. 6, pp. 5339-5356, Nov. 2020.

[5] R. Wang, Q. Sun, D. Ma *et al*., "The small-signal stability analysis of the droop-controlled converter in electromagnetic timescale," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1459-1469, Jul. 2019.

[6] Z. Yi, Y. Xu, W. Gu *et al*., "A multi-time-scale economic scheduling strategy for virtual power plant based on deferrable loads aggregation and disaggregation," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 3, pp. 1332-1346, Jul. 2020.

[7] J. Zhou, Y. Xu, H. Sun *et al*., "Distributed event-triggered H_∞ consensus based current sharing control of DC microgrids considering uncertainties," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7413-7425, Dec. 2020.

[8] Y. Li, D. W. Gao, W. Gao *et al*., "A distributed double-Newton descent algorithm for cooperative energy management of multiple energy bodies in energy internet," *IEEE Transactions on Industrial Informatics*, doi: 10.1109/TII.2020.3029974

[9] Q. Zhong and G. Weiss, "Synchronverters: inverters that mimic synchronous generators," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 4, pp. 1259-1267, Apr. 2011.

[10] Q. Zhong, "Virtual synchronous machines: a unified interface for grid integration," *IEEE Power Electronics Magazine*, vol. 3, no. 4, pp. 18-27, Dec. 2016.

[11] J. Chen and T. O'Donnell, "Parameter constraints for virtual synchronous generator considering stability," *IEEE Transactions on Power Systems*, vol. 34, no. 3, pp. 2479-2481, May 2019.

[12] Z. Yi, Y. Xu, J. Zhou *et al*., "Bi-level programming for optimal operation of an active distribution network with multiple virtual power plants," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 4, pp. 2855-2869, Oct. 2020.

[13] J. Lee, G. Jang, E. Muljadi *et al*., "Stable short-term frequency support using adaptive gains for a DFIG-based wind power plant," *IEEE Transactions on Energy Conversion*, vol. 31, no. 3, pp. 6289-6297, Sep. 2016.

[14] D. Li, Q. Zhu, S. Lin *et al*., "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, Mar. 2017.

[15] F. Wang, L. Zhang, X. Feng *et al*., "An adaptive control strategy for virtual synchronous generator," *IEEE Transactions on Industry Applications*, vol. 54, no. 5, pp. 5124-5133, Sept. 2018.

[16] J. Li, B. Wen, and H. Wang, "Adaptive virtual inertia control strategy of VSG for micro-grid based on improved bang-bang control strategy," *IEEE Access*, vol. 7, pp. 39509-39514, Mar. 2019.

[17] H. Wu, X. Ruan, D. Yang *et al*., "Small-signal modeling and parameters design for virtual synchronous generators," *IEEE Transactions on Industrial Electronics*, vol. 63, no. 7, pp. 4292-4303, Jul. 2016.

[18] M. Li, W. Huang, N. Tai *et al*., "A dual-adaptivity inertia control strategy for virtual synchronous generator," *IEEE Transactions on Power Systems*, vol. 35, no. 1, pp. 594-604, Jan. 2020.

[19] H. Wu and X. Wang, "A mode-adaptive power-angle control method for transient stability enhancement of virtual synchronous generators," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 2, pp. 1034-1049, Jun. 2020.

[20] 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. 1463-1471, Mar. 2018.

[21] W. Du, Q. Fu, and H. Wang, "Power system small-signal angular stability affected by virtual synchronous generators," *IEEE Transactions on Power Systems*, vol. 34, no. 4, pp. 3209-3219, Jul. 2019.

[22] U. Markovic, Z. Chu, P. Aristidou *et al*., "Fast frequency control scheme through adaptive virtual inertia emulation," in *Proceedings of 2018 IEEE Innovative Smart Grid Technologies – Asia*, Singapore, Singapore, Mar. 2018, pp. 787-792.

[23] U. Markovic, Z. Chu, P. Aristidou *et al*., "LQR-based adaptive virtual synchronous machine for power systems with high inverter penetration," *IEEE Transactions on Sustainable Energy*, vol. 10, no. 3, pp. 1501-1511, Jul. 2019.

[24] W. Cjch and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279-292, May 1992.

[25] D. Silver, G. Lever, N. Heess *et al*., "Deterministic policy gradient algorithms," in *Proceedings of the 31st International Conference on Machine Learning*, Beijing, China, Jun. 2014, pp. 387-395.

[26] V. Mnih, K. Kavukcuoglu, D. Silver *et al*., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529-533, Feb. 2015.

[27] T. P. Lillicrap, J. J. Hunt, A. Pritzel *et al*. (2019, Jul.). Continuous control with deep reinforcement learning. [Online]. Available: https://arxiv.org/abs/1509.02971v2

[28] J. Schulman, P. Moritz, S. Levine *et al*. (2018, Oct.). High-dimensional continuous control using generalized advantage estimation. [Online]. Available: https://arxiv.org/abs/1506.02438

[29] Y. Li. (2018, Nov.). Deep reinforcement learning: an overview. [Online]. Available: https://arxiv.org/abs/1701.07274

[30] X. Lian, W. Zhang, C. Zhang *et al.* (2018, Sept.). Asynchronous decentralized parallel stochastic gradient descent. [Online]. Available: https://arxiv.org/abs/1710.06952

[31] W. Du, Z. Chen, K. P. Schneider *et al*., "A comparative study of two widely used grid-forming droop controls on microgrid small signal stability," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 8, no. 2, pp. 963-975, Jun. 2020.

[32] M. I. Jordan and T. M. Mitchell, "Machine learning: trends, perspectives, and prospects," *Science*, vol. 349, no. 6245, pp. 255-260, Jul. 2015.

[33] R. Wang, Q. Sun, P. Zhang *et al*., "Reduced-order transfer function model of the droop-controlled inverter via Jordan continued-fraction expansion," *IEEE Transactions on Energy Conversion*, vol. 35, no. 3, pp. 1585-1595, Sept. 2020.

[34] W. Yan, L. Cheng, S. Yan *et al*., "Enabling and evaluation of inertial control for PMSG-WTG using synchronverter with multiple virtual rotating masses in microgrid," *IEEE Transactions on Sustainable Energy*, vol. 11, no. 2, pp. 1078-1088, Apr. 2020.

[35] P. Wawrzynski, "Control policy with autocorrelated noise in reinforcement learning for robotics," *International Journal of Machine Learning and Computing*, vol. 5, no. 2, pp. 91-95, Apr. 2015.

Yushuai Li received the B.S. degree in electrical engineering and automation, and the Ph.D. degree in control theory and control engineering from the Northeastern University, Shenyang, China, in 2014 and 2019, respectively. He serves as an Editor for Frontiers in Energy Research, and a Guest Editor for Complexity on the Special Issue "Theory and Applications of Cyber-Physical Systems". His main research interests include distributed modelling, control, energy management and optimization of Energy Internet and multi-energy systems, as well as distributed machine learning algorithm with applications in microgrids.

Wei Gao received his B.S. degree in automation from Hebei University of Technology, Tianjin, China, in 2017. He is pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, University of Denver, Denver, USA. His research interests include microgrid control, renewable energy, and power system stability.

Weihang Yan received the B.S. and M.S. degrees in control theory and engineering from Northeastern University, Shenyang, China, in 2014 and 2016, respectively, and the Ph.D. degree in Electrical and Computer Engineering from University of Denver, Denver, USA, in 2020. He is currently a

Postdoctoral Researcher with the Power Systems Engineering Center, National Renewable Energy Laboratory (NREL), Golden, USA. His research interests include stability analysis and control of power systems with high penetration of renewable generation.

Shuo Huang received the B. S. degree from City College of Science and Technology, Chongqing University, Chongqing, China, in 2018. He is currently pursuing the master's degree in electrical and computer engineering, University of Denver, Denver, USA. His research interests include adaptive control and stability analysis of power system.

Rui Wang received the B. S. degree in Northeastern University, Shenyang, China, in 2016. He is currently pursuing the Ph. D. degree in the School of Information Science and Engineering, Institute of Automation, Northeastern University, Shenyang, China. Since 2019, he has become a Visiting Scholar with the Energy Research Institute, Nanyang Technological University, Singapore, Singapore. His current research interests include collaborative optimization of distributed generation, and stability analysis of electromagnetic time scale in Energy Internet.

Vahan Gevorgian received the Ph. D. degree in electrical engineering from the State Engineering University of Armenia, Yerevan, Armenia, in 1993. He joined National Renewable Energy Laboratory (NREL), Golden, USA, in October 1994. He is currently working with the Power Systems Engineering Center focusing on renewable energy impacts on transmission and interconnection issues and dynamic modeling of variable generation systems. His research interests include dynamometer and field testing of large and small wind turbines, dynamometer testing of wind turbine drivetrain components, development of advanced data acquisition systems, and wind turbine power quality.

David Wenzhong Gao received his M. S. and Ph. D. degrees in electrical and computer engineering, specializing in electric power engineering, from Georgia Institute of Technology, Atlanta, USA, in 1999 and 2002, respectively. He is currently with the Department of Electrical and Computer Engineering, University of Denver, Denver, USA. He is an Associate Editor for IEEE Journal of Emerging and Selected Topics in Power Electronics, and Journal of Modern Power Systems and Clean Energy. He was an Editor of IEEE Transactions on Sustainable Energy. He is the General Chair for the 48th North American Power Symposium (NAPS 2016) and the IEEE Symposium on Power Electronics and Machines in Wind Applications (PEMWA 2012). His current teaching and research interests include renewable energy and distributed generation, microgrid, smart grid, power system protection, power electronics applications in power systems, power system modeling and simulation, and hybrid electric propulsion systems.