Mechanism Analysis and Real-time Control of Energy Storage Based Grid Power Oscillation Damping

A Soft Actor-Critic Approach

Li, Tao; Hu, Weihao; Zhang, Bin; Zhang, Guozhou; Li, Jian; Chen, Zhe; Blaabjerg, Frede

Published in:
I E E E Transactions on Sustainable Energy

DOI (link to publication from Publisher):
10.1109/TSTE.2021.3071268

Publication date:
2021

Document Version
Accepted author manuscript, peer reviewed version

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

Mechanism Analysis and Real-time Control of Energy Storage Based Grid Power Oscillation Damping: A Soft Actor-Critic Approach

Tao Li, Weihao Hu, *Senior Member, IEEE*, Bin Zhang, Guozhou Zhang, Jian Li, *Member, IEEE*, Zhe Chen, *Fellow, IEEE*, *Frede Blaabjerg*, *Fellow, IEEE*

***Abstract*—In this paper, the mechanism of energy storage (ES)-based power oscillation damping is derived by the small signal and the classical electric torque method. And then, by cooperating PI with an integral reduction loop, a controller is designed to form a novel PI-IR controller to guarantee that the energy variation of ES damper is zero at the end of one oscillation. Furthermore, for the controller parameters tuning, the conventional model-based methods require a forecasting model on the uncertainty disturbances. To this end, this problem is formulated as a finite Markov decision process with unknown transition probability, and introduce a deep reinforcement learning (DRL) based model-free agent, the soft actor-critic, to obtain the real-time optimal control strategy. After numerous training, the well-trained agent can act as an experienced decision maker to provide the real-time near-optimal parameters setting for PI-IR control under different operating conditions. Time-domain and eigenvalue analysis results demonstrate the effectiveness of the proposed PI-IR controller and the superiority of the employed DRL based model-free method.**

***Index Terms*—Energy storage, power system stability, PI, PI-IR, deep reinforcement learning.**

I. INTRODUCTION

TO accelerate the transformation of energy structure, the large-scale renewable energy (RE) is penetrating into the power system. This transformation, due to the uncertainty characteristics of RE, has brought great challenges to the safety and stability of the power system operation and control [1], [2]. Some researches consider the impact of grid-connected RE on power system stability, and the results show that the fluctuation power will greatly affect the power system including frequency stability [3], voltage stability [4] and transient stability [5]. When the RE is incorporated into the power system, the proportion of traditional synchronous generator (SG) decreases, resulting in a reduction in the equivalent inertia of the power system. The lower inertia further reduces the stability margin of the system [6], and triggers power oscillations more frequently.

The traditional power system, consists of synchronous generator, can cope with second-level power fluctuation, the change rate of load is relatively slow. But the output fluctuation of renewable energy is basically millisecond-level and uncontrollable. This difference in regulation ability would lead to the imbalance between power supply and demand of the high renewable energy penetration power system, and furtherly causes great challenges to the power system operation and control.

Traditionally, the power oscillation suppression device is the power system stabilizer (PSS), a linearized controller around the normal operating point [7], which improves the damping capacity of the power system by controlling the excitation of SG. Apart from the above PSSs–based suppression approaches, the flexible AC transmission systems (FACTS) and energy storage (ES)-based methods are also widely used to deal with this problem [8]-[10]. For instance, A. Chakrabortty [8] proposed a wide-area damping control method to mitigate the electromechanical oscillation in large power systems by Thyristor controlled series compensators.

The essential factor causing the security and stability of power system is insufficient system damping and power imbalance. As an alternative solution, the energy storage device, is connected to the power system through power converter and appropriate control strategy, has the characteristic of fast response speed, and can compensate the power fluctuation of renewable energy in time. With the wide application of energy storage technologies, which provides the security protection for the access of large-scale renewable energy to traditional power system. For example, a deterministic and an interval unit commitment co-optimization of controllable power source and pumped hydro energy storage is proposed in [9]. Y. Zhu et al. [10] studied the battery energy storage to improve the power stability from the view of both the placement and controller parameters optimization.

To ensure the effectiveness of the controllers, some design methods established on the intelligence heuristic algorithms are proposed. For instance, the non-dominated sorting in genetic algorithms-II (NSGA-II) was used to tune the proportional-integral-derivate (PID) controller parameters to enhance the performance of a FACTS-based stabilizer [11]. X. Sui et al. [12] proposed an ES–based method to damp the inter-area

This work was supported by the Sichuan Science and Technology Program under Grant 2020JDJQ0037 and 2020YFG0312.

(Corresponding author: Weihao Hu)

Tao Li, Weihao Hu, Bin Zhang, Guozhou Zhang and Jian Li are with the School of Mechanical and Electrical Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: tli@std.uestc.edu.cn; whu@uestc.edu.cn; sven@std.uestc.edu.cn; zgz@std.uestc.edu.cn; Leejian@uestc.edu.cn;

Zhe Chen and Frede Blaabjerg are with the Department of Energy Technology, Aalborg University, Aalborg, Denmark (e-mail: zch@et.aau.dk; fbl@et.aau.dk).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

oscillation, and the particle swarm optimization (PSO) algorithm was applied to tune the power oscillation dampers (POD) parameters. Similarly, M. Beza et al. [13] designed an adoptive POD controller for static synchronous compensator (STATCOM) equipped with ES. The modified imperialist competitive algorithm (MICA) combined with a probabilistic eigenvalue approach was proposed in [14] to optimize the parameters of the PSS, the POD of doubly fed induction generators (DFIGs) and STATCOM controllers. The advantage of the heuristic method is that no model information is required. However, the optimal parameters obtained by this method are based on a fixed operating point. The high penetration RE will cause the operating point of the system to vary within a relatively large range. The aforementioned methods achieved some success in power oscillation, but these methods may be unsuitable for real-time operation where the variations in RE output and the stochastic fault are much more complicated.

In recently, developing the emerging real-time control strategy has been recognized as an important way to handle the time-varying operating point caused by intermittent RE output. For example, A. S. Mir et al. [15] developed a deep neural network (DNN) –based actor-critic (AC) algorithm for real-time power oscillation control of interconnected power systems. Similarly, a DNN was applied to tune the parameters for STATCOM to enhance the low-frequency oscillation damping in [16]. Y. Guo et al. [17] developed an adaptive gain scheduling droop voltage control strategy for wind power plant based on the data-driven real-time system equivalent approach to achieve the voltage/reactive power control. The above approaches aim to formulate the power system stability problem as a model-based control problem, and can obtain a good response performance for the power system stability control. However, the model-based control strategies place too much reliance on the accurate model of the power system. Unfortunately, the mathematical model and parameters of the system are not always known accurately. In this condition, the model-based control strategies may be invalid, and the reliability of the system cannot be guaranteed.

Lately, the model-free methods, which are independent on any system model information, have been implemented significantly success in complicated decision-making application [18]. Inspired by [18], the development of model-free methods for smart grid applications have attracted a lot of attention recently [19]-[22]. The advantage of the model-free method over the model-based method is that it can learn a top-quality control policy based on the deep reinforcement learning (DRL) technique and does not depend on the model of the system. For instance, G. Z. Zhang et al. [19] introduced the Deep Deterministic Policy Gradient (DDPG) algorithm to train the agent on learning the real-time control strategy for STATCOM -based additional damping controller for wind farm. Similarly, C. Chen et al. [20] developed a DDPG–based approach to learn an emergency frequency strategy. Y. Hashmy et al. [21] proposed a faster exploration -based DDPG approach to overcome communication delays and other non-linearity challenges in a wide-area system for low-frequency oscillation damping control. To effectively suppress the ultra-low frequency oscillations, G. Z. Zhang et al. [22] proposed a novel proportional resonance (PR) based PR-PSS controller. Furthermore, the DRL algorithm asynchronous advantage actor-critic (A3C), is employed to set the real-time parameters of PR-PSS under the uncertainty scenario. These DRL model-free approaches have achieved an effective response performance for the modern power system operation and control. However, to the best of author's knowledge, this is the first study to investigate the mechanism analysis and apply the state-of-the-art DRL approach to real-time control of ES-based grid power oscillation damping problem.

In this paper, the mechanism of ES-based damper is explicitly analyzed via the small signal and classical electric torque method. And then, a novel proportional integral controller with integral reduction (PI-IR) suitable for ES to suppress power oscillation is designed. Finally, the real-time control problem of an ES-based damper is formulated as a finite Markov decision process (MDP). The objective is to quickly suppress the grid power oscillation while finding the cost-efficient charging/discharging scheme by tuning the PI-IR controller parameters. A model-free approach is introduced to tune the optimal parameters in the real-world scenarios. Specifically, the developed approach takes the oscillation duration and the integral of ES charging/discharging power as input, and outputs the real-time parameters of PI-IR controller.

The contribution of this paper can be summarized:

1) The mechanism analysis of ES-based grid power oscillation damping is designed and implemented via small signal and damping torque coefficient method.

2) A novel PI-IR controller is proposed to damp power oscillation of SG and reduce energy deviation of ES.

3) A soft actor-critic (SAC) model-free algorithm which does not require any model information is proposed to tune the real-time optimal parameters for the PI-IR controller.

The rest of this paper is organized as follows. The mechanism analysis of ES to suppress grid power oscillation and controller design is presented in Section II. Then, the state-of-the-art algorithm is introduced in Section III to solve the controller parameters tuning problem. In Section IV, the simulation is carried out to validate the effectiveness of the proposed approach. Finally, Section V gives the conclusion.

II. Mechanism Anslysis and Controller Design

The single-machine model connected to a strong power grid is presented in Fig. 1, deriving the mathematical model of electromagnetic power oscillation. It consists of seven parts: synchronous generator (SG), transmission line, strong grid, battery bank, voltage source converter (VSC), the LC-filter as well as grid-connected control part. These seven parts are connected by two loops: the electricity loop and the control loop (current control inner-loop and speed control outer-loop). When it comes to the grid power oscillation, the ES can supply the variable real power to compensate electromagnetic power of the SG based on the feed-back signal and the appropriate control strategy. As the ES interacts with SG through the VSC, the ES can be regarded as a controllable current source. To mitigate the real power oscillation, the reactive current of ES can be set to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

zero, that is, the phase of the ES charging/discharging current is in phase with grid-connected point of common coupling (PCC) voltage $V\angle 0$.

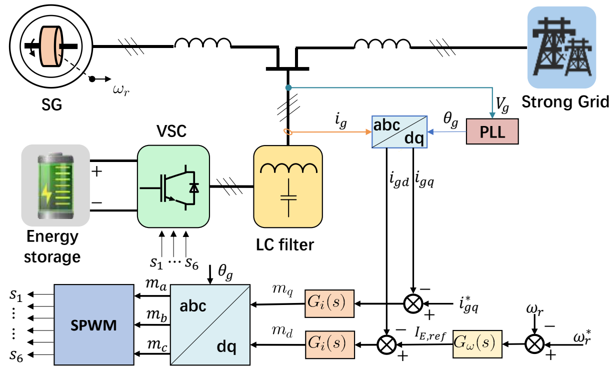


Fig. 1. Structure of the modern flexible power system with an ES.

To analyze the influence of ES on the power system oscillation, second-order model of the SG is introduced. It can be represented by an equivalent voltage source $E\angle\delta$ behand the *d*-axis equivalent reactance X_d, and its magnitude is assumed to remain constant value at the pre-perturbation. The strong grid and transmission line are modeled as a constant $U\angle-\theta$ and reactance X_T. Specifically, the simplified equivalent system is shown as Fig. 2.

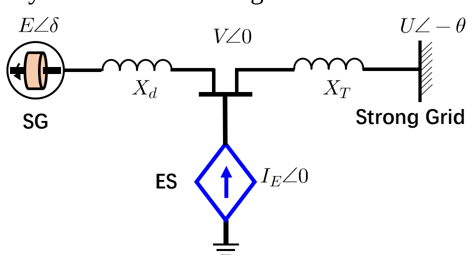


Fig. 2. Simplified equivalent system of the Fig. 1.

A. Electromagnetic Power Formulation of SG-ES

Application of Kirchhoff's law of current to the grid-connected PCC from Fig. 2 results in:

$$\frac{E\angle\delta - V\angle 0}{jX_d} + I_E\angle 0 = \frac{V\angle 0 - U\angle-\theta}{jX_T} \tag{1}$$

Separating the real and imaginary parts from (1):

$$\begin{cases} V\ X_d + X_T\ - EX_T\cos\delta = UX_d\cos\theta \\ X_dX_TI_E + EX_T\sin\delta = UX_d\sin\theta \end{cases} \tag{2}$$

In general, eliminating θ based on the equation $\sin\theta^2 + \cos\theta^2 = 1$ in (2), it can be simplified as follows:

$$\begin{aligned}(UX_d)^2 &= [V\ X_d + X_T\ - EX_T\cos\delta]^2 \\ &+ [X_dX_TI_E + EX_T\sin\delta]^2\end{aligned} \tag{3}$$

Linearizing (3) around the pre-disturbance operating point represented by $\delta = \delta_0$ and $I_E = I_{E0}$ results in:

$$\begin{aligned}\Delta V &= \frac{[k_1EV_0\sin\delta_0 + X_dX_T{}^2EI_{E0}\cos\delta_0]\Delta\delta}{k_1E\cos\delta_0 - k_2V_0} \\ &+ \frac{[X_d{}^2X_T{}^2I_{E0} + X_dX_T^2E\sin\delta_0]\Delta I_E}{k_1E\cos\delta_0 - k_2V_0}\end{aligned} \tag{4}$$

where "Δ" denotes the small variation. k_1 and k_2 are the simplified reactance coefficient, $k_1 = X_T(X_d + X_T)$, $k_2 = (X_d + X_T)^2$.

Based on Fig. 2, the electromagnetic power generation of SG is formulated as:

$$P_e = \frac{EV}{X_d}\sin\delta \tag{5}$$

Substituting for ΔV in linearized (5) yields the electromagnetic power formulation of SG-ES:

$$\Delta P_e = k_g\Delta\delta - k_E\Delta I_E \tag{6}$$

Equation (6) reveals that the charging or discharging current of ES will affect the grid power oscillation process. where k_g reveals the synchronous and stable operation capability of SG, named utility grid synchronization coefficient. k_E represents the effect of ES on the dynamic characteristic of SG under ES grid-connected strategy, named ES grid-connected coefficient. k_g and k_E are constant value in (7) and (8), derived from the pre-disturbance operating condition.

$$\begin{aligned}k_g = \frac{E}{X_d}\Bigg(&\frac{k_1EV_0 + X_dX_T{}^2EI_{E0}\sin\delta_0\cos\delta_0}{k_1E\cos\delta_0 - k_2V_0} \\ &- \frac{k_2V_0^2\cos\delta_0}{k_1E\cos\delta_0 - k_2V_0}\Bigg)\end{aligned} \tag{7}$$

$$k_E = X_T^2E\,sin\,\delta_0\frac{X_dI_{E0} + E\sin\delta_0}{k_2V_0 - k_1E\cos\delta_0} \tag{8}$$

B. Mechanism Analysis in Single Machine systems

As the rapid increase of ES technology, there are more RE plants, which are equipped with utility-scale ES. The utility-scale ES can effectively reduce the energy curtailment and flexibly participate in the power market. Moreover, considering the expensive initial investment of ES, its function has been further developed to supply the ancillary service for power system oscillation damping. ES would affect the power system oscillation process by means of injecting or absorbing active current based on the electromagnetic power formulation of SG-ES. That is, the interactive timescale between the SG and ES is electromechanical level. Hence, the current control inner-loop of ES can be neglected. In addition, the charging or discharging active current quantity of ES can be controlled by the feed-back rotor speed, frequency, or rotor angle signal. However, the rotor angle is more difficult to measure than the rotor speed and frequency. The rotor speed is measured as the feed-back control signal to achieve mechanism analysis. In addition, it should be noted that feed-back frequency control also has equivalent effect, since the real power is highly correlated with the system frequency[23].

The classical second-order formula of SG is used to analyze the interaction mechanism as well as the dynamic characteristic between the SG and ES. The linearized rotor swing equation is expressed as a set of differential equations in per unit [7]:

$$\begin{cases} \dfrac{d\Delta\delta}{dt} = \omega_0 \Delta\omega_r \\ 2H\dfrac{d\Delta\omega_r}{dt} = \Delta P_m - \Delta P_e - D\Delta\omega_r \end{cases} \tag{9}$$

where rotor angle δ is in electrical radians, time t is in seconds, ω_0 is the base rotor electrical speed in radians per second, ω_r is the actual rotor speed in per unit, H is inertia constant, P_m is prime mover input mechanical power, P_e is output electromagnetic power, and D is damping coefficient.

To analyze the mechanism of ES mitigating the grid power oscillation, the PI controller is used to implement the control of the ES, and provide active reference current $I_{E,ref}$:

$$I_{E,ref} = \left(k_p + \frac{k_i}{s}\right)(\omega_r^* - \omega_r) \tag{10}$$

where s represents the *Laplace* operator, ω_r^* indicates the reference rotor speed.

Equation (10) shows if the feed-back speed of the generator is less than or greater than the rated synchronous rotor speed, the ES can generate or absorb real power to/from the utility grid to compensate the imbalance electromagnetic power. Thus, if the PI controller has a good performance, the reference rotor speed ω_r^* is followed by the actual rotor speed ω_r, and the power oscillation process is mitigated by the ES. Moreover, the actual output current I_E of ES is approximately equal to the inner-loop reference value $I_{E,ref}$ by ignoring the current loop respond time.

Linearizing (10) becomes:

$$\Delta I_E \approx \Delta I_{E,ref} = -\left(k_p + \frac{k_i}{s}\right)\Delta\omega_r \tag{11}$$

Considering $s\Delta\delta = \omega_0\Delta\omega_r$, (11) can be rearranged into:

$$\Delta I_E = -k_p\Delta\omega_r - \frac{k_i}{\omega_0}\Delta\delta \tag{12}$$

Substituting for (12) in (6), yields:

$$\Delta P_e = \left(k_g + k_E {}^{k_i}\!/_{\omega_0}\right)\Delta\delta + k_E k_p \Delta\omega_r \tag{13}$$

Furthermore, equation (9) can be rewritten as:

$$\begin{cases} \dfrac{d\Delta\delta}{dt} = \omega_0\Delta\omega_r \\ K_J\dfrac{d\Delta\omega_r}{dt} = \Delta T_m - K_S\Delta\delta - K_D\Delta\omega_r \end{cases} \tag{14}$$

where K_J is inertia time constant, T_m denotes mechanical torque, K_S is the synchronizing torque coefficient, and K_D is the damping coefficient.

Substituting for (13) in (9) and comparing with (14), yields:

$$\begin{cases} \Delta T_m = \Delta P_m \\ K_J = 2H \\ K_S = k_g + k_E {}^{k_i}\!/_{\omega_0} \\ K_D = D + k_E k_p \end{cases} \tag{15}$$

Equation (15) reveals that the PI controller's gains can directly affect the damping coefficient (affected by k_p) and synchronizing torque coefficient (affected by k_i), respectively.

C. *Mechanism Analysis in Multi-Machine systems*

In the multi-machine system, the ES can be installed near the generators. As shown in Fig. 3. P_{ei}, P_{Gi} and P_{Ei} are the electromagnetic power, network injection real power and the ES output power of the ith generator, respectively. The swing equation of multi-machine system can be simplified as follows:

$$\begin{cases} \dfrac{d\boldsymbol{\delta}}{dt} = \boldsymbol{\Omega_0}(\boldsymbol{\omega_r} - \omega_r^*) \\ \boldsymbol{2H}\dfrac{d\boldsymbol{\omega_r}}{dt} = \boldsymbol{P_m} - \boldsymbol{P_e} - \boldsymbol{D}(\boldsymbol{\omega_r} - \omega_r^*) \\ \boldsymbol{P_G} = \boldsymbol{P_e} + \boldsymbol{P_E} \end{cases} \tag{16}$$

where both $\boldsymbol{\delta}$, $\boldsymbol{\omega_r}$, $\boldsymbol{P_m}$, $\boldsymbol{P_e}$, $\boldsymbol{P_G}$, and $\boldsymbol{P_E}$ are n-dimensional column vectors; $\boldsymbol{\Omega_0}$, $\boldsymbol{H}$, and $\boldsymbol{D}$ are the diagonal matrix. Specifically, $\boldsymbol{P_G} = [P_{Gi}]_{n\times 1}$, P_{Gi} can be approximately calculated by [24]:

$$P_{Gi} = \sum_{j=1}^{n} |E_i||V_j||Y_{ij}|\cos(\varphi_{ij} - \delta_i + \vartheta_j) \tag{17}$$

where $|V_j|$ and ϑ_j are the voltage and phase angle at jth bus; $|Y_{ij}|$ and φ_{ij} are modulus and phase angle of the admittance, respectively. In addition, $\boldsymbol{P_E} = [P_{Ei}]_{n\times 1}$ can be controlled by rotor speed signal of corresponding generator by PI controller:

$$P_{Ei} = \left(k_{pi} + \frac{k_{ii}}{s}\right)(\omega_r^* - \omega_r) \tag{18}$$

where k_{pi} and k_{ii} are the gains of ith PI controller.

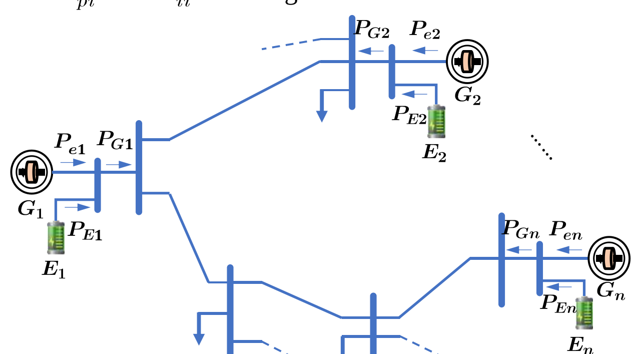


Fig. 3. Multi-machine system illustration with ES integration.

The linearized expression of (16)-(18) around the fixed point can be derived:

$$\begin{cases} \dfrac{d\Delta\boldsymbol{\delta}}{dt} = \boldsymbol{\Omega_0}\Delta\boldsymbol{\omega_r} \\ \boldsymbol{2H}\dfrac{d\Delta\boldsymbol{\omega_r}}{dt} = \Delta\boldsymbol{P_m} - \Delta\boldsymbol{P_G} + \Delta\boldsymbol{P_E} - \boldsymbol{D}\Delta\boldsymbol{\omega_r} \end{cases} \tag{19}$$

Specifically,

$$\begin{cases} \Delta\boldsymbol{P_G} = \left[\dfrac{\boldsymbol{\partial} P_{Gi}}{\boldsymbol{\partial}\delta_i}\right]_{n\times n}\Delta\boldsymbol{\delta} = \boldsymbol{J_P}\Delta\boldsymbol{\delta} \\ \Delta\boldsymbol{P_E} = -\boldsymbol{K_p}\Delta\boldsymbol{\omega_r} - \boldsymbol{K_i}\boldsymbol{\Omega_0}^{-\mathbf{1}}\Delta\boldsymbol{\delta} \end{cases} \tag{20}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

where $\boldsymbol{J_P}$ is the *Jacobian* matrix; $\boldsymbol{K_p}$ and $\boldsymbol{K_i}$ are the diagonal matrix for the PI controller parameters.

Substituting for (20) into (19), yields:

$$\boldsymbol{2H}\frac{d\Delta\boldsymbol{\omega_r}}{dt} = \Delta\boldsymbol{P_m} - (\boldsymbol{J_P} + \boldsymbol{K_i}\boldsymbol{\Omega_0}^{-1})\Delta\boldsymbol{\delta} - (\boldsymbol{D} + \boldsymbol{K_p})\Delta\boldsymbol{\omega_r} \quad (21)$$

Comparing with (14), results in:

$$\begin{cases}\Delta\boldsymbol{T_m} = \Delta\boldsymbol{P_m}\\ \boldsymbol{K_J} = \boldsymbol{2H}\\ \boldsymbol{K_S} = \boldsymbol{J_P} + \boldsymbol{K_i}\boldsymbol{\Omega_0}^{-1}\\ \boldsymbol{K_D} = \boldsymbol{D} + \boldsymbol{K_p}\end{cases} \quad (22)$$

Similarly, it can conclude that the ES controlled by PI controller in multi-machine system can equivalently affect synchronizing torque and damping coefficient, respectively.

The essence of ES to suppress power oscillation is that it can provide appropriate active power support through PI controller to the power system in time. That is, the cost of ES to suppress power oscillation is its own energy deviation (ΔE_{es}) after participating in the suppression of power oscillation. The greater the energy deviation of ES, the easier it is to cause the deep charging or discharging. However, the ES controlled by traditional PI controller cannot ensure that the energy deviation equals to zero in the steady state of power oscillation control (Equation (A-1) -(A-8) in Appendix show the quantitative analysis).

In order to reduce the energy deviation while retaining the high performance of the PI controller, a novel PI-IR is proposed in Section II-D to damp power oscillation of SG and reduce energy deviation of ES

D. Controller Design

Inspired by aforementioned issue, a novel PI-IR controller is designed to implement oscillation suppression, while making sure that the energy deviation equals to zero at the end of an oscillation process. The structure of the PI-IR controller is shown in Fig. 4. Additionally, the transfer function $G_\omega(s)$ can be derived:

$$G_\omega(s) = \frac{sk_p + k_i}{s + k_i k_b} \quad (23)$$

where the k_p, k_i and k_b ($k_b \neq 0$) are the controller gains.

Compared with the traditional PI controller, the proposed PI-IR controller can eliminate the accumulative integral error by the additional negative feed-back loop to ensure that the energy deviation of ES equals to zero in the steady-state of power oscillation control. Equation (A-9) -(A-12) in Appendix show that why the integral reduction can suppress the energy deviation.

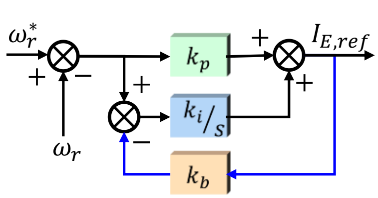


Fig. 4. The structure of the PI-IR controller.

The mechanism of the ES with the proposed PI-IR controller to suppress the power oscillation is analyzed by referring the derivation process of (11) -(15), yields:

$$\begin{cases}\Delta T_m = \Delta P_m - k_E k_i k_b \Delta SoC\\ K_J = 2H\\ K_S = k_g + k_E{}^{k_i}/_{\omega_0}\\ K_D = D + k_E k_p\end{cases} \quad (24)$$

Equation (24) reveals that the PI-IR controller's gains can directly affect the damping coefficient (affected by k_p), synchronizing torque coefficient (affected by k_i), and mechanical torque (affected by k_i and k_b), respectively. That is, when it comes to feed-back rotor speed as well as ES controlled by PI-IR controller, the synchronous and oscillation damping ability of the synchronous power system will be equivalently dominated. In addition, they are also affected by the power system structure parameters and the steady-state operating point. However, adjusting the PI-IR controller parameters is the most significant and flexible way to dynamically change the synchronous and damping feature.

III. Problem and Approach Formulation

Although the optimal and fixed controller parameters can be obtained based on the classical control theory at a certain grid structure and steady-state operating point in the traditional control field, the steady-state operating point would be changed in real-time due to all kinds of uncertainties. This situation would be unfavorable for the performance of the controller with fixed parameters. Hence, real-time tuning method is introduced in this section.

A. Problem Formulation

To ensure the response performance of the SG-ES integrated system for any possible disturbance, the optimal PI-IR parameters tuning can be converted to an online making-decision problem in uncertainty environment. The finite MDP is applied to reformulate the PI-IR controller parameters tuning problem. MDP is an important way to build the reinforcement learning (RL) framework, which describes that the next state of the system is not only related to the current state, but also associated with the current action taken[25]. Specifically, there are five crucial elements in the MDP that make up a tuple, namely <s_t, a_t, p_t, r_t, γ>.

1) State: The state at time step t (the meaning of a time step is an oscillation process in this research) is defined as a vector, $s_t = (t_o, E_{es})$. t_o denotes the oscillation duration; E_{es} indicates the energy consumption or accumulation for one oscillation process.

$$E_{es} = \int_{t_0}^{Ts} p_{es}(\tau)d\tau \quad (25)$$

where t_0 is the initial time, Ts is the time-domain simulation time, and p_{es} is the output active power.

2) Action: the action a_t represents the PI-IR parameters for given the state s_t. The parameters are constrained as follows:

$$\begin{cases} k_{p,min} \le k_p \le k_{p,max} \\ k_{i,min} \le k_i \le k_{i,max} \\ k_{b,min} \le k_b \le k_{b,max} \end{cases} \tag{26}$$

where the $k_{x,min}$ and $k_{x,max}$ $(x = p, i, b)$ are the allowed minimum and maximum value, respectively.

3) State transition: The state transition probability function is related to the action and the randomness of the system, and the state transition process from s_t to s_{t+1} can be represented as $s_{t+1} \sim p_t(s_t, a_t, \varepsilon_t)$. The state transition process is subject to uncertainty since the oscillation duration and energy variation are unknown. Besides, because the randomness ε_t is closely linked with many factors, it is an intractable task to find an accurate distribution model. In this paper, in order to better describe this uncertainty, a model-free approach is introduced to learn the state transition like shown in Section III-B.

4) Reward: the reward r_t should be carefully designed to accurately evaluate the action-value for each time step t. The multimodal reward is defined as:

$$r_t = -\int_{t_0}^{Ts} \tau |\Delta\omega_r(\tau)| d\tau - \int_{t_0}^{Ts} \frac{1}{\tau} |p_{es}(\tau)| d\tau - \varphi |E_{es}| \tag{27}$$

where φ is a penalty coefficient to ensure the minimum energy variation of ES.

5) Action-value function: At each step, the agent takes the state s_t as input and outputs an action a_t based on the policy $\pi(s_t|a_t)$. The performance of the action under the given state is evaluated by the expected accumulative reward for one trajectory, which is represented as [26]:

$$Q^\pi\ s_t, a_t\ = \mathbb{E}_\pi\left[R_t + \gamma \mathbb{E}_{a_{t+1}\sim\pi}[Q^\pi\ s_{t+1}, a_{t+1}\ ]\right] \tag{28}$$

where

$$R_t = r_t + \gamma[r_{t+1} + \gamma^1 r_{t+2} \cdots + \gamma^{T-t-1} r_T] \tag{29}$$

where $Q^\pi(s_t, a_t)$ is named action-value function, T is the finite MDP steps, and $\gamma \in [0{,}1]$ is a discount factor, which is utilized to balance the importance of current and future reward.

Thus, the objective of the real-time parameters tuning problem is to obtain optimal policy π^* to maximize the action-value function.

B. SAC Algorithm for PI-IR Real-time Control

In this paper, an innovative DRL algorithm, SAC, is employed to solve the PI-IR real-time control problem, which is formulated as MDP. Since new samples are required at each gradient step, A3C [22] and PPO [27], [28] as commonly used DRL algorithms, are of a notoriously low sampling efficiency. Although DDPG [19], as an off-policy algorithm, is proposed to improve the utilization of samples, it is too sensitive to the hyper-parameters in the training process, which will also have a negative impact on the training efficiency. To address these drawbacks, the off-policy maximum-entropy deep RL algorithm, SAC, is developed to provide a robust and sample-efficient training performance. SAC algorithm incorporates three key tricks: Actor-Critic (AC) architecture (two DNNs approximate policy and value function, respectively), entropy maximization, which enables to guarantee stability and encourage exploration, and off-policy to improve the utilization efficiency of samples.

1) Actor-Critic Method: AC method is the distinct framework of the proposed SAC algorithm. Two DNNs are established in AC method, named Actor and Critic, for policy estimation and policy improvement. At each iteration, the Actor μ, parameterized by θ^μ, is employed to generate a next-state action $\mu(s_{t+1} \mid \theta^\mu)$ based on the current policy function; then, the Critic performs the policy evaluation task to estimate Q-values $Q(s_t, a_t \mid \theta^Q)$ of corresponding actions; and temporal difference (TD) learning is to guarantee the estimation accuracy and update mechanism of the Critic simultaneously by minimizing the following loss function L^Q , which is presented as follows[29]:

$$\begin{gathered} \Delta Q_t = r_t + \gamma Q(s_{t+1}, \mu(s_{t+1} \mid \theta^\mu) \mid \theta^Q) - Q(s_t, a_t \mid \theta^Q) \\ L^Q(s_t, a_t \mid \theta^Q) = (\Delta Q_t)^2 \end{gathered} \tag{30}$$

where ΔQ_t is TD error at time step t; in order to achieve higher Q-value, gradient $\nabla_a Q(s_t, a_t \mid \theta^Q)$ is employed in the Critic; besides, $\nabla_{\theta^\mu}\mu(s \mid \theta^\mu)|_{s_t}$ indicates how the parameters of the Actor μ affect the direction of action selection; at last, as shown in (31), the resulting gradient $\nabla_{\theta^\mu}\mu$, which contains the value function information, is provided as reference for the Actor parameters update and action selection [19].

$$\nabla_{\theta^\mu}\mu|s_t = \nabla_a Q(s, a \mid \theta^Q)|_{s_t, \mu(s_t)} \nabla_{\theta^\mu}\mu(s \mid \theta^\mu)|_{s_t} \tag{31}$$

2) Maximum-Entropy RL Framework: Generally, the objective of RL is to learn a policy to maximize the expected value of a cumulative reward, which is presented in (32).

$$\pi^* = \underset{\pi}{argmax} \mathbb{E}\left\{\sum_{t=0}^{\infty} \gamma^t r_t\right\} \tag{32}$$

However, for the advantages of encouraging exploration and learning more near-optimal behaviors to improve learning speed, maximum-entropy RL framework has a comprehensive learning target. Specifically, in addition to obtain higher cumulative rewards, it also requires an entropy term $H(\pi(a_t \mid s_t)) = -\log \pi(a \mid s)$ of each output action of the policy. The optimization objective formula can be transformed into:

$$J(\pi) = \mathbb{E}\left\{\sum_{t=0}^{\infty} \gamma^t [r_t - \alpha \log\ \pi(a_t \mid s_t)] \mid \pi\right\} \tag{33}$$

where the parameter α is a temperature coefficient [30], which is used to control whether the agent's target is to focus on reward or entropy.

3) Soft Actor-Critic Algorithm: Based on the above theory, soft policy iteration can converge in the case of tabular. Because the approximation method based on DNNs can deal with the high-dimensional problems, which cannot be solved by the tabular, it can be better to apply DNNs to approximate soft Q-function and policy. Specifically, one network $Q_\theta(s,a)$ parameterized by θ approximates soft Q-function, and another $\pi_\phi(\cdot|\ s)$ network parameterized by ϕ learns the mean and covariance of Gaussian distribution policy.

As expressed in (34), the update mechanism of soft Q-function (parameter θ) is the same as Q-learning, which updates Bellman residuals [31], except that the value function here contains the entropy item.

$$\begin{aligned} J_Q(\theta) &= \mathop{\mathbb{E}}_{\substack{s_t,a_t,s_{t+1}\sim M \\ a_{t+1}\sim\pi_\phi}}\left[\frac{1}{2}\left(Q_\theta(s_t,a_t)-y^2\right)\right] \\ y &= r(s_t,a_t)-\gamma[Q_\theta(s_{t+1},a_{t+1}) \\ &\quad -\alpha\log\ \pi_\phi(a_{t+1}\mid s_{t+1})] \\ &\theta \leftarrow \theta-\eta_c\nabla_\theta J_Q(\theta) \end{aligned} \tag{34}$$

Note that, a target critic network parameterized by $\bar{\theta}$ is utilized to improve the training stability. When training $Q_\theta(s_t,a_t)$, (s_t,a_t) is extracted from the replay buffer, but a_{t+1} is temporarily collected from policy π_ϕ during training. The policy network $\pi_\phi(\cdot|\ s)$ (parameter ϕ) is updated by minimizing the Kullback-Leibler (KL) divergence [32], which is shown in (35). The details of the SAC algorithm are presented in Table I.

$$\begin{aligned} J_\pi(\phi) &= \mathop{\mathbb{E}}_{s_t\sim M}\left[\mathop{\mathbb{E}}_{a_t\sim\pi_t}[\alpha\log\pi_\phi(a_t|s_t)-Q_\theta(s_t,a_t)]\right] \\ &\phi \leftarrow \phi-\eta_a\nabla_\phi J_\pi(\phi) \end{aligned} \tag{35}$$

TABLE I
THE FLOW OF SOFT ACTOR-CRITIC BASED PARAMETERS TUNING ALGORITHM

```
// Starting Training
Initialize network weights θ, θ̄, φ
For each episode do
  // Generating training data
  For each time step in the environment do
    Choose action a_t based on π_φ(·| s)
    Execute a_t, and observe r_t, s_{t+1}
    Store (s_t, a_t, r_t, s_{t+1}) in replay buffer M
    End For
  // Training neural networks
    For each gradient step do:
      Uniformly sample m batches from replay buffer
      Update soft Q-value function based on (34)
      Update the actor network according to (35)
      Update target network weights according to
                    θ̄ ← τθ + (1 − τ)θ̄
    End For
End
```

IV. CASE STUDY

In this section, the authors aim to evaluate the proposed real-time tuning approach on multiple case studies and illustrate its performance through the time-domain simulation analysis. The details about the time-domain simulation setup are given in Section IV-A. The superiority of the designed PI-IR controller is validated in Section IV-B. Then, the training process is presented in Section IV-C where only the wind power fluctuation is considered as the unknown disturbance. Finally, the generalization of the well-trained agent is tested on single and multiple disturbances in IV-D.

A. Simulation Setup

The simulation is carried out on a modified four-machine two-area system, which consists of two areas linked two 120 kV of 110 km length. The simplified single line diagram is shown in Fig. 5. Although small in scale, it closely mimics the behavior of the typical power systems in practice operation. There are three identical round rotor generators rated 13.8 kV/150 MVA, two of them in Area #1, and an external equivalent grid rated 120 kV/200 MVA; the Area #2 is equipped with one round rotor generators, an ES rated 10 MW/1 kWh. Besides, a 25 MW wind farm is connected into the Area #2. The load in Area #1 is represented as constant impedances.

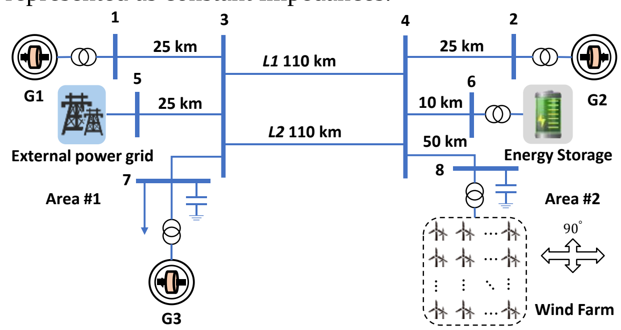


Fig. 5. The simplified single line diagram of the test system.

The performance of the designed controller and approach are evaluated under the time-domain simulation with a real-world scenario. The real-world hourly wind power data for one year is downloaded from *Energinet* [33] as the disturbance. The 7 days test data are randomly sampled from the raw data to act as test data set, and the residual data are applied for training. The hyper-parameters of the training algorithm are shown in Table II. Besides, the structure of the Actor and Critic network are predetermined; the neurons of input and output layer are equal to the number of states and actions, respectively; the number of the neurons in its three hidden layers are 64, 128 and 64.

TABLE II
HYPER-PARAMETERS SETTING FOR TRAINING SAC ALGORITHM

Name	Value
Replay size M	48 000
Discount factor γ	0.9
Mini-batch size m	64
Training eposides N	7 000
Step in each episode n	24
Soft update coefficient τ	5×10^{-3}
Learning rate for value learning η_c	2×10^{-3}
Learning rate for policy learning η_a	2×10^{-4}
Entropy regularization coefficient α	0.2

B. Effectiveness Verification

One of the contributions of this paper is to design a controller suitable for ES to suppress power oscillation while ensuring that the energy deviation of ES is zero at the end of the oscillation. Hence, the effectiveness of the designed PI-IR controller is evaluated by comparing with several traditional controller, including P and PI controller. Considering the wind power fluctuation as the disturbance, Fig. 6 shows that the energy deviation curve of ES corresponding to different controllers. This time-domain simulation shows that both the P and PI controller cannot guarantee that the energy deviation is zero at the end of the oscillation. However, the proposed PI-IR controller would always make sure the energy deviation is zero. The advantage of the proposed PI-IR controller is that it can reduce the investment cost of ES at the planning stage while making it an additional damping controller without affecting its participation in the economic dispatch at the operation stage.

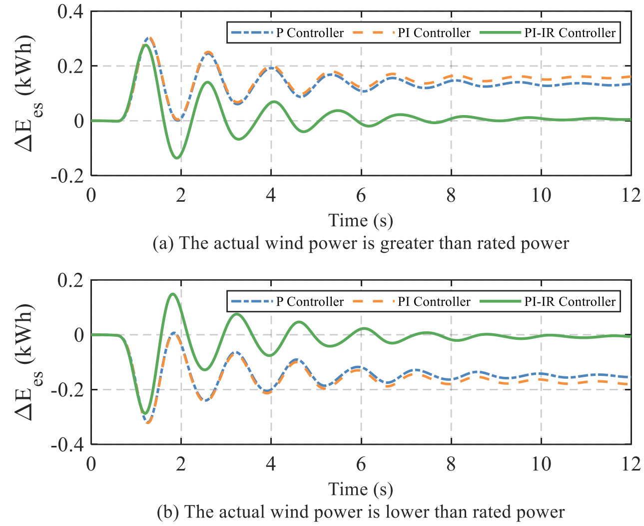


Fig. 6. Energy deviation curve of ES corresponding to different controllers.

C. Training Process of SAC Agent

The training and simulation are implemented in *Tensorflow* 1.8.0 with *Python* 3.6.5 and *MATLAB*, the hardware is a 64-bit computer with one Intel(R) Core(TM) i9-9820X CPU @ 3.30GHz. During the training process, regarding history wind power fluctuation as the disturbance, the agent obtains current state s_t from the test system and then returns action a_t based on the policy π. After executing action a_t, the reward r_t and next state s_{t+1} can be received. As the training episodes increase, the state-action mapping become more accurate.

As a comparison, the double deep Q network (DDQN)–based PI-IR parameters tuning mechanism is also performed. The DDQN is similar with SAC, and it is also the real-time strategy based on the state information (see the Ref. [34] for the detailed algorithm). The cumulative reward over 7000 episodes is shown in Fig. 7.

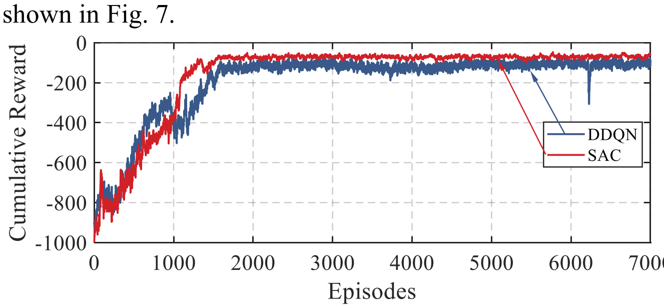


Fig. 7. Cumulative reward per episode during the training process.

The reward of both the DDQN and SAC are lower in the initial stage. However, as the better experiences are stored in the replay buffer, the cumulative reward begins to increase gradually, and then it converges to around -70.29 for SAC agent with small oscillation. For DDQN, the reward nearly has same convergence speed as SAC, it finally converges around -108.48. However, the training curve of DDQN fluctuates greatly, and its steady-state reward value is lower than that of SAC. The reward curve demonstrates that the proposed SAC approach can learn the policy more effectively to maximize the cumulative reward than then DDQN method.

D. Online Application and Performance Evaluation

In this part, the proposed SAC approach is evaluated and compared with several commonly solutions, including the DDQN and linear model predictive control (MPC, more details can be found in [35]) method. Taking the negative reward function as the cost function of the MPC method, the MPC–based PI-IR parameters tuning method can obtain the optimal parameters through real-time rolling calculation. Moreover, to evaluate the generalization of the well-trained agent, two cases are introduced.

Case 1: Taking the wind power fluctuation as the disturbance.

Case 2: On the basis of Case 1, the load on Bus 7 increases by 0.1 p.u. at 4s.

So far, three methods and two cases are proposed, and applied to the test system by time-domain simulation. The mode analysis of Case 1 is carried out by *Prony* identification. The eigenvalues of the dominant mode and the probability density function (PDF) of the corresponding damping ratio are plotted in Fig. 8. It can be seen from the Fig. 8 that all the eigenvalues are located in the left plane of the real axis under the three methods. However, it is clearly shown that the linear MPC–based method has the lowest stability margin, followed by DDQN, and SAC has the highest stability margin. The reason is that it is difficult for the PI-IR controller of the linear MPC–based tuning method to adapt the significantly changing of the wind power; note that the policy formulated by DDQN–based real-time parameters tuning method is discrete, that is, the action cannot change continuously for each state, which causes the information loss. Thus, this drawback makes it easy to fall into a local optimum; compared with DDQN, SAC is a continuous stochastic policy and developed for maximum entropy reinforcement learning, which is, the agent not only learns one way to solve the oscillation, but all possibilities.

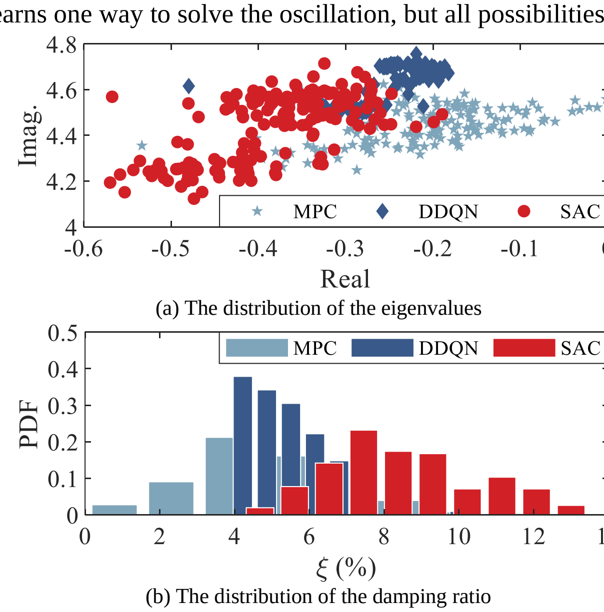
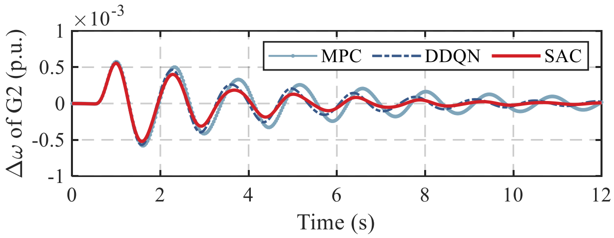


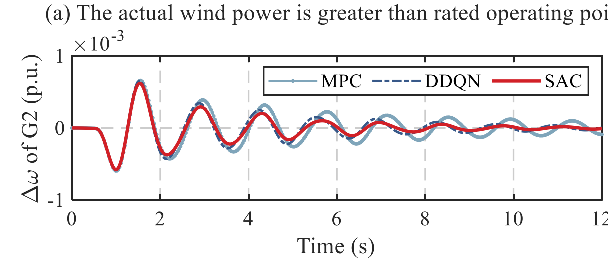
Fig. 8. Distribution of the eigenvalues of the dominant mode and the PDF of the damping ratio.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

To evaluate the transient process of the speed oscillations in details, the time-domain simulation curve is extracted. For case 1, the average of the rotor speed deviation of G2 is presented in Fig. 9. It is clearly shown that all three methods can effectively mitigate the rotor oscillation. Moreover, comparing with the linear MPC and DDQN –based method, the SAC method has a faster speed to mitigate the oscillation, and the rotor speed deviation is completely suppressed after 10s.



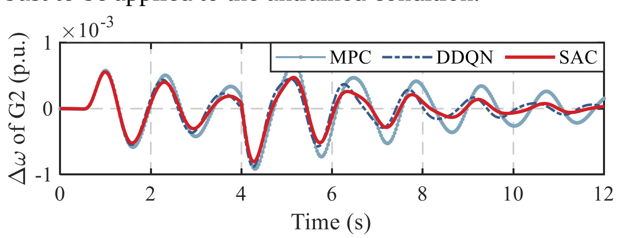
(a) The actual wind power is greater than rated operating point



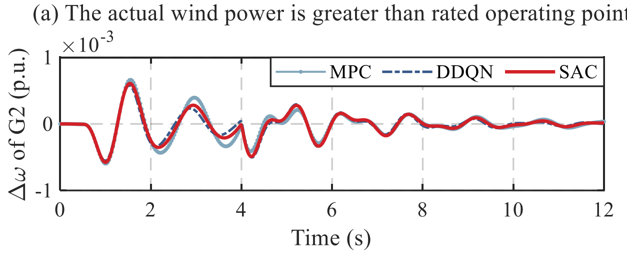
(b) The actual wind power is lower than rated operating point

Fig. 9. Rotor speed deviation of G2 in Case 1.

The adaptability comparison of the three methods is further done with the multiple disturbances in Case 2, and the oscillation curves are shown in Fig. 10. Under the continuous disturbances condition, the test system with the SAC–based agent reaches the steady-state in the shortest time. These experimental results reveal that the SAC–based agent is more robust to be applied to the untrained condition.



(a) The actual wind power is greater than rated operating point



(b) The actual wind power is lower than rated operating point

Fig. 10. Rotor speed deviation of G2 in Case 2.

Specifically, the average action results are listed in Table III for two cases. The load changing will also aggravate the power oscillation of the power system, so the agent will enhance the adjustable ability of the controller to further mitigate the power oscillation. This is why the k_p and k_i parameters value in Case 2 are greater than that in Case 1. It is worthing noting that, with the load demand increasing, the function of the additional integral reduction loop will be decreased to enhance the synchronization ability of the generator.

TABLE III
THE AVERAGE ACTION RESULTS

	k_p	k_i	k_b
Case 1	433.701	11.042	0.0811
Case 2	464.618	33.929	0.0438

Moreover, in order to intuitively show the quantitative analysis of the mentioned three methods, the mean value and standard deviation of power oscillation modes of the test system under the test data set is shown in Table IV. It can conclude that, it is difficult for the linear MPC–based tuning method to adapt the significantly changing of the wind power, so that the power system has a lower stability margin. However, the SAC-based agent has learned the probability distribution characteristic of the wind power, and can deal with the arbitrary oscillation caused by wind power to make the power oscillation mode move more to the left side of the complex plane, to ensure that the power system has a larger stability margin.

TABLE IV
THE COMPARISON OF POWER OSCILLATION MODES

Method	Real part	Imag. part	Damping (%)	Standard deviation of damping
MPC	-0.21	4.45	4.71	1.97
DDQN	-0.25	4.62	5.40	1.07
SAC	-0.38	4.44	8.53	1.98

V. CONCLUSION AND FUTURE WORK

In this paper, the mechanism of ES damper suppressing the grid power oscillation damping via small signal and damping torque coefficient method is analyzed. Then, a novel PI-IR controller is proposed to mitigate power oscillation while ensuring that the energy deviation equals to zero in the steady state of power oscillation, and the controller parameters tuning problem is formulated as a MDP with unknown transition probability. In this tuning process, the fluctuated wind power is only used as a disturbance signal to cause the power oscillation in power system, and the SAC–based model-free method is employed to determine the near-optimal real-time parameters. Finally, the modified four-machine two-area system with ES and wind farm is used as the test system, and then the effectiveness of the proposed PI-IR controller has been compared with the conventional P and PI controller in the test system, and the results show that only the proposed PI-IR controller can make sure that the energy variation of ES is zero while effectively suppressing the power oscillation. In addition, the test system has also been applied for comparative study of MPC–based tuning, DDQN–based tuning and SAC–based tuning. The time-domain simulation results under trained and untrained disturbances demonstrated that the performance of the SAC method is better than the other two compared methods. In current research, the load changing is not considered during the training phase, to this end, the multiple disturbance will be considered to train the agent in our future work.

APPENDIX

Proof: Let k_p, k_i, and k_b be the constant. In order to facilitate theoretical analysis, the variation of the state of charge

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

(ΔSoC) is used instead of energy deviation (ΔE_{es}) of ES.

The current variation ΔI_E of the ES with PI controller can be formulated:

$$\Delta I_E(s) = (k_p + \frac{k_i}{s})\Delta\omega_r(s) \tag{A-1}$$

The variation of the state of charge (ΔSoC) is defined as:

$$\begin{aligned}\Delta SoC(s) &= \frac{1}{s}\Delta I_E(s) = \frac{1}{s}(k_p + \frac{k_i}{s})\Delta\omega_r(s) \\ &= G(s)\Delta\omega_r(s)\end{aligned} \tag{A-2}$$

The inverse *Laplace* transform of $G(s)$ can be calculated:

$$\mathcal{L}^{-1}[G(s)] = k_p + tk_i \tag{A-3}$$

Based on *Prony* analysis, the time-domain form of rotor speed deviation $\Delta\omega_r(t)$ is as follows:

$$\begin{aligned}\Delta\omega_r(t) &= \mathcal{L}^{-1}[\Delta\omega_r(s)] \\ &= \sum_{n=1}^{N} A_n e^{\alpha_n t}\cos(\theta_n + 2\pi f_n t)\end{aligned} \tag{A-4}$$

where N indicates the number of dominant oscillation signals, A_n represents the amplitude of the signal component n, α_n is the decay factor of the signal component n, θ_n is the initial phase of the signal component n, f_n is the undamped natural frequency of the signal component n.

$$\begin{aligned}\Delta SoC(t) &= \mathcal{L}^{-1}\left[\frac{1}{s}\Delta I_E(s)\right] \\ &= (k_p + tk_i)\sum_{n=1}^{N} A_n e^{\alpha_n t}\cos(\theta_n \\ &\quad + 2\pi f_n t)\end{aligned} \tag{A-5}$$

In (A-5), when $\alpha_n < 0, \forall n$ and the oscillation duration time $T_s \geq {-5}/{\max\{\alpha_1, \alpha_2, \dots, \alpha_n\}}$:

$$\begin{aligned}\lim_{t\to T_s}\sum_{n=1}^{N} k_p A_n e^{\alpha_n t}\cos(\theta_n + 2\pi f_n t) = 0, \forall\alpha_i, (i \\ = 1,2, \dots, n)\end{aligned} \tag{A-6}$$

$$\begin{aligned}\lim_{t\to T_s}\sum_{n=1}^{N} tk_i A_n e^{\alpha_n t}\cos(\theta_n + 2\pi f_n t) \neq 0, \exists\alpha_i, (i \\ = 1,2, \dots, n)\end{aligned} \tag{A-7}$$

Hence $\exists\alpha_i, (i = 1,2, \dots, n)$, we would get:

$$\begin{cases}\lim_{t\to T_s}\Delta\omega_r(t) = 0 \\ \lim_{t\to T_s}\Delta SoC(t) \neq 0\end{cases} \tag{A-8}$$

Equation (A-2) -(A-8) show that the ES with PI control strategy can suppress the power oscillation without ensuring that the energy deviation of ES is zero.

Similarly, the ΔSoC of the ES with PI-IR controller can be formulated:

$$\begin{aligned}\Delta SoC(s) &= \frac{1}{s}\Delta I_E(s) = \frac{1}{s}(\frac{sk_p + k_i}{s + k_i k_b})\Delta\omega_r(s) \\ &= G(\mathrm{s})\Delta\omega_r(s)\end{aligned} \tag{A-9}$$

The inverse *Laplace* transform of $G(s)$ is ($k_b \neq 0$ and $k_i k_b > 0$):

$$\mathcal{L}^{-1}[G(\mathrm{s})] = \frac{1}{k_b} + (k_p - \frac{1}{k_b})e^{-k_i k_b t} \tag{A-10}$$

$$\begin{aligned}\Delta SoC(t) &= \sum_{n=1}^{N}\frac{1}{k_b}A_n e^{\alpha_n t}\cos(\theta_n + 2\pi f_n t) \\ &+ \sum_{n=1}^{N}(k_p - \frac{1}{k_b})A_n e^{(\alpha_n - k_i k_b)t}\cos(\theta_n + 2\pi f_n t)\end{aligned} \tag{A-11}$$

Due to $\alpha_n < 0, \forall n$ and $k_i k_b > 0$, $\alpha_n - k_i k_b < 0$. To this end,

$$\begin{cases}\lim_{t\to T_s}\Delta\omega_r(t) = 0 \\ \lim_{t\to T_s}\Delta SoC(t) = 0\end{cases} \tag{A-12}$$

Hence, the ES with PI-IR controller can mitigate the power oscillation and make sure zero energy deviation of ES at the same time.

REFERENCES

[1] X. Li, D. Hui and X. Lai, "Battery Energy Storage Station (BESS)-Based Smoothing Control of Photovoltaic (PV) and Wind Power Generation Fluctuations," *IEEE Trans. Sustain. Energy*, vol. 4, no. 2, pp. 464-473, April 2013.

[2] X. Xie, W. Liu, H. Liu, et al., "A System-Wide Protection Against Unstable SSCI in Series-Compensated Wind Power Systems," *IEEE Trans. Power Del.*, vol. 33, no. 6, pp. 3095-3104, Dec. 2018.

[3] Y. Gu, J. Liu, T. C. Green, et. al., "Motion-Induction Compensation to Mitigate Sub-Synchronous Oscillation in Wind Farms," *IEEE Trans. Sustain. Energy*, vol. 11, no. 3, pp. 1247-1256, July 2020.

[4] H. Nazaripouya, H. R. Pota, C. Chu, et. al., "Real-Time Model-Free Coordination of Active and Reactive Powers of Distributed Energy Resources to Improve Voltage Regulation in Distribution Systems," *IEEE Trans. Sustain. Energy*, vol. 11, no. 3, pp. 1483-1494, July 2020.

[5] M. Ayar, S. Obuz, R. D. Trevizan, et al., "A Distributed Control Approach for Enhancing Smart Grid Transient Stability and Resilience," *IEEE Trans. Smart Grid*, vol. 8, no. 6, pp. 3035-3044, Nov. 2017.

[6] L. S. Xiong, F. Zhou and F. Wang et al., "Static Synchronous Generator Model: A New Perspective to Investigate Dynamic Characteristics and Stability Issues of Grid-Tied PWM Inverter," *IEEE Trans. Power Electron*, vol. 31, no. 9, pp. 6264-6280, Sept. 2016.

[7] P. Kundur, Power System Stability and Control. New York, NY, USA: McGraw-Hill, 1994.

[8] A. Chakrabortty, "Wide-Area Damping Control of Power Systems Using Dynamic Clustering and TCSC-Based Redesigns," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1503-1514, Sept. 2012.

[9] K. Bruninx, Y. Dvorkin, and E. Delarue et al., "Coupling Pumped Hydro Energy Storage with Unit Commitment," *IEEE Trans. Sustain. Energy*, vol. 7, no. 2, pp. 786-796, April 2016.

[10] Y. Zhu, C. Liu, K. Sun, D. Shi et al., "Optimization of Battery Energy Storage to Improve Power System Oscillation Damping," *IEEE Trans. Sustain. Energy*, vol. 10, no. 3, pp. 1015-1024, July 2019.

[11] S. Panda, "Multi-objective PID controller tuning for a FACTS-based damping stabilizer using Non-dominated Sorting Genetic Algorithm-II," *Int. J. Electr. Power Energy Syst*. 2011; 33:1296–308.

[12] X. Sui, Y. Tang, H. He et al., "Energy-Storage-Based Low-Frequency Oscillation Damping Control Using Particle Swarm Optimization and Heuristic Dynamic Programming," *IEEE Trans. Power Syst.*, vol. 29, no. 5, pp. 2539-2548, Sept. 2014.

[13] M. Beza and M. Bongiorno, "An Adaptive Power Oscillation Damping Controller by STATCOM With Energy Storage," *IEEE Trans. Power Syst.*, vol. 30, no. 1, pp. 484-493, Jan. 2015.

[14] M. J. Morshed and A. Fekih, "A Probabilistic Robust Coordinated Approach to Stabilize Power Oscillations in DFIG-Based Power Systems," *IEEE Trans. Ind. Inform.*, vol. 15, no. 10, pp. 5599-5612, Oct. 2019.

[15] A. S. Mir, S. Bhasin and N. Senroy, "Decentralized Nonlinear Adaptive Optimal Control Scheme for Enhancement of Power System Stability," *IEEE Trans. Power Syst*., vol. 35, no. 2, pp. 1400-1410, Mar. 2020.

[16] T. K. Chau, S. S. Yu, T. Fernando, et al., "A Load-Forecasting-Based Adaptive Parameter Optimization Strategy of STATCOM Using ANNs for Enhancement of LFOD in Power Systems," *IEEE Trans. Ind. Inform.*, vol. 14, no. 6, pp. 2463-2472, June 2018.

[17] Y. Guo and H. Gao, "Data-Driven Online System Equivalent for Self-Adaptive Droop Voltage Control of Wind Power Plants," *IEEE Trans. Energy Conver.*, vol. 35, no. 1, pp. 302-305, March 2020.

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

[19] G. Z. Zhang, W. H. Hu et al. "A data-driven approach for designing STATCOM additional damping controller for wind farms," *Int. J. Electr. Power Energy Syst.*, vol. 117, 2020, pp. 1-13.

[20] C. Chen, M. Cui, F. F. Li, et al., "Model-Free Emergency Frequency Control Based on Reinforcement Learning," *IEEE Trans. Ind. Inform.*, early access.

[21] Y. Hashmy, Z. Yu, D. Shi et al., "Wide-area Measurement System-based Low Frequency Oscillation Damping Control through Reinforcement Learning," *IEEE Trans. Smart Grid,* early access.

[22] G. Z. Zhang, W. H. Hu et al. "Deep Reinforcement Learning Based Approach for Proportional Resonance Power System Stabilizer to Prevent Ultra-Low-Frequency Oscillations," *IEEE Trans. Smart Grid,* vol. 11, no. 6, pp. 5260-5272, Nov. 2020.

[23] Anderson PM (2002) Power system control and stability. WileyIEEE, Piscataway.

[24] H. Saadat, Power System Analysis. New York, NY, USA: McGraw-Hill, 1999.

[25] Olivier S, Olivier B. Markov Decision Process in Artificial Intelligence, Hoboken, NJ USA: John Wiley & Sons, Inc, 2013.

[26] D. Cao, W. H. Hu et al "Bidding strategy for trading wind energy and purchasing reserve of wind power producer – A DRL based approach", *Int. J. Electr. Power Energy Syst.*, 117, 2020, pp. 1-10.

[27] B. Zhang, W. H. Hu, D. Cao, et.al. "Deep Reinforcement Learning–based Approach for Optimizing Energy Conversion in Integrated Electrical and Heating System with Renewable Energy." *Energy Convers Manage*, vol. 202, no.15 Dec. 2019.

[28] D. Cao, W. Hu, J. Zhao, et al. "Reinforcement learning and its applications in modern power and energy systems: a review." *J. M. Power Syst. Clean Energy*, vol. 8, no. 6, pp. 1029-1042, 2020.

[29] B. Zhang, W. H. Hu, J. H, Li, et.al. "Dynamic Energy Conversion and Management Strategy for an Integrated Electricity and Natural Gas System with Renewable Energy: Deep Reinforcement Learning Approach." *Energy Convers Manage*, vol. 220, no.15 Sept. 2020.

[30] Bozorg-Haddad, Omid, Solgi, Mohammad & Loáiciga, Hugo A., 2017. Simulated Annealing. In Meta-Heuristic and Evolutionary Algorithms for Engineering Optimization. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 69–78.

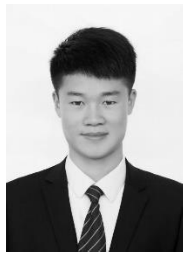
[31] Watkins, C. & Dayan, J., 1992. Q -learning. Machine Learning, 8(3-4), pp.279–292.

[32] W. Wang, N. Yu, Y. Gao *et al*., "Safe off-policy deep reinforcement learning algorithm for volt-var control in power distribution systems," *IEEE Transactions on Smart Grid*, vol. 11, no. 4, pp. 3008-3018, Dec. 2019.

[33] Energinet, "www.energinet.dk," 2020. [Online]. Available: www.energinet.dk.

[34] V. Bui, A. Hussain and H. Kim, "Double Deep Q -Learning-Based Distributed Operation of Battery Energy Storage System Considering Uncertainties," *IEEE Trans. Smart Grid*, vol. 11, no. 1, pp. 457-469, Jan. 2020.

[35] N. Sockeel, J. Gafford, B. Papari and M. Mazzola, "Virtual Inertia Emulator-Based Model Predictive Control for Grid Frequency Regulation Considering High Penetration of Inverter-Based Energy Storage System," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2932-2939, Oct. 2020.



Tao Li received the B.S. from Sichuan Normal University, Chengdu, China, in 2018. He is currently pursuing the M.S. degree in electrical engineering at the University of Electronic Science and Technology of China (UESTC). His research interests include power system operation & control and application of deep reinforcement learning.

Weihao Hu (S'06–M'13–SM'15) received the B.Eng. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 2004 and 2007, respectively, both in electrical engineering, and Ph. D. degree from Aalborg University, Denmark, in 2012.

He is currently a Full Professor and the Director of Institute of Smart Power and Energy Systems (ISPES) at the University of Electronics Science and Technology of China (UESTC). He was an Associate Professor at the Department of Energy Technology, Aalborg University, Denmark and the Vice Program Leader of Wind Power System Research Program at the same department. His research interests include artificial intelligence in modern power systems and renewable power generation. He has led/participated in more than 15 national and international research projects and he has more than 170 publications in his technical field.

He is an Associate Editor for IET Renewable Power Generation, a Guest Editor-in-Chief for Journal of Modern Power Systems and Clean Energy Special Issue on Applications of Artificial Intelligence in Modern Power Systems, a Guest Editor-in-Chief for Transactions of China Electrical Technology Special Issue on Planning and operation of multiple renewable energy complementary power generation systems, and a Guest Editor for the IEEE TRANSACTIONS ON POWER SYSTEM Special Section on Enabling very high penetration renewable energy integration into future power systems. He was serving as the Technical Program Chair (TPC) for IEEE Innovative Smart Grid Technologies (ISGT) Asia 2019 and is serving as the Conference Chair for the Asia Energy and Electrical Engineering Symposium (AEEES 2020). He is currently serving as Chair for IEEE Chengdu Section PELS Chapter. He is a Fellow of the Institution of Engineering and Technology, London, U.K. and an IEEE Senior Member.

Bin Zhang is currently working toward the M.S. degree in electrical engineering at the University of Electronic Science and Technology of China (UESTC), Chengdu. His research interest includes optimization of integrated energy system and application of deep reinforcement learning.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSTE.2021.3071268, IEEE Transactions on Sustainable Energy

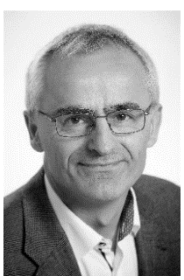
Guozhou Zhang received the B.S. from Chongqing University of Technology, Chongqing, China, in 2016, the M. S. degree from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2019. He is currently working toward the Ph.D. degree in control science and engineering at the UESTC. His research interest includes power system analysis and control.

Jian Li (M'14-SM'20) received his BSc, MSc, and Ph.D degrees in Detection Technology and Automatic Engineering from University of Electronic Science and Technology of China (UESTC) in 2007, 2010, and 2014, respectively. In 2011-2013, he was awarded a scholarship by China Scholarship Council (CSC) to visit the Telerobotic and Biorobotic Systems Group of the University of Alberta as a visiting doctoral student supervised by Dr. Mahdi Tavakoli. He is currently a professor at UESTC. His current research and academic interests include power system analysis and control, renewable energy integration, and big data.

Zhe Chen (M'95-SM'98-F'19) received the B.Eng. and M.Sc. degrees from Northeast China Institute of Electric Power Engineering, Jilin, China, and the Ph.D. degree from University of Durham, Durham, U.K.

He is a Full Professor with the Department of Energy Technology, Aalborg University, Denmark. He is the Leader of Wind Power System Research Program in the Department of Energy Technology, Aalborg University and the Danish Principal Investigator for Wind Energy of Sino-Danish Centre for Education and Research. His research areas include power systems, power electronics and electric machines; and his main current research interests are wind energy and modern power systems. He has led many research projects and has more than 400 publications in his technical field.

Dr. Chen is an Editor of the IEEE TRANSACTIONS ON POWER SYSTEMS, an Associate Editor of the IEEE TRANSACTIONS ON POWER ELECTRONICS, a Fellow of the Institution of Engineering and Technology, London, U.K., a Chartered Engineer in the U.K., and a Fellow of the IEEE.



Frede Blaabjerg (S'86–M'88–SM'97–F'03) was with ABB-Scandia, Randers, Denmark, from 1987 to 1988. From 1988 to 1992, he got the PhD degree in Electrical Engineering at Aalborg University in 1995. He became an Assistant Professor in 1992, an Associate Professor in 1996, and a Full Professor of power electronics and drives in 1998. From 2017 he became a Villum Investigator. He is honoris causa at University Politehnica Timisoara (UPT), Romania and Tallinn Technical University (TTU) in Estonia.

His current research interests include power electronics and its applications such as in wind turbines, PV systems, reliability, harmonics and adjustable speed drives. He has published more than 600 journal papers in the fields of power electronics and its applications. He is the co-author of four monographs and editor of ten books in power electronics and its applications.

He has received 32 IEEE Prize Paper Awards, the IEEE PELS Distinguished Service Award in 2009, the EPE-PEMC Council Award in 2010, the IEEE William E. Newell Power Electronics Award 2014, the Villum Kann Rasmussen Research Award 2014, the Global Energy Prize in 2019 and the 2020 IEEE Edison Medal. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON POWER ELECTRONICS from 2006 to 2012. He has been Distinguished Lecturer for the IEEE Power Electronics Society from 2005 to 2007 and for the IEEE Industry Applications Society from 2010 to 2011 as well as 2017 to 2018. In 2019-2020 he serves as President of IEEE Power Electronics Society. He is Vice-President of the Danish Academy of Technical Sciences too. He is nominated in 2014-2019 by Thomson Reuters to be between the most 250 cited researchers in Engineering in the world.