

Research Article

Research on Hierarchical and Distributed Control for Smart Generation Based on Virtual Wolf Pack Strategy

Lei Xi , Lang Liu , Yuehua Huang , Yanchun Xu , and Yunning Zhang 

College of Electrical Engineering and New Energy, China Three Gorges University, Yichang 443002, China

Correspondence should be addressed to Yuehua Huang; hyh@ctgu.edu.cn and Yanchun Xu; xyc7309@163.com

Received 9 April 2018; Accepted 13 May 2018; Published 2 July 2018

Academic Editor: Yimin Zhou

Copyright © 2018 Lei Xi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Nowadays, haze has become a big trouble in our society. One of the significant solutions is to introduce renewable energy on a large scale. How to ensure that power system can adapt to the integration and consumption of new energy very well has become a scientific issue. A smart generation control which is called hierarchical and distributed control based on virtual wolf pack strategy is explored in this study. The proposed method is based on multiagent system stochastic consensus game principle. Meanwhile, it is also integrated into the new win-lose judgment criterion and eligibility trace. The simulations, conducted on the modified power system model based on the IEEE two-area load frequency control and Hubei power grid model in China, demonstrate that the proposed method can obtain the optimal collaborative control of AGC units in a given regional power grid. Compared with some smart methods, the proposed one can improve the closed-loop system performances and reduce the carbon emission. Meanwhile, a faster convergence speed and stronger robustness are also achieved.

1. Introduction

Recently, the thermal power generation makes the environmental pollution more serious, especially the air pollution. Therefore, more and more clean energies such as wind and photovoltaics are continuously merged into the strongly coupling interconnected power grid [1]. However, new troubles, such as voltage over limit and power fluctuations as well as frequency instability [2–4], are brought out. Meanwhile, the safe operation of the power grid is also affected. The traditional centralized automatic generation control (AGC) cannot obtain the similar control performance with the decentralized AGC since the energy distributions are more dispersed. It will be an inevitable trend for the future smart grid to research the decentralized AGC.

In recent years, many scholars have devoted to the optimal control strategy of decentralized AGC [5–13]. Authors in [6] put forward the concept of optimal AGC by using the original dual transformation method, which is based on the optimal control theory. It showed that the dynamic equation and the constructed AGC control strategy of the

interconnected system could realize multiarea decentralized optimal AGC control. However, the used optimal AGC controller needed to feedback all the state variables which were difficult to be obtained directly in the actual system. In [8], a new method was proposed based on the model predictive control. It focused on a decentralized optimal AGC control strategy based on cooperative synchronous power grid. While the stability and robustness of the multivariable predictive control method including the application in actual AGC system needed to be further studied, the method was a great amount of calculation and time-consuming. Yu et al. [11] demonstrated that an optimal AGC can be achieved under the circumstance that the agents are in small number. However, the algorithm is only applicable to systems with a small number of agents and its application is limited. In the same way, the decentralized control has been studied by the author in the early stage, namely, decentralized correlated equilibrium $Q(\lambda)$ -learning (DCEQ(λ)) [12] based on multiagent (MA). It can solve the complex stochastic dynamic characteristics and optimal coordination control of AGC after the access of distributed energy. Nevertheless, if the number of

MA increases, the searching time for the MA equilibrium solution is geometric growth, which will limit the application of DCEQ(λ) in larger systems. Therefore, the decentralized win or learn fast policy hill-climbing(λ) (DWoLF-PHC(λ)) [13] based on MA was developed, in which by using average mixed strategy instead of equilibrium strategy. Thus, the dynamic characteristics of the system are effectively improved, and the dynamic optimization control of the total power is also obtained. However, the DWoLF-PHC(λ) still has multisolution problem. It results in system instability when the number of MA increases sharply.

The above literatures have some limitations that they only focus on the control strategy of the total power in the AGC. However, the dynamic optimal allocation of the total power is not involved. In fact, the modern power grid has gradually been developed into a hierarchical and distributed control (HDC) structure, which integrates the large-scale new energy. For this reason, a single control strategy is difficult to meet the requirements of control performance standards (CPS). Therefore, a hierarchical and distributed control based on virtual wolf pack strategy (HDC-VWPS) is proposed in order to attenuate the stochastic disturbance problem caused by massive integration of new energy to the power grid. The proposed strategy is based on multiagent system stochastic consensus game (MAS-SCG). It is divided into two parts. The first part is an AGC optimal control method which combines a new win-lose judgment criterion, policy hill-climbing algorithm (PHC) [14], and eligibility trace [15]. Especially, the new win-lose judgment criterion is named as policy dynamics-based WoLF (PDWoLF) [16]. Moreover, the control method, which is called PDWoLF-PHC(λ), is based on multiagent system stochastic game (MAS-SG) theory. Meanwhile, the second part is the collaborative consensus (CC) algorithm [17] which is based on multi-agent system collaborative consensus game (MAS-CC) theory. This algorithm is used to distribute the total power dynamically and optimally. Consequently, the perfect combination of AGC control and distribution is realized. At the same time, the intelligence from the whole to the branch is truly obtained. The significant difference between smart generation control (SGC) and AGC is that the original proportional-integral (PI) control in AGC is replaced by the smart control in SGC.

The rest of the paper is as follows. The SGC framework based on HDC structure is proposed in Section 2. The HDC-VWPS is expounded in Section 3. Meanwhile, Section 4 is the AGC design based on HDC-VWPS. Section 5 covers the case study, and Section 6 summarizes the full text, respectively.

2. SGC Framework Based on HDC Structure

Hierarchical reinforcement learning (HRL) [18] is a hierarchical control method that can solve the problem of “curse of dimensionality” in traditional reinforcement learning effectively. A new method, namely, HDC-VWPS, is put forward to obtain the optimal total power and its optimal dispatch dynamically. The term “virtual wolf pack” is a generator set group (GSGs) of a certain control area. The PDWoLF-PHC(λ) with the win or learn fast (WoLF)

attribute based on heterogeneous MAS-SG theory is adopted to obtain the total power of each GSG. Meanwhile, the ramp time CC algorithm based on homogeneous MAS-CC theory is used to distribute the total power to each unit dynamically in order to achieve the optimal coordination control of each GSG. The “leader” of virtual wolf pack refers to a new dispatcher who is responsible for communicating, contacting and cooperating with the leaders of the other GSGs, and sending the instructions to each unit in the GSGs. Each GSG only has one leader. The SGC framework based on HDC structure is shown in Figure 1, where ΔP_{tie} is the tie-line exchange power, Δf is the interconnected power grid frequency error, ΔP_i is the total power of GSG i ($i = 1, 2, \dots, n$), and ΔP_{iu} is the regulation power of the u th unit in GSG i .

3. HDC-VWPS

A HDC-VWPS is designed to coordinate and optimize the operation of GSGs in the SGC system with HDC structure through the integration of MAS-SG and MAS-CC.

3.1. MAS-SG Framework. Based on the MAS-SG framework, a PDWoLF-PHC(λ) algorithm is proposed to the game among GSGs to obtain total power command of each GSG.

The WoLF principle can meet the convergence requirement by changing the learning rate without sacrificing rationality, namely, learn quickly when losing and cautiously when winning [14]. However, in more than 2×2 games, the players cannot accurately calculate the win-lose criterion and can only rely on the estimation. Therefore, an improved WoLF version, PDWoLF, whose judgment criterion can be accurately computable in more than 2×2 games, was explored in [16]. Also, it can converge to Nash equilibrium in more than 2 action games.

It indicates that PHC algorithm can meet the requirement of the rationality in [14]. Therefore, PDWoLF-PHC can satisfy the requirements of the convergence and the rationality at the same time. It also converges faster with a higher learning rate ratio [16]. The PDWoLF-PHC is the extension of the classical Q-learning [19]. It combines the multistep backtracking idea of the SARSA(λ) [15] to search the optimal action-value function through the continuous trial and error dynamically. The parameter λ refers to the use of an eligibility trace. It can solve the temporal credit assignment of time-delayed reinforcement learning. The optimal value function $V^{\pi^*}(s)$ and strategy $\pi^*(s)$ are as follows.

$$\begin{aligned} V^{\pi^*}(s) &= \max_{a \in A} Q(s, a), \\ \pi^*(s) &= \operatorname{argmax}_{a \in A} Q(s, a), \end{aligned} \quad (1)$$

where A is the set of possible actions under state s .

The eligibility trace is updated by

$$e_{k+1}(s, a) = \begin{cases} \gamma \lambda e_k(s, a) + 1, & \text{if } (s, a) = (s_k, a_k), \\ \gamma \lambda e_k(s, a), & \text{otherwise,} \end{cases} \quad (2)$$

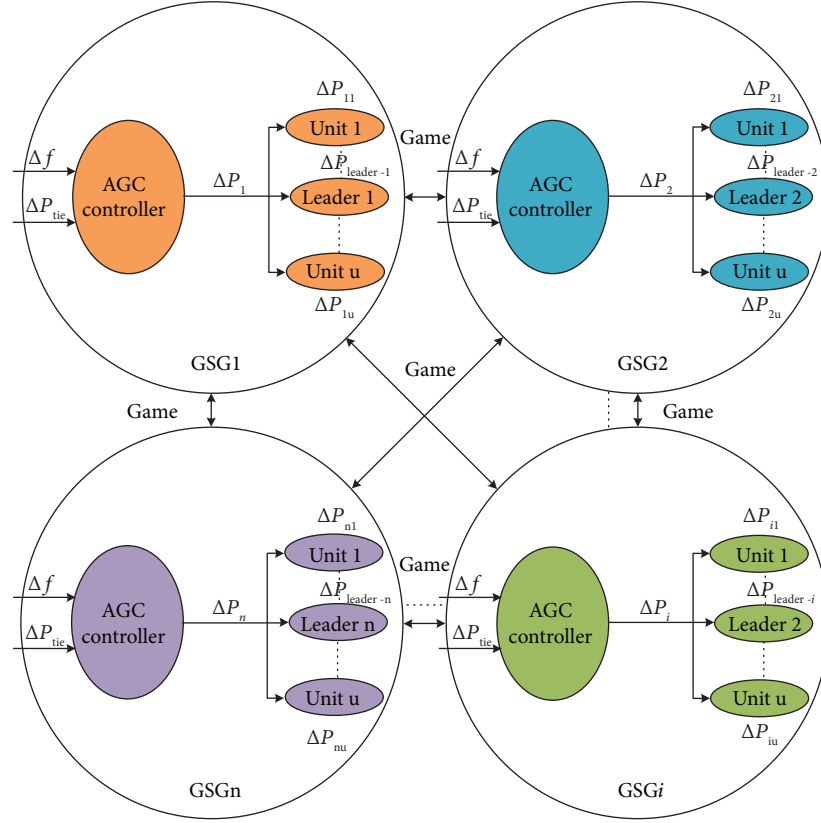


FIGURE 1: The SGC framework based on HDC structure.

where $e_k(s, a)$ denotes the eligibility trace at the k th step iteration under state s and action a , γ is the discount factor, and λ is the trace-attenuation factor.

The Q function will be iteratively updated according to

$$Q_{k+1}(s_k, a_k) = Q_k(s_k, a_k) + \alpha \cdot \left[R(s_k, s_{k+1}, a_k) + \gamma \max_{a' \in A} Q_k(s_{k+1}, a') - Q_k(s_k, a_k) \right] e_k(s_k, a_k), \quad (3)$$

where $0 < \alpha < 1$ is the Q -learning rate and $R(s_k, s_{k+1}, a_k)$ is the reward function value from state s_k to s_{k+1} under the selected action a_k . $Q(s, a)$ is the Q value function when executing action a in state s , which uses look-up table method. a' is a greedy action. After sufficient trial and error iterations are done, the state-value function $Q(s, a)$ will converge to the Q^* matrix with the probability of one. Finally, an optimal control strategy, represented by the optimal Q function (Q^* matrix), can be obtained.

The win-lose criterion of PDWoLF-PHC(λ) is determined by two parameters δ_{win} and δ_{lose} for a given agent. Strategy $\pi(s_k, a_k)$ is updated for an agent according to (4) in the state state-action pair (s_k, a_k) .

$$\pi_{k+1}(s_k, a_k) = \pi_k(s_k, a_k) + \Delta_{s_k a_k}, \quad (4)$$

where $\Delta_{s_k a_k}$ is the variable quantity of the updating strategy. The updating rule is described as follows.

$$\Delta_{s_k a_k} = \begin{cases} -\delta_{s_k a_k}(k), & \text{if } a_k \neq \arg \max_{a'} Q_{k+1}(s_k, a'), \\ \sum_{a' \neq a_k} \delta_{s_k a'}(k), & \text{otherwise,} \end{cases} \quad (5)$$

$$\delta_{s_k a_k}(k) = \min \left(\pi_k(s_k, a_k), \frac{\delta}{|A| - 1} \right). \quad (6)$$

In (6), $|A|$ is the number of possible actions. δ is the variable learning rate and $\delta_{\text{win}} < \delta_{\text{lose}} \in (0, 1]$. Also, $\varphi = \delta_{\text{lose}} / \delta_{\text{win}}$ is defined as the variable learning rate ratio. δ is updated by

$$\delta = \begin{cases} \delta_{\text{win}}, & \text{if } \Delta_k(s_k, a_k) \cdot \Delta_k^2(s_k, a_k) < 0, \\ \delta_{\text{lose}}, & \text{otherwise,} \end{cases} \quad (7)$$

where $\Delta_k^2(s_k, a_k)$ is the decision space slope value and $\Delta_k(s_k, a_k)$ is the decision change rate at the k th step iteration. Meanwhile, $\Delta^2(s_k, a_k)$ and $\Delta(s_k, a_k)$ are updated by

$$\begin{aligned} \Delta_{k+1}^2(s_k, a_k) &= \Delta s_k a_k - \Delta_k(s_k, a_k), \\ \Delta_{k+1}(s_k, a_k) &= \Delta_{s_k a_k}. \end{aligned} \quad (8)$$

3.2. *MAS-CC Framework.* The MAS-CC framework is introduced into the HDC-VWPS to dynamically allocate the total power command to each unit.

3.2.1. *Graph Theory.* The topology of MAS can be expressed as a directed graph $G = (V, E, B)$ with a set of nodes $V = \{v_1, v_2, \dots, v_n\}$, edges $E \subset V \times V$, and a weighted adjacency matrix $B = [b_{ij}] \hat{I}R^{n \times n}$. Among them v_i denotes the i th agent, edge means the relationship among agents, and constant $b_{ij} \geq 0$ is the weight factor between v_i and v_j . If there is a connection between any two vertices, then the graph G is called a strongly connected graph. The Laplacian matrix $L = [l_{ij}] \hat{I}R^{n \times n}$ of graph G can be written as follows.

$$l_{ii} = \sum_{j=1, j \neq i}^n b_{ij}, l_{ij} = -b_{ij}, \forall i \neq j, \quad (9)$$

where the matrix L reflects the topology of the MA network.

3.2.2. *Collaborative Consensus.* In a MAS, it is usually called collaborative consensus (CC) [20] while an agent interacts with the adjacent one to reach the consensus. A MAS consisting of n autonomous agents is regarded as a node in a directed graph G . The purpose of CC is to obtain a consensus in each agent and to update state in real time after communicating with neighboring agents. Due to the communication delay among agents, the first-order CC algorithm of a discrete system is chosen as follows.

$$\psi_i[k+1] = \sum_{j=1}^n d_{ij}[k] \psi_j[k], \quad (10)$$

where ψ_i is the state of the i th agent, k represents the discrete time series, and $d_{ij}[k]$ denotes the (i, j) entry of the row stochastic matrix $D = [d_{ij}] \hat{I}R^{n \times n}$ at discrete time. $k.d_{ij}[k]$ is given by

$$d_{ij}[k] = \frac{|l_{ij}|}{\sum_{j=1}^n |l_{ij}|}. \quad (11)$$

The CC algorithm can be achieved if and only if the directed graph is strongly connected on the condition of the continuous communication and constant gain b_{ij} .

3.2.3. *Ramp Time Collaborative Consensus.* The ramp time is chosen as the consensus variable among all units in a GSG. A unit which has a higher ramp rate will be distributed with more disturbances. The ramp time of the u th unit in GSGi can be obtained as follows.

$$t_{iu} = \frac{\Delta P_{iu}}{\Delta P_{iu}^{\text{rate}}}, \quad (12)$$

where ΔP_{iu} is the regulation power of the u th unit in GSGi. $\Delta P_{iu}^{\text{rate}}$ is the ramp rate of the unit and is calculated as follows.

$$\Delta P_{iu}^{\text{rate}} = \begin{cases} P_{iu}^{\text{rate}+}, & \Delta P_i > 0, \\ P_{iu}^{\text{rate}-}, & \Delta P_i < 0, \end{cases} \quad (13)$$

where $\Delta P_{iu}^{\text{rate}+}$ and $\Delta P_{iu}^{\text{rate}-}$ are the upper and lower bounds of the ramp rate, respectively.

The ramp time of the u th unit in GSGi can be updated according to (10) as follows.

$$t_{iu}[k+1] = \sum_{v=1}^{U_i} d_{uv}^{(i)}[k] t_{iv}[k], \quad (14)$$

where U_i is the total number of units in GSGi. $D^{(i)} = [d_{uv}^{(i)}] \in R^{U_i \times U_i}$ is the row stochastic matrix.

Then the ramp time of the GSGi leader can be updated as follows.

$$t_{iu}[k+1] = \begin{cases} \sum_{v=1}^{U_i} d_{uv}^{(i)}[k] t_{iv}[k] + \xi_i \Delta P_{\text{error}-i}, & \Delta P_i > 0, \\ \sum_{v=1}^{U_i} d_{uv}^{(i)}[k] t_{iv}[k] - \xi_i \Delta P_{\text{error}-i}, & \Delta P_i < 0, \end{cases} \quad (15)$$

where $\xi_i > 0$ represents the GSGi's adjustment factor of the power error. $\Delta P_{\text{error}-i}$ denotes the power error between the GSGi total power and the total power of all units. It is obtained from

$$\Delta P_{\text{error}-i} = \Delta P_i - \sum_{u=1}^{U_i} \Delta P_{iu}. \quad (16)$$

In the condition of the total power command $\Delta P_i > 0$, if $\Delta P_{\text{error}-i} > 0$, the ramp time t_{iu} needs to be increased; otherwise t_{iu} needs to be reduced. Oppositely, t_{iu} will be increased or decreased in condition that $\Delta P_i < 0$.

As a ramp time CC algorithm among units is adopted, the power of some units may exceed their maximum power. At the same time, the smaller the unit maximum ramp time t_{iu}^{max} is adopted, the faster the power limit is reached. While the power limit is reached, the u th unit's power and ramp time are as follows.

$$\Delta P_{iu} = \begin{cases} \Delta P_{iu}^{\text{max}}, & \Delta P_{iu} > \Delta P_{iu}^{\text{max}}, \\ \Delta P_{iu}^{\text{min}}, & \Delta P_{iu} < \Delta P_{iu}^{\text{min}}, \end{cases} \quad (17)$$

$$t_{iu} = t_{iu}^{\text{max}} = \begin{cases} \Delta \frac{P_{iu}^{\text{max}}}{P_{iu}^{\text{rate}+}}, & \Delta P_{iu} > \Delta P_{iu}^{\text{max}}, \\ \Delta \frac{P_{iu}^{\text{min}}}{P_{iu}^{\text{rate}-}}, & \Delta P_{iu} < \Delta P_{iu}^{\text{min}}, \end{cases}$$

where $\Delta P_{iu}^{\text{max}}$ and $\Delta P_{iu}^{\text{min}}$ are the maximum and minimum reserve capacity of the u th unit in GSGi, respectively. Furthermore, if the power ΔP_{iu} of the u th unit exceeds its limit, the weight factor becomes as follows.

$$b_{uv}^{(i)} = 0, v = 1, 2, \dots, U_i. \quad (18)$$

Initialize $Q_0(s, a)$, $\pi_0(s, a)$ and $e_0(s, a)$, for all $s \in S$, $a \in A$;
 Set parameters δ_{win} , δ_{lose} , φ , γ , λ , α , and $T_{step} = AGC$ decision time;
 Give the initial state s_0 , $k = 0$;
 Repeat
 (1) Choose an exploration action a_k based on the mixed strategy set $\pi(s_k, a_k)$;
 (2) Execute the exploration action a_k to AGC units and run LFC system for the next T_{step} sec;
 (3) Observe a new state s_{k+1} via CPS1 and ACE;
 (4) Obtain a short-term reward $R(s_k, s_{k+1}, a_k)$ using Eq. (19);
 (5) Update eligibility trace according to Eq. (2);
 (6) Update Q function using Eq. (3);
 (7) Select variable learning rate δ with Eq. (7);
 (8) Compute $\Delta_{s_k a_k}$ by Eq. (5) and Eq. (6);
 (9) Calculate $\Delta^2(s_k, a_k)$ and $\Delta(s_k, a_k)$ according to Eq. (8);
 (10) Update the mixed strategy $\pi(s_k, a_k)$ according to Eq. (4);
 (11) Obtain the total power ΔP_{iu} of the GSGi;
 (12) Determine the ramp rate according to Eq. (13);
 (13) Execute CC algorithm according to Eq. (14) and Eq. (15);
 (14) Calculate the u th unit power ΔP_{iu} in GSGi;
 (15) If the power limit is not exceeded, then execute step 17;
 (16) Calculate ΔP_{iu} and t_{iu} according to Eq. (17). And update $D^{(i)}$ using Eq. (9), Eq. (11) and Eq. (18);
 (17) Calculate the power error $\Delta P_{error-i}$ according to Eq. (16);
 (18) If $|\Delta P_{error-i}| < \Delta P_{error-i}^{\max}$ is not satisfied, execute step 13;
 (19) Output the u th unit power ΔP_{iu} ;
 (20) Set $k = k + 1$, and return to step 1.
End

ALGORITHM 1: Execution steps of the HDC-VWPS.

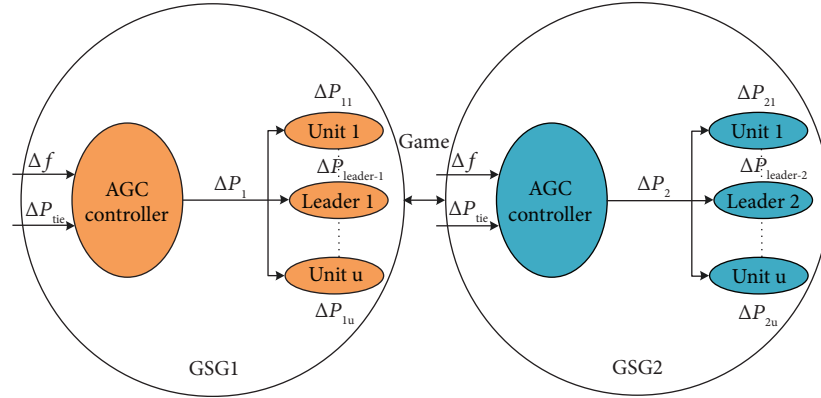


FIGURE 2: Modified power system model based on IEEE two-area LFC.

where $B = [b_{uv}^{(i)}] \in R^{U_i \times U_i}$ is the weighted adjacency matrix of the GSGi.

4. AGC Design Based on HDC-VWPS

4.1. Reward Function Selection. The impact of energy management system (EMS) on the environment is considered, and carbon emission (CE) as part of the reward function is also introduced. Meanwhile, in the load frequency control (LFC), each regional power grid will control the generator set in this area according to its own area control

error (ACE). The main purpose is the ACE is zero when the steady state is reached. Therefore, in the reward function, the weighted sum of CE and ACE is taken as the objective function. The reward function in GSGi is defined as follows.

$$R(s_k, s_{k+1}, a_k) = -\eta \frac{\left(\sum_{u=1}^{U_i} C_{iu} [\Delta P_{iu}(k)] \right)}{1000} - (1 - \eta) [ACE(k)]^2$$

s.t. $\Delta P_{ui}^{\min} \leq \Delta P_{iu}(k) \leq \Delta P_{iu}^{\max}$,

(19)

TABLE 1: Model parameters of GSG units in the Hubei power grid.

GSG number	Plant types	Unit number	ΔP_{iu}^{\max} (MW)	ΔP_{iu}^{\min} (MW)	GRC (MW/min)	C_{iu} (kg/kWh)
GSG1	Coal-fired power plants	G1	200	-200	5	0.99
		G2	200	-200	5	0.99
		G3	200	-200	5	0.99
		G4	176.5	176.5	4.4	0.89
		G5	300	-300	7	0.99
		G6	300	-300	7	0.99
		G7	300	-300	7	0.99
		G8	350	-350	8.2	0.89
		G9	185	-185	4.6	0.99
	Pumped storage power plant	G10	300	0	300	0
		G11	300	0	300	0
		G12	300	0	300	0
		G13	300	0	300	0
GSG2	Coal-fired power plants	G14	220	-220	5.5	0.89
		G15	220	-220	5.5	0.89
		G16	600	-600	12	0.89
		G17	600	-600	12	0.89
		G18	300	-300	6	0.99
		G19	300	-300	6	0.99
GSG3	Coal-fired power plants	G20	300	-300	6	0.99
		G21	300	-300	6	0.99
		G22	300	-300	6	0.99
		G23	300	-300	6	0.99
		G24	600	-600	12	0.89
		G25	600	-600	12	0.89
		G26	150	0	150	0
		G27	150	0	150	0
	Hydropower plants	G28	150	0	150	0
		G29	150	0	150	0
		G30	150	0	150	0
		G31	150	0	150	0
		G32	170	0	170	0
GSG4	Coal-fired power plants	G33	170	0	170	0
		G34	300	-300	6	0.99
		G35	300	-300	6	0.99
		G36	300	-300	6	0.99
		G37	300	-300	6	0.99
		G38	1000	-1000	20	0.87
	Hydropower plants	G39	1000	-1000	20	0.87
		G40	300	0	300	0
		G41	300	0	300	0
		G42	300	0	300	0
G43	300	0	300	0		

where $\Delta P_{iu}(k)$ is the actual output power of the iu th unit in GSG i at the k th iteration; $ACE(k)$ indicates the instantaneous value of ACE at the k th iteration; η and $(1-\eta)$ represent the weight factor of controlled area's CE and

ACE, respectively. Here, η equals 0.5. C_{iu} is the CE intensity coefficient of the iu th unit in GSG i , whose unit is kg/kWh. ΔP_{iu}^{\max} and ΔP_{iu}^{\min} are the upper and lower bounds of the iu th unit's capacity in GSG i , respectively. The CE

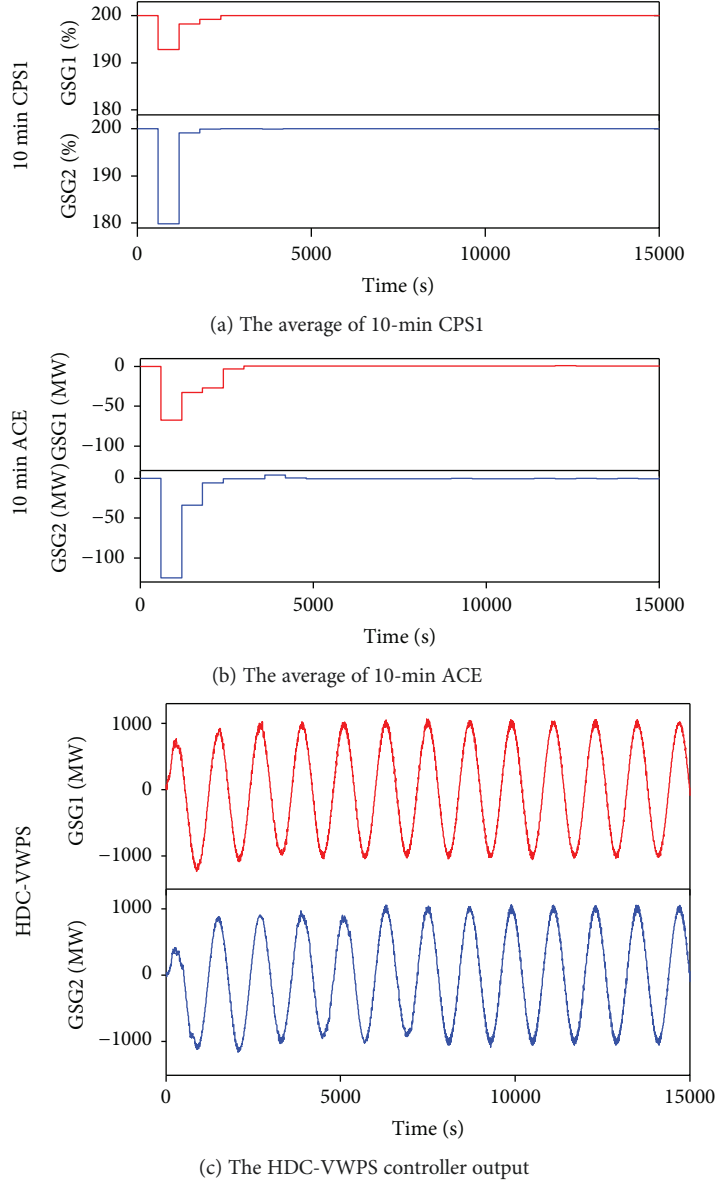


FIGURE 3: The prelearning of HDC-VWPS obtained in each GSG.

intensity coefficients for each type of generator set are as follows.

$$C_{iu}^{\text{water}} = 0, \quad C_{iu}^{\text{coal-fired}} = \begin{cases} 0.87, & \Delta P_{iu} > 600, \\ 0.89, & 300 < \Delta P_{iu} \leq 600, \\ 0.99, & \Delta P_{iu} \leq 300, \end{cases} \quad (20)$$

where ΔP_{iu} is the u th unit regulation power of the GSG i in MW.

4.2. Parameter Setting. A reasonable set of six parameters $\lambda, \gamma, \alpha, \delta, \varphi$, and ξ_i is required in the design of the control system.

The trace-attenuation factor λ allocates the credits among state-action pairs. Usually, the parameter λ is located

between 0 and 1. It determines the convergence rate and the non-Markov decision process (MDP) effects for large time-delay systems. Generally, the factor λ can be interpreted as a time scaling element in the backtracking. For Q-function errors, a small λ means that few credit will be given to the historical state-action pairs while a large λ denotes that much credit will be assigned. Through trial and error, it shows that $0.7 < \lambda < 0.95$ is acceptable. Here, $\lambda = 0.9$ is selected.

The discount factor γ is between 0 and 1, which discounts the future rewards in Q functions. A value close to 1 should be chosen as the latest rewards in the thermal-dominated LFC process which is the most important. Experiments demonstrate that $0.6 < \gamma < 0.95$ is proper. Here, $\gamma = 0.9$ is chosen.

The Q-learning rate α is set between 0 and 1, which weighs the convergence rate of the Q-functions, namely, algorithm stability. Note that a larger α can accelerate the learning rate, while a smaller α can enhance the system

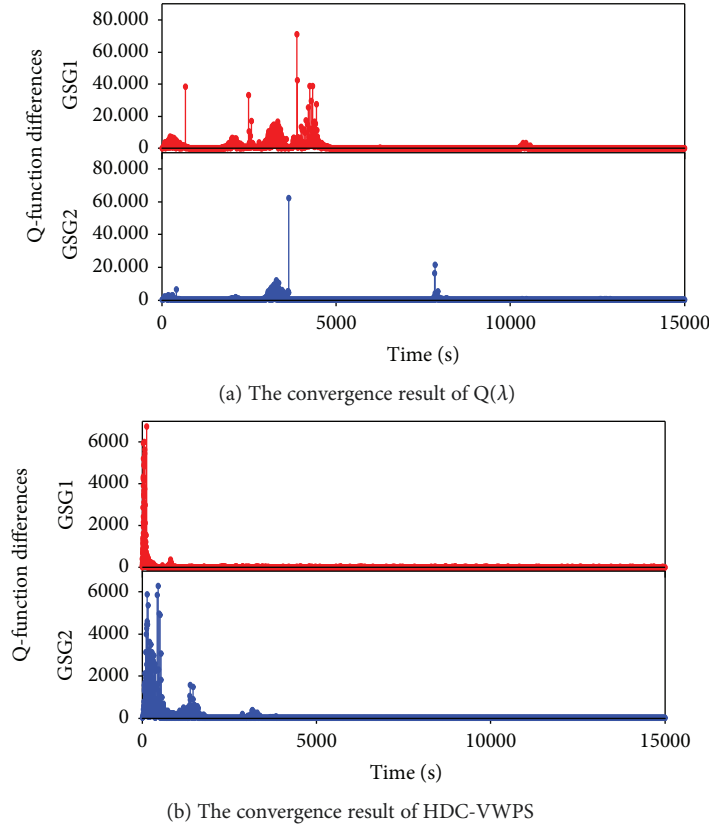


FIGURE 4: The convergence result of Q-function differences obtained in each GSG during prelearning.

stability. In the prelearning process, the initial value of α is chosen to be 0.1 to obtain the overall search. After that, in order to gradually increase the stability of the system, it will be reduced in a linear way.

The variable learning rate δ is between 0 and 1, which derives an optimal policy by maximizing the action value. Especially, the algorithm will be degraded into Q-learning if δ equals 1. The main reason is that a maximum value action is permanently executed in every iteration. For a fast convergence rate, the greedy strategy with a variable learning rate ratio $\varphi = \delta_{\text{lose}}/\delta_{\text{win}} = 4$ is selected in a stochastic game. Through trial and error, it shows that $\delta_{\text{win}} = 0.06$ can obtain stable control characteristics.

The value of power error adjustment factor ξ_i in GSG i is related to ΔP_i , which is shown in

$$\xi_i = \begin{cases} 0.0025, & \Delta P_i > 500, \\ 0.005, & 0 < \Delta P_i < 500, \\ 0.008, & -500 < \Delta P_i < 0, \\ 0.02, & -1000 < \Delta P_i < -500, \\ 0.03, & \Delta P_i < -1000. \end{cases} \quad (21)$$

ΔP_i is the total power of GSG i in MW.

4.3. *HDC-VWPS Procedure.* The Overall HDC-VWPS Procedure Is Described in Algorithm 1.

5. Case Study

5.1. *The Modified Model with Two-Area LFC Power System in IEEE.* In order to test the control performance of the proposed strategy, an IEEE-modified model with two-area LFC power system [21] is selected as the simulation object, whose framework is shown in Figure 2. The system parameters are taken from [22], and those of GSG1 and GSG2 are provided in Table 1.

The work cycle of the AGC is set to be 4 s. Note that HDC-VWPS has to undergo a sufficient prelearning through off-line trial and error before the final online implementation. It includes extensive explorations in CPS state space for the optimization of Q-functions and state-value functions [23]. Figure 3 presents the prelearning of each area produced by a continuous 10 min sinusoidal disturbance. It is obvious that the HDC-VWPS can converge to the optimal strategy in each GSG with qualified CPS1 (the average of 10 min CPS1) and $E_{\text{AVE:10 min}}$ (the average of 10 min ACE).

Furthermore, a Q matrix $\|Q_{ik}(s, a) - Q_{i(k-1)}(s, a)\|_2 \leq \varsigma$ with 2 norms is used as the criterion for the prelearning termination of an optimal strategy [24]. $\varsigma = 0.1$ is a specified positive constant. Both the Q value and look-up table will be automatically saved after the prelearning, such that HDC-VWPS can be applied into a real power system. The convergence result of Q-function differences is given in Figure 4. The result is obtained in each GSG during the prelearning, in which the HDC-VWPS can accelerate the convergence rate by nearly 26.7%~40% over that of $Q(\lambda)$.

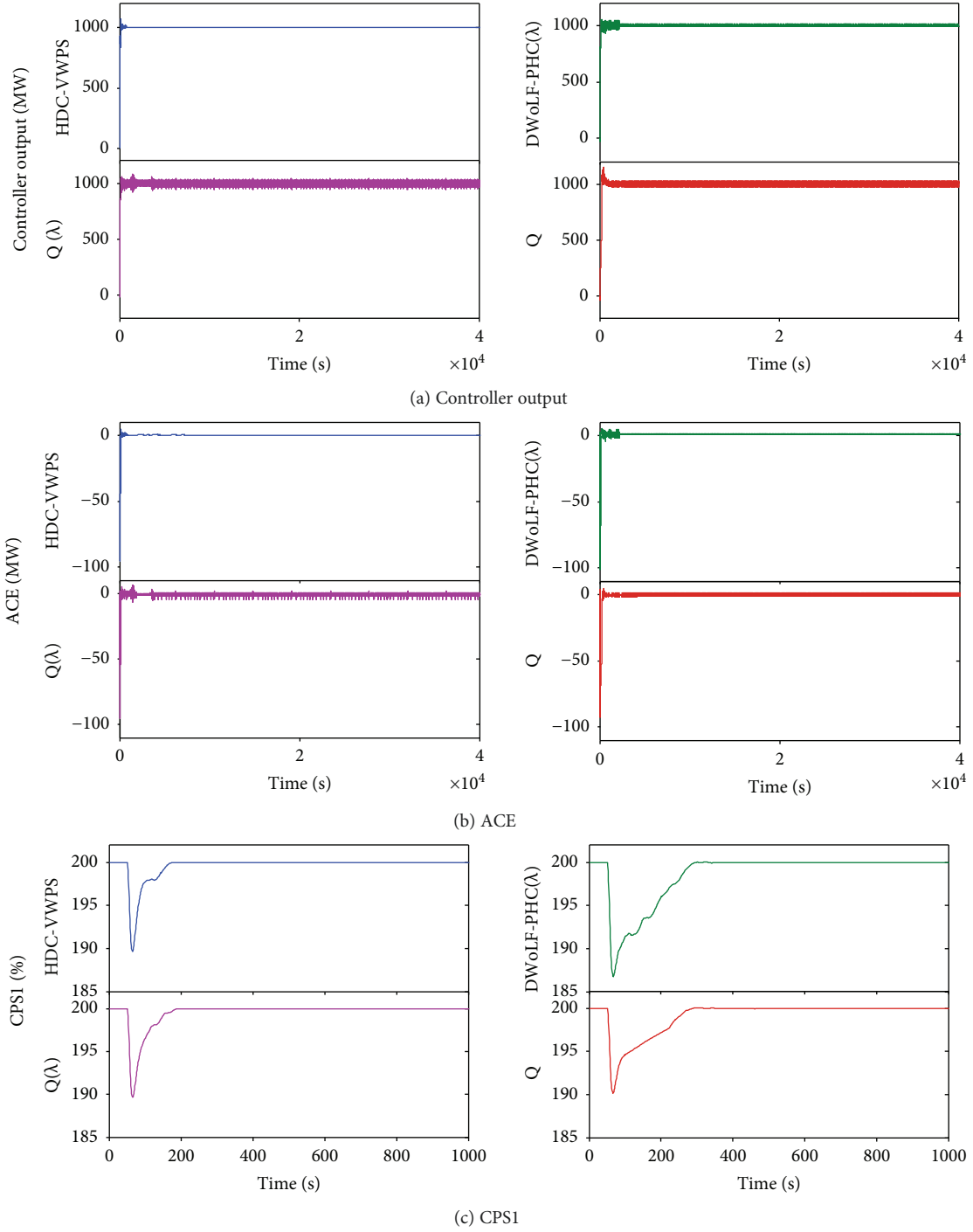


FIGURE 5: Control performance of four AGC controllers under a step load disturbance.

In order to evaluate the robustness of each algorithm, the control performances of DWoLF-PHC(λ), $Q(\lambda)$, and Q -learning are compared with that of HDC-VWPS under a step and a stochastic load disturbance in GSG1. The simulation results under a step load disturbance are shown in Figure 5. In Figure 5(a) it is shown that the overshoots are around 6.3758%, 4.907%, 7.2614%, and 13.0435%, respectively. Meanwhile, in Figure 5(b), it refers that the average values of ACE are 0.1261 MW, 1.0682 MW, 1.2216 MW, and

1.0438 MW, respectively. In addition, in Figure 5(c), it is illustrated that the minimum CPS1 is 189.6487%, 186.7696%, 189.6426%, and 190.1703%, respectively. In the meantime, the simulation results under a stochastic load disturbance are described in Figure 6. In Figure 6(a), it is demonstrated that HDC-VWPS has the strongest robustness. Besides, in Figure 6(b), it refers that the average values of ACE are 22.7175 MW, 45.1846 MW, 66.6484 MW, and 75.7486 MW, respectively. Moreover, in Figure 6(c), it is presented that

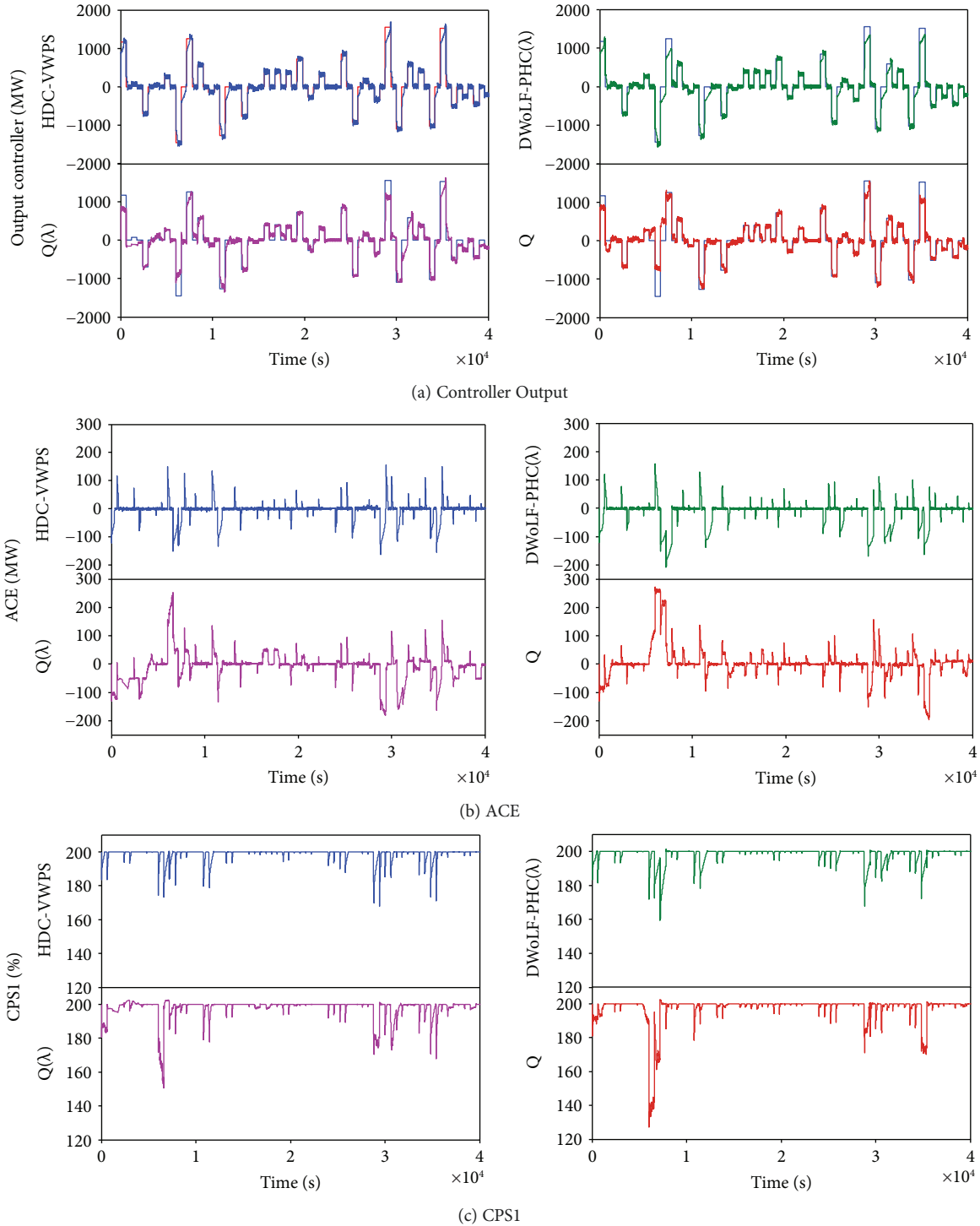


FIGURE 6: Control performance of four AGC controllers under a stochastic load disturbance.

the minimum CPS1 is 167.7471%, 159.4400%, 150.6757%, and 127.3168%, respectively. Therefore, HDC-VWPS can provide better control performances for AGC units.

The stochastic white noise is used as the load disturbance after the prelearning process, in which the control performance of each algorithm obtained in each GSG is summarized in Figure 7. CE, $|\Delta f|$ (average values of the frequency deviation), $|E_{AVE:1 \min}|$ (average values of 1 min

ACE), and CPS1 are the average values over 24h. It can be seen from Figure 7 that compared with the other methods, HDC-VWPS can reduce CE by 1.21%~1.51%, $|\Delta f|$ by $4.5647 \times 10^{-4} \sim 7.5851 \times 10^{-4}$ Hz, and $|E_{AVE:1 \min}|$ by 5.79%~44.22% and increase CPS1 by 0.0007%~0.02%.

5.2. Four-Area Model of Hubei Power Grid. Four-area model of Hubei power grid is shown in Figure 8. As shown in

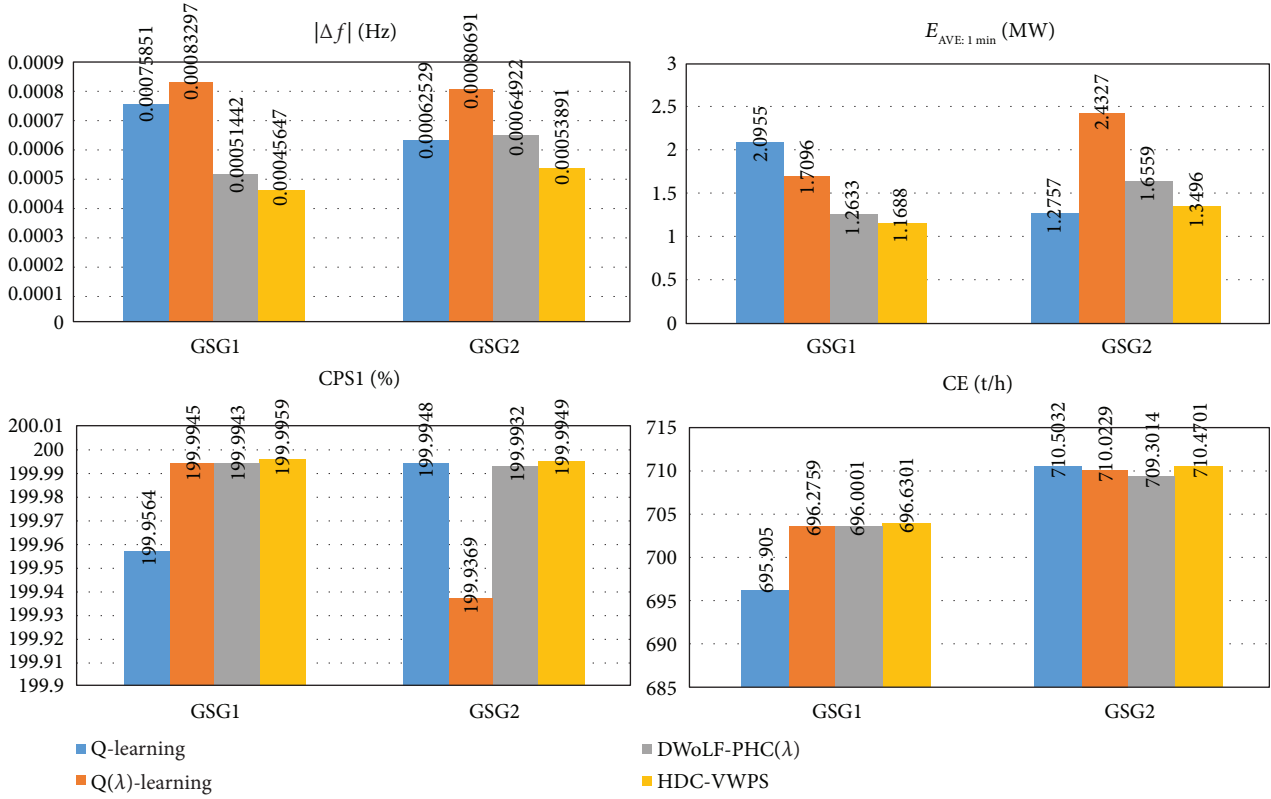


FIGURE 7: Statistic performance of each GSG in the two-area LFC modified power system model.

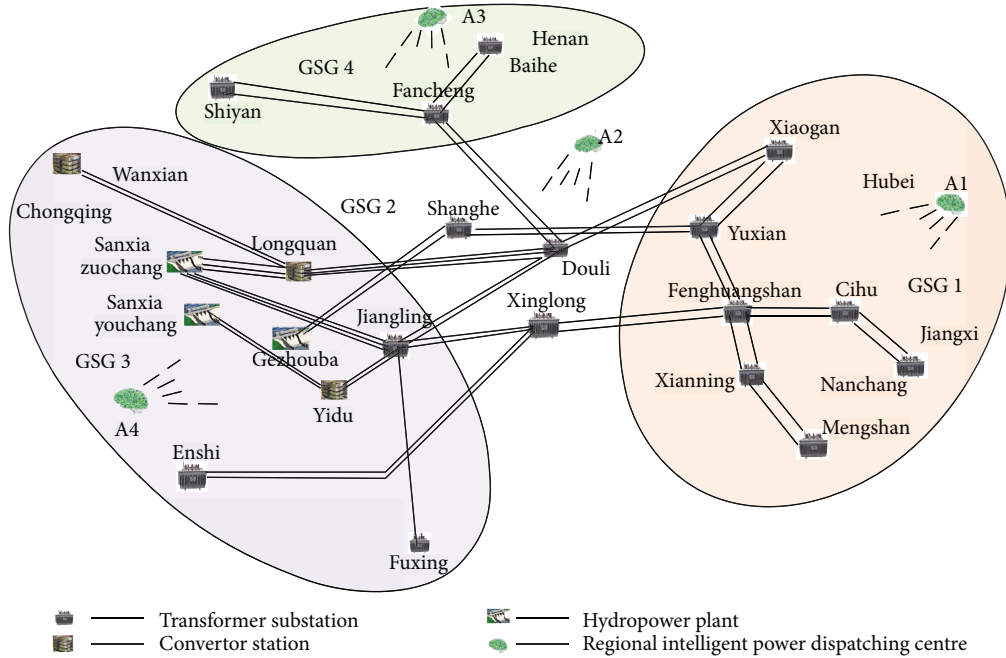


FIGURE 8: The interconnected network of Hubei power grid model in China.

Figure 9, an AC/DC hybrid Hubei power grid model, which consists of totally 43 units of four GSGs, is analyzed in the paper. The control performance is CPS, and the work cycle of AGC is set to be 4 s. The L_{10} of Hubei power grid model is 118 MW. ΔX_g is the governor output, and ΔP_g is the turbine

output. At the same time, T_g is the time constant of the governor, T_t is the time constant of the turbine, and $K_p/(1 + sT_p)$ is the equivalent function of AC frequency response, respectively. Related parameters are set as follows. $T_g = 0.078$, $T_t = 0.32$, $K_{p1} = 0.0027$, $K_{p2} = 0.0029$, $K_{p3} = 0.0027$,

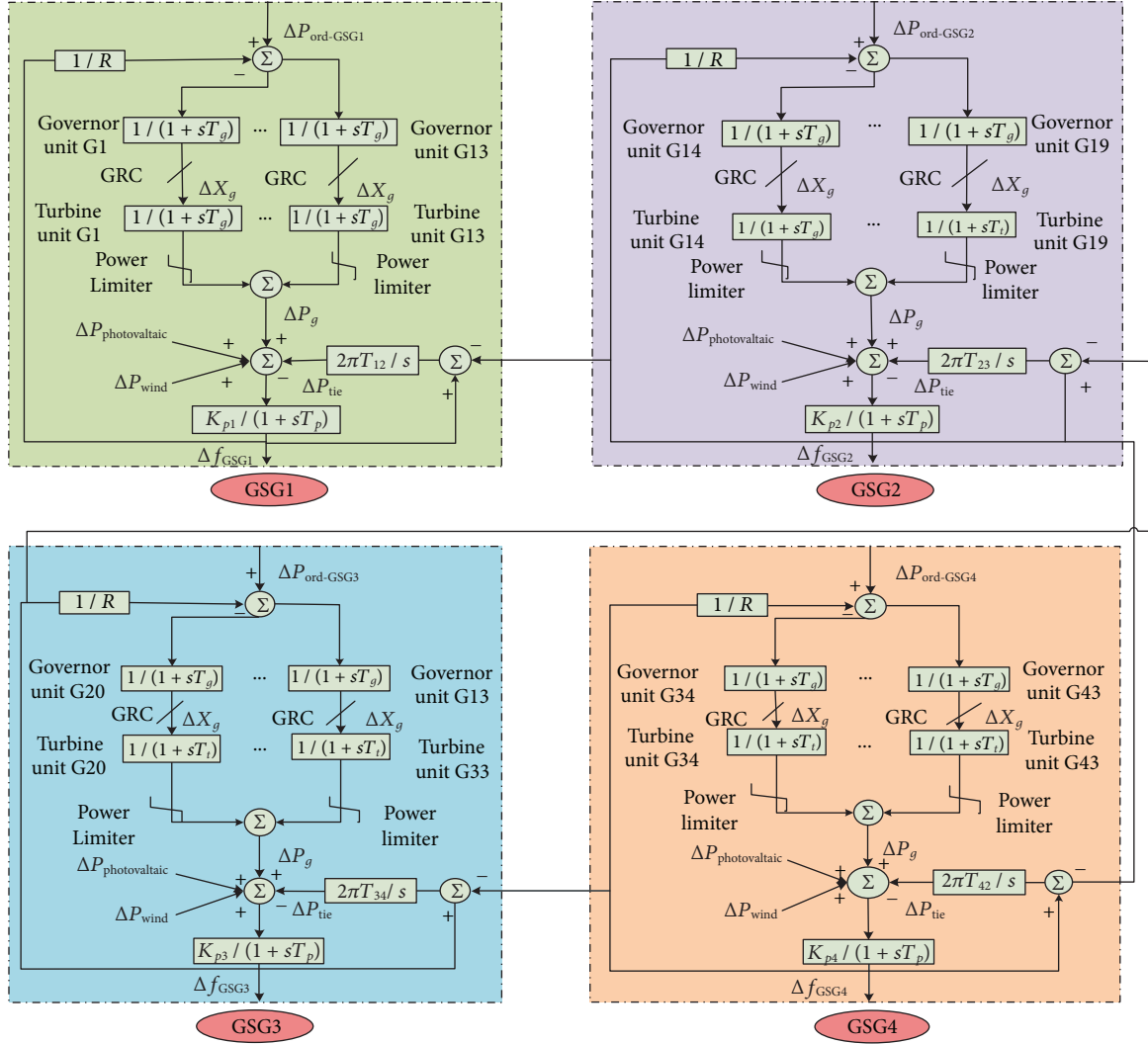


FIGURE 9: Hubei power grid model.

$K_{p4} = 0.0025$, and $T_p = 20$. Generation rate constraint (GRC) is the $|P_{iu}^{rate+}/P_{iu}^{rate-}|$ in this study. GRC and all the other system parameters are given in Table 1.

The system includes coal-fired power plants, hydro-power plants, and pumped storage power plants. The output of each plant is relative to its own governor, and the setting point of AGC is obtained according to the optimal dispatch. The long-term AGC control performance based on MA is evaluated by a statistic experiment with 30-day stochastic load disturbance. Four types of controllers are simulated, that is, Q-learning, $Q(\lambda)$, DWoLF-PHC(λ), and HDC-VWPS. The statistic experiment results obtained under the impulsive perturbations and stochastic white noise load fluctuation are showed in Figures 10 and 11, respectively. Especially, $|\Delta f|$ and $|ACE|$ are the average values of the frequency deviation and ACE. CPS1, CPS2, and CPS are the monthly compliance percentages. The same weight of HDC-VWPS in each GSG is chosen, which has a more effective joint cooperation than other policies. As a result, a higher scalability and self-learning efficiency can be achieved.

Figure 10 shows that the HDC-VWPS in GSG1, compared with other methods, reduces CE by 11.48%~29.45%, $|\Delta f|$ by 0.237~0.0325 Hz, and ACE by 8.57%~90.37% and increases CPS1 by 1.03%~29.9%.

Figure 11 shows that the HDC-VWPS in GSG1, compared with other methods, reduces CE by 0.17%~20.24%, $|\Delta f|$ by 0.003~0.078 Hz, ACE by 45%~94% and increases CPS1 by 0.03%~4%. Similar results can be obtained in other GSGs.

It can be seen from the simulation results that the HDC-VWPS has stronger adaptability and better control performance than that of other three methods. In each GSG area, the win-lose criteria of the unit depend on the sign of the product of $\Delta(s_k, a_k)$ and $\Delta^2(s_k, a_k)$. By determining the "lose" or "win" of an agent, the corresponding variable learning rate is selected to obtain the optimal Q function through updating the Q value dynamically. Meanwhile, the variable quantity is determined in the mix strategy updating. Finally, the optimal mixed strategy is gained by the dynamic updating continuously. The results also demonstrate that the proposed

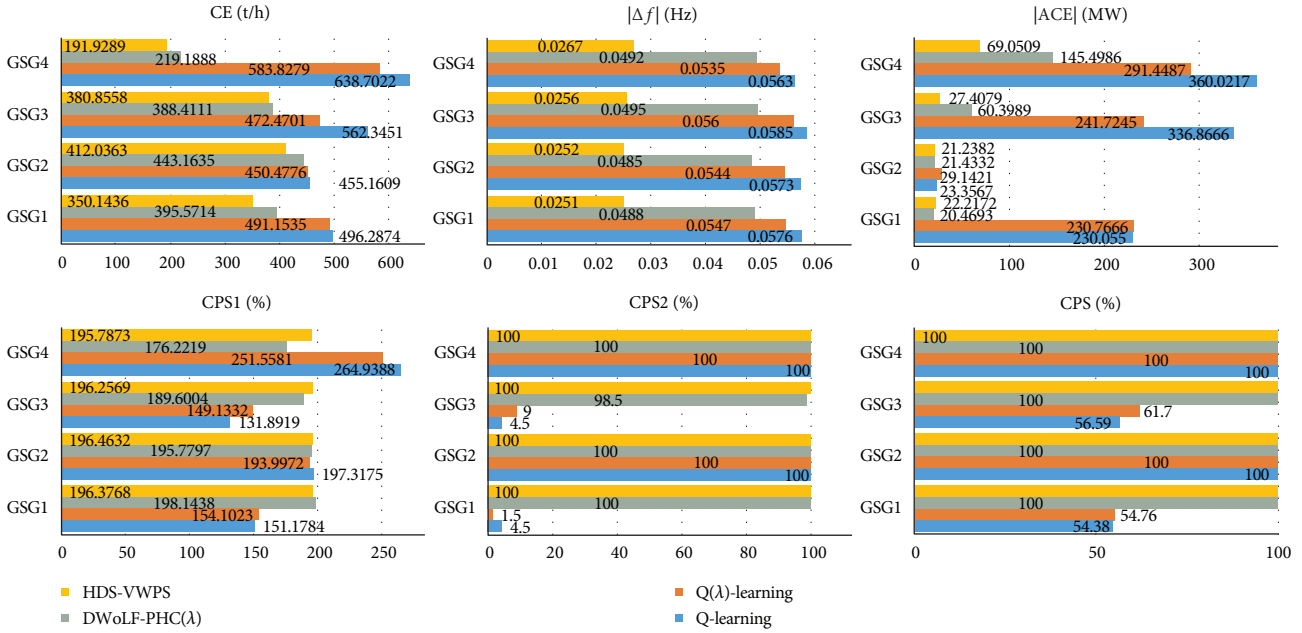


FIGURE 10: Statistic experiment results obtained under the impulsive perturbation in the Hubei power grid model.

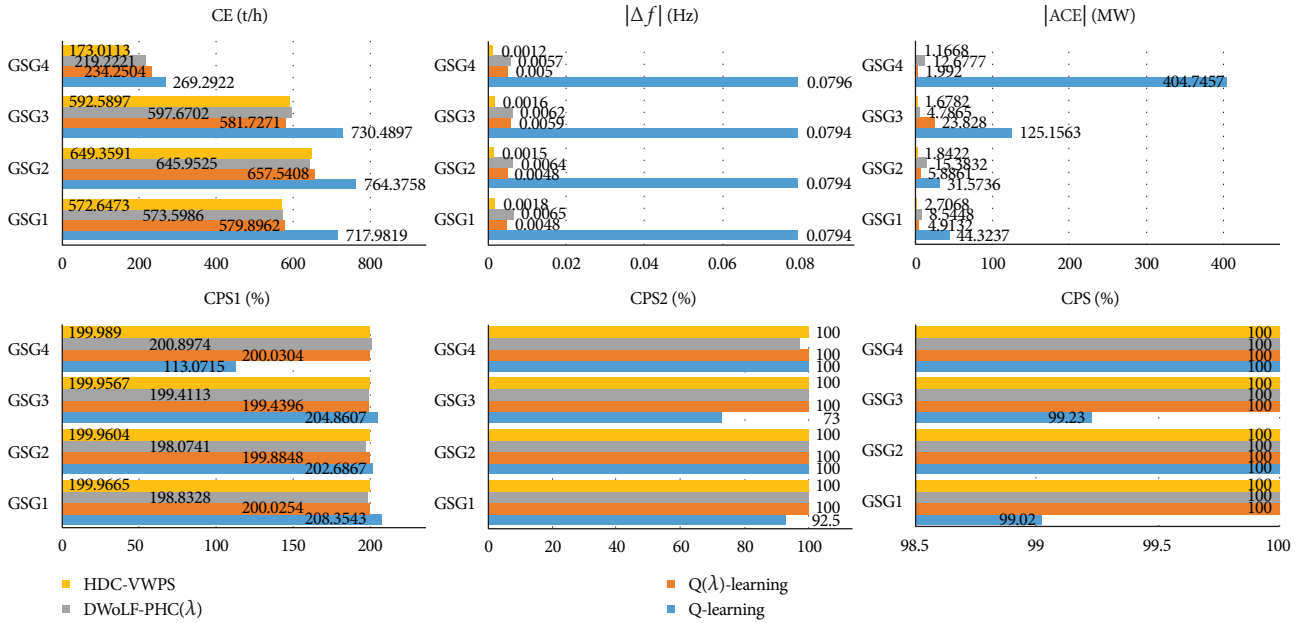


FIGURE 11: Statistic experiment results obtained under the white noise load fluctuation in the Hubei power grid model.

strategy can effectively reduce the CE and improve the utilization rate of new energy.

6. Conclusion

Based on the MAS-SCG theory, a novel HDC-VWPS method with new win-lose judgment criterion and eligibility trace is proposed to dynamically obtain the optimal total power and its optimal dispatch. Also, it can attenuate the stochastic disturbance caused by massive integration of new energy to the power grid.

Based on MAS-SG, a PDWoLF-PHC(λ) algorithm is proposed to solve the universality problem which usually a strict knowledge system is required for agents under the traditional MAS-SG system. It also solved the problem which the agents cannot accurately calculate the judgment criterion and converge to Nash equilibrium slowly in more than 2×2 games. Based on MAS-CC theory, the ramp time CC algorithm is used to allocate the total power command to each unit dynamically.

The simulation results verify the effectiveness of the proposed strategy using modified power system model in the IEEE two-area LFC and Hubei power grid model in

China. Compared with other four smart methods, the proposed one can satisfy the CPS requirements and improve the performance of the closed-loop system. Also, it can reduce the CE and maximize the utilization rate of energy.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

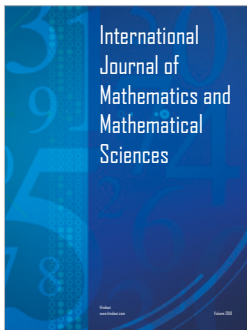
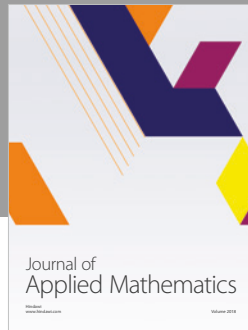
The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51707102 and 61603212).

References

- [1] O. Dobzhanskyi, P. Gottipati, E. Karaman, X. Luo, E. A. Mendrela, and A. M. Trzynadlowski, "Multilayer-winding versus switched-flux permanent-magnet AC machines for gearless applications in clean-energy systems," *IEEE Transactions on Industry Applications*, vol. 48, no. 6, pp. 2296–2302, 2012.
- [2] A. Milicua, G. Abad, and M. A. Rodriguez Vidal, "Online reference limitation method of shunt-connected converters to the grid to avoid exceeding voltage and current limits under unbalanced operation—part I: theory," *IEEE Transactions on Energy Conversion*, vol. 30, no. 3, pp. 852–863, 2015.
- [3] A. Noruzi, T. Banki, O. Abedinia, and N. Ghadimi, "A new method for probabilistic assessments in power systems, combining Monte Carlo and stochastic-algebraic methods," *Complexity*, vol. 21, no. 2, 110 pages, 2015.
- [4] Y. A.-R. I. Mohamed, "Suppression of low and high-frequency instabilities and grid-induced disturbances in distributed generation inverters," *IEEE Transactions on Power Electronics*, vol. 26, no. 12, pp. 3790–3803, 2011.
- [5] M. Mohammadi, A. Danandeh, H. Nasir Aghdam, and N. Ojaroudi, "Wavelet neural network based on islanding detection via inverter-based DG," *Complexity*, vol. 21, no. 2, 324 pages, 2015.
- [6] O. Elgerd and C. Fosha, "Optimum megawatt-frequency control of multi-area electric energy systems," *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-89, no. 4, pp. 556–563, 1970.
- [7] N. Ghadimi, "An adaptive neuro-fuzzy inference system for islanding detection in wind turbine as distributed generation," *Complexity*, vol. 21, no. 1, 20 pages, 2015.
- [8] N. Atic, D. Rerkpreedapong, A. Hasanovic, and A. Feliachi, "NERC compliant decentralized load frequency control design using model predictive control," in *2003 IEEE Power Engineering Society General Meeting (IEEE Cat. No.03CH37491)*, vol. 559, Toronto, Ontario, Canada, July 2003.
- [9] O. Abedinia, N. Amjady, and A. Ghasemi, "A new metaheuristic algorithm based on shark smell optimization," *Complexity*, vol. 21, no. 5, 116 pages, 2016.
- [10] A. Ahadi, N. Ghadimi, and D. Mirabbasi, "An analytical methodology for assessment of smart monitoring impact on future electric power distribution system reliability," *Complexity*, vol. 21, no. 1, 113 pages, 2015.
- [11] T. Yu, J. Liu, K. W. Chan, and J. J. Wang, "Distributed multi-step $Q(\lambda)$ learning for optimal power flow of large-scale power grids," *International Journal of Electrical Power & Energy Systems*, vol. 42, no. 1, pp. 614–620, 2012.
- [12] T. Yu, L. Xi, B. Yang, Z. Xu, and L. Jiang, "Multiagent Stochastic Dynamic Game for Smart Generation Control," *Journal of Energy Engineering*, vol. 142, no. 1, article 04015012, 2016.
- [13] L. Xi, T. Yu, B. Yang, and X. Zhang, "A novel multi-agent decentralized win or learn fast policy hill-climbing with eligibility trace algorithm for smart generation control of interconnected complex power grids ☆," *Energy Conversion and Management*, vol. 103, pp. 82–93, 2015.
- [14] M. Bowling and M. Veloso, "Multi-agent learning using a variable learning rate," *Artificial Intelligence*, vol. 136, no. 2, pp. 215–250, 2002.
- [15] R. S. Sutton and A. G. Barto, *Reinforcement Learning: an Introduction*, MIT Press, Cambridge, 1998.
- [16] B. Banerjee and J. Peng, "Adaptive policy gradient in multi-agent learning," in *Proceedings of the second international joint conference on Autonomous agents and multiagent systems - AAMAS '03*, pp. 686–692, New York, NY, USA, July 2003.
- [17] X. Zhang, H. Xu, T. Yu, B. Yang, and M. Xu, "Robust collaborative consensus algorithm for decentralized economic dispatch with a practical communication network," *Electric Power Systems Research*, vol. 140, pp. 597–610, 2016.
- [18] M. Ghavamzadeh, S. Mahadevan, and R. Makar, "Hierarchical multi-agent reinforcement learning," *Autonomous Agents and Multi-Agent Systems*, vol. 13, no. 2, pp. 197–229, 2006.
- [19] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3-4, pp. 279–292, 1992.
- [20] S. K. Gupta, K. Kar, S. Mishra, and J. T. Wen, "Collaborative energy and thermal comfort management through distributed consensus algorithms," *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 4, pp. 1285–1296, 2015.
- [21] G. Ray, A. N. Prasad, and G. D. Prasad, "A new approach to the design of robust load-frequency controller for large scale power systems," *Electric Power Systems Research*, vol. 51, no. 1, pp. 13–22, 1999.
- [22] O. I. Elgerd and H. H. Happ, "Electric energy systems theory: an introduction," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-2, no. 2, pp. 296–297, 1972.
- [23] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.
- [24] T. Yu, B. Zhou, K. W. Chan, L. Chen, and B. Yang, "Stochastic optimal relaxed automatic generation control in non-Markov environment based on multi-step $Q(\lambda)$ learning," *IEEE Transactions on Power Systems*, vol. 26, no. 3, pp. 1272–1282, 2011.



Hindawi

Submit your manuscripts at
www.hindawi.com

