A Reinforcement Learning based Power System Stabilizer for a Grid Connected Wind Energy Conversion System

Rahul Kosuru, Pengcheng Chen, and Shichao Liu, *Senior Member, IEEE*
Department of Electronic, Carleton University
Ottawa, Ontario, Canada

2020 IEEE Electric Power and Energy Conference (EPEC) | 978-1-7281-6489-2/20/$31.00 ©2020 IEEE | DOI: 10.1109/EPEC48502.2020.9320127

***Abstract*— When connecting renewable sources wind turbines to a power grid, low frequency oscillations caused by wind turbines may threaten the stability of the entire electrical power system. Power system stabilizers (PSSs) are used to damp the low frequency oscillations. However, these PSSs are usually designed based on small-signal models around a fixed wind speed and their performances could be degraded when wind speed varies in a real-time pattern. In this paper, a reinforcement learning (RL) based power system stabilizer is designed for a grid-connected double-fed induction generator (DFIG) based wind system to enable the online optimization of control gains when wind speed varies. In specific, the Q-learning based PSS is designed in the rotor-side controller of the DFIG based wind system. In this method, the active power change is defined as the state, and the control output of the rotor side controller (RSC) is used as the action. A grid-connected DFIG based wind system is simulated and the results show that the Q-learning based PSS can quickly adjust the control parameters online and damp the low frequency oscillation under a time-varying wind speed condition.**

Index Terms: Reinforcement Learning (RL), Rotor Side Controller RSC), Grid Side Controller (GSC), Power System Stabilizer (PSS), Doubly Fed Induction Generator (DFIG).

I. INTRODUCTION

Any power system network (grid) experiences a lot of disturbances, as there may be a great deal of variations at the load end or at the power generating stations (renewable or non-renewable).The critical control issues that need to be considered are power system stability and voltage regulation [1]. The importance of power system stability can be illustrated clearly by viewing [2], which presents a clear description of power system instability as the primary ground for any major black out. The main objective of any power system stabilizer (PSS) is to provide stable power to the grid and to improve the damping of the oscillations. Especially when connecting any renewable energy resource to the conventional grid, supplying a stable power is always challenging due to fluctuating frequencies and voltages.

In a grid-connected double-fed induction generator (DFIG) based wind system shown in Fig. 1, the decoupled controllers are used to control the rotor side active power and reactive power, the generator speed, and track the maximum power. They also help maintain DC voltage and regulate grid side active and reactive powers.

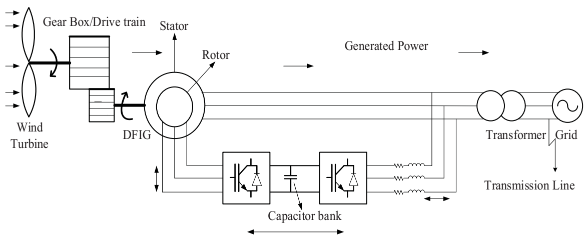

Fig. 1 A grid connected DFIG based wind system

Power system stabilizers can be applied at the output or the input of the controller. A PSS can be employed for any DFIG that is influenced by the network oscillations such as the rotor speed, the stator electrical power, and the network voltage frequency [3]. Various oscillations that are damped by using PSSs include electromechanical [4], inter-area power system oscillations [5], [6], network damping capability [7], oscillations caused by DFIG when integrated in a network [8]. It is clearly illustrated from these works a PSS is needed to damp the oscillations and to improve the stability of the grid integrated wind system.

Most of the conventional PSSs employed to wind energy systems are classical designs where the system is linearized around an operating point. There are also PSSs developed by using artificial intelligence (AI) techniques for the design and stability studies of non-renewable power systems. The most commonly used AI techniques for a PSS design are ANN [9] and Fuzzy Logic [10]. The usage of AI techniques to solve stability issues has been defined into three separate methodologies based on the techniques used, including supervised learning, unsupervised learning and reinforcement learning (RL) [11].

The current research work has used reinforcement learning in controlling the PSS to damp oscillations in power systems. One of the main reasons for using RL control method is its capability of adopting itself to evolving generation levels, load levels, operating uncertainties and response to arbitrary disturbances [12]. In [12], a RL method is used in online mode and applied to control a thyristor controller series capacitor (TCSC) aimed to damp power systems oscillations. RL controllers are deigned to stabilize the closed loop system after severe disturbances in [13]. A specific RL algorithm called Q-Learning is utilized to control and adjust the gain of conventional PSS in [14]. RL algorithms are also used in generation control and voltage and reactive power control [15-17]. However, regarding the implementations of reinforcement learning PSSs in grid-connected wind systems, only few works have been existing in the literature. In [18], the adaptive dynamic programming based PSS is designed for the DFIG wind turbine. In [19], an energy function-based controller has been extended to an adaptive version by using reinforcement learning.

In this work, a Q-learning based power system stabilizer is proposed to damp the low frequency oscillations in a grid-connected DFIG based wind system under a time-varying wind condition. While conventional power system stabilizers are designed by conducting small-signal analysis with respect to a fixed speed, time-varying wind speed conditions are commonly seen in the practical implementations. The designed Q-learning based power system stabilizer can find the optimal control gains in real-time corresponding to variable wind speeds. Simulation results verify the Q-learning based power system stabilizer can fast damp the low frequency oscillations in the system when wind speed varies.

II. Reinforcement Learning

Reinforcement Learning (RL) algorithms are focused on goal-directed learning from iterations. They are mainly used to solve closed-loop problems that use the actions from learning systems that influences the later inputs [11]. An RL algorithm consists of a discrete set of environment states S, a discrete set of agent actions A, and a set of scalar reinforcement signals R. Here, the agent is connected to its environment through action and agent receives current state as input and then the agent chooses an action to generate an output. With the obtained actions the state of the environment changes through the reinforcement signal which communicates with the agents. The final goal of any RL algorithm is to increase the long-run sum of values of the reinforcement signals which can learn over the time by trial and error method and solve the problem [20].

Among the various existing RL algorithms, the Q-learning algorithm is considered as simple and easier to implement due to its simple way for agents to learn and act optimally in controlled Markovian domains. The other main advantage of Q-learning algorithm is that it is exploration insensitive: Q values will converge to the optimal values, independent of how the agent behaves while the data is being collected [20]-[21]. With these advantages, the current paper is using Q-learning algorithm to provide inputs to the PSS for variable wind speeds. Assuming the best action is taken initially, the Q-learning optimal value function is taken as

$$Q(s,a) = R(s,a) + \gamma \sum_{s' \varepsilon S} T(s,a,s') \max Q(s',a') \quad (1)$$

In the above equation, $Q(s,a)$ is the expected discounted reinforcement of taking action a in state s. Once the action is taken the agent would be given a reward R, as per the effectiveness of the action by observing the resulting state S of the environment. Here T is the probability of action a applied to state s which change the state s to s'. For each action executed the Q values will converge with probability 1 to Q^* and when the Q values are nearly converged to their optimal values, the agent will act greedily by taking the action with the highest Q value, from this greedy policy the optimal action is determined. $\gamma\ (0 \leq \gamma \leq 0)$ is the discount factor which discounts the rewards exponentially in the future. Typically, an agent will look up the Q-memory look-up table which has state s and action a and is updated as per [20-21].

$$Q(s,a) = (1-\alpha).Q(s,a) + \alpha(r + \gamma \max(Q(s',a'))) \quad (2)$$

The parameter $\alpha(0 \leq \alpha \leq 1)$ in the above equation updates the Q-memory and affects the number of iterations [17].

III. Design of Q-learning Based PSS

For this paper, the PSS is designed based on the transformation technique. The block diagram for PSS with transformation technique is shown in the Fig. 2.

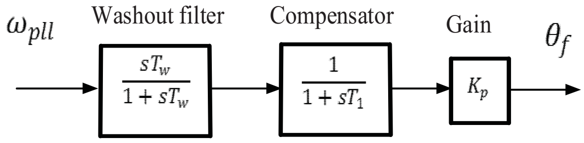

Fig. 2 Block diagram for PSS with transformation technique.

The signal θ_f is obtained from PLL frequency (ω_{pll}):

$$\theta_f = (K_t \omega_{pll}) \left(\frac{sT_w}{1+sT_w}\right)\left(\frac{1}{1+sT_1}\right) \quad (3)$$

The transformation is defined as [16]:

$$i_{dr}^* = \cos(\theta_f)\, i_{drref} - \sin(\theta_f)\, i_{qrref} \quad (4)$$

$$i_{qr}^* = \sin(\theta_f)\, i_{drref} + \cos(\theta_f)\, i_{qrref} \quad (5)$$

The developed PSS with transformation technique is implemented on the inner current controller of RSC on both d and q axis control loops. The block diagram with transformation technique is given in Fig. 3.

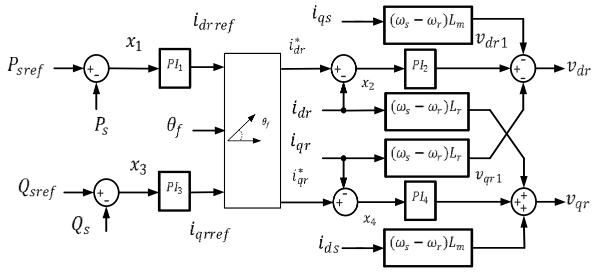

Fig. 3 RSC with the transformation-based PSS

The conventional Q-learning algorithm is used in this paper with small changes to make it feasible to the power system built. The main motive of the Q-learning algorithm used is to suppress low frequency oscillations generated by variable wind speeds and search for the optimal control action in each state. The Q-learning algorithm developed is used to optimize the controller parameter K_{p1} of the rotor side, which can suppress the low-frequency oscillation of the power system and obtain stable power output. To ensure the continuous exploration and avoid tending to local optimum, a pursuit algorithm based on the learning automata algorithm is utilized to the action policy. Initially actions are determined with the uniform probability distribution. When the Q-table is updated, the probabilities of actions are updated as follows.

$$\begin{cases} P_s^{k+1}(a_g) = P_s^k(a_g) + \beta\left(1 - P_s^k(a_g)\right) \\ P_s^{k+1}(a) = (1-\beta)P_s^k(a_g)\ \forall a \in A, a \neq a_g \\ P_{\tilde{s}}^{k+1}(a) = P_{\tilde{s}}^k(a)\ \forall a \in A, \forall \tilde{s} \in S, \tilde{s} \neq s \end{cases} \quad (5)$$

where $P_s^k(a_g)$ denotes the probability with action-state pair (a_g, s) is selected in iteration k, and β represent the action exploration rate. After the effect of this algorithm, Q^k tends to Q^* for a sufficiently large k and an optimal policy is obtained. The specific Q-learning based control parameter on-line optimization algorithm implemented in this paper is summarized as follows:

Algorithm 1 Q-learning based Control Parameter Optimization Algorithm in the Rotor Side

For each episode **do**
Initialize $Q^0(s,a) = 0, \forall a \in A, \forall s \in S$
Initialize $P_s^0(a) = \frac{1}{|A|}, \forall a \in A, \forall s \in S$
For each step of episode **do**
Choose a from s based on the current distribution $P_s(a)$
Take action a, observe r, s'
Update $Q(s,a)$ according to (2)
Update $P_s(a)$ according to (5)
$s \leftarrow s', a \leftarrow a'$
End for
End for

Since the problems considered in the reinforcement learning can be generally modeled as Markov decision process (MDP) models, we transform the adaptive parameter problem into a MDP model by designing a five-tuples (S, A, P, γ, r) as follows:

Design of state space S: S is a set of states which represent configurations of the system. It is assumed that all possible states are finite. In order to damp the low-frequency oscillation of the power system and obtain stable power output, we use active power ΔP as the state information.

The states of the MDP are described as follow: ΔP state-space (in per-unit value) is discretized into the 11 spaces: $(-\infty, -0.5]$, $(-0.5, -0.3]$, $(-0.3, -0.1]$, $(-0.1, -0.05]$, $(-0.05, -0.02]$, $(-0.02, -0.01]$, $(-0.01, -0.005]$, $(-0.005, -0.002]$, $(-0.002, 0.002]$, $(0.002, 0.01]$, $(0.01, \infty)$. We can see that the state distribution is unbalanced around zero, because the deviation of the active power at the rated value is also unbalanced with wind fluctuation.

Design of action space A: A is a set of actions which are executed by the agent to influence the environment. As mentioned before, the output of agent should be the controller parameter P_{k1} at rotor side. In this paper, we get the discrete action as follows: A= [-0.025 -0.02 -0.015 -0.01 -0.005 0.005 0.01 0.015 0.02 0.025].

$P: S \times A \rightarrow \theta(S)$ is the state transition function that shows the distribution of the next state s_{k+1} after executing an action a k to the environment with current state s_k . Since the parameter variation model is unknown, we can use temporal difference method to train the adaptive parameter policy. In this paper, we apply Q-learning algorithm to optimize the parameter of the rotor side controller.

Design of reward function $r(s,a)$: $r(s,a)$ is the function that maps the state-action pair (s,a) to a scalar which represents the immediate reward after applying an action a to the environment with state s. In this paper, $r_k = -k \times |P - P_{ref}|$. This means that the more the active power deviates from the reference value, the smaller the immediate reward, which prompts adjustment of the controller parameters so that the active power reaches the reference value.

$\gamma \in [0,1]$ is the discount parameter and determines the extent of the concern for the long-term reward. It is set as $\lambda = 0.1$ in this paper. The parameter β used in (4.1) is used for updating the probability distribution. The value of β is 0.1. The value of α is 0.2.

IV. Simulation Studies

The grid connected DFIG wind system shown in Fig. 1 is simulated and studied in Matlab R2016a/SimPower. A small signal model is developed to identify the state variables that affect the stability of the overall system. From small signal model, the sensitive variables observed are i_{dr},i_{qr},i_{qs}, x_7, v_{ds} and θ_{pll}, the eigen values are listed in table 1.

Table 1 Eigenvalues of the System without PSS

	$\lambda = \sigma \pm j\omega$	State Variables ΔX
λ_1, λ_2	$-18.13 \pm 507.09i$	I_{ds} and I_{qs}
λ_3, λ_4	$-24.38 \pm 326.16i$	I_{dr} and I_{qr}
λ_5, λ_6	$-0.51 \pm 0.70i$	ω_r and ω_t
λ_7	-0.06	V_{dc}
λ_8	-10.0	θ_{PLL}

The input given to the power system stabilizer is the frequency ω_{pll}. One of the main observations from small signal stability analysis is that, the state variables associated with the inner current control loop tend to move faster towards the real axis making the system unstable. This is observed with the large signal model as well by applying small faults. The shift of the eigen values with the new PSS implemented is noted in the Table 2 below:

Table 2 Eigenvalues of the System with PSS

	$\lambda = \sigma \pm j\omega$	State Variables ΔX
λ_4, λ_5	$-5.02 \pm 389.67i$	i_{dr} and i_{qr}
λ_6, λ_7	$-73.54 \pm 376.36i$	i_{qs} and x_7
λ_8, λ_9	$-64 \pm 37.08i$	V_{dc} and x_5
$\lambda_{12}, \lambda_{13}$	$-10.34 \pm 8.47i$	v_{qs} and θ_{pll}
$\lambda_{19}, \lambda_{20}$	$-0.0019 \pm 0.93i$	x_2, x_4
$\lambda_{11}, \lambda_{21}$	$-16.16 \pm 1.79i$	x_1 and θ_f

The highlighted eigenvalues of (E, A) x_1 and θ_f are newly added signal input from the PSS. The time constants for the washout filter and compensator are determined from the small signal model by observing the stability of the system.

It can be observed from the Fig. 4 and Fig. 5 that there is an increase in the damping of the oscillation for both active power and generator speed with a PSS implemented with transformation technique.

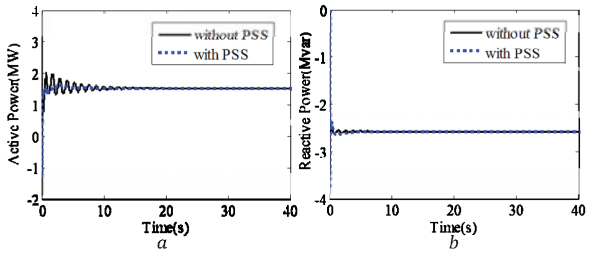

Fig. 4: (a). Active power and (b).reactive power with and without PSS

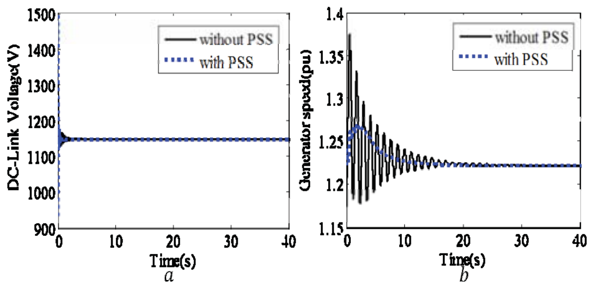

Fig. 5 (a). DC-link voltage and (b). Generator speed with PSS

The plot with transformation PSS with a three-phase short circuit fault condition triggered at t=30 second and ended at t=32 second is shown in Fig. 6. It can be observed that the oscillations are damped even during the fault condition. This can be more clearly observed from the power plot and in which the oscillations are damped the settling time is also very less. it can be concluded that both the PSS is working effectively during the fault condition.

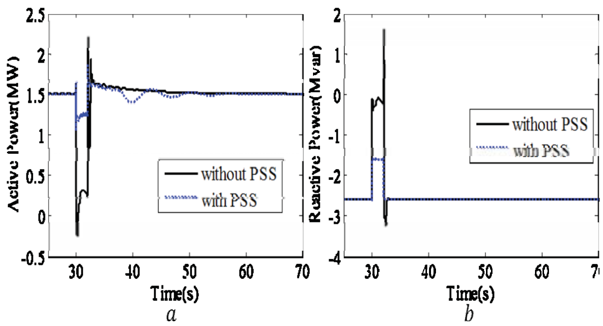

Fig. 6: (a). Active power and (b).reactive power with and without PSS during fault condition

Then, the introduced Q-learning algorithm is implemented as a Q-learning agent on the developed PSS under a time-varying wind speed. The algorithm model is applied at P_{sref}. Since the control objective is to stabilize the output of the power system with active and reactive power under low frequency oscillations, the active power change of the power system is defined as the state of the agent, and the control output of the controller is used as the action of the agent to train the appropriate control Strategy. During the operation of the system, the wind speed changes are shown in Fig. 7.

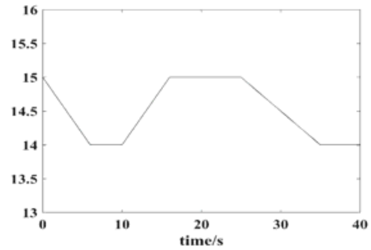

Fig. 7 Wind speed change

When the wind speed fluctuates, in the Q-learning agent training period, Fig. 8 to Fig. 10 show the simulation results of active power in the power system under the various numbers of iterations of the reinforcement-learning controller. It can be observed from these results that the reinforcement learning based PSS can quickly damp the low frequency oscillation and sustain a smooth active power performance after 12 iterations.

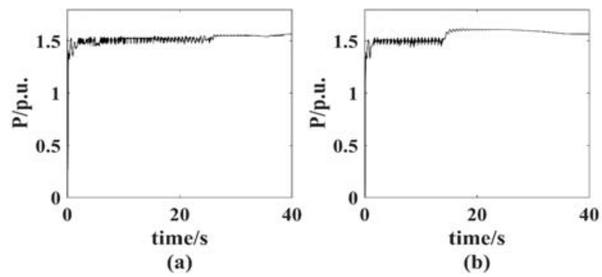

Fig. 8: Active power response of (a) iterate one times (b) iterate three times

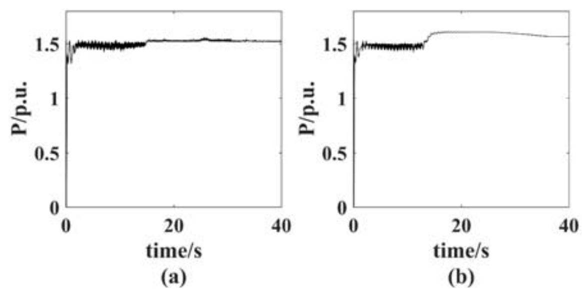

Fig. 9 Active power response of (a) iterate five times (b) iterate seven times

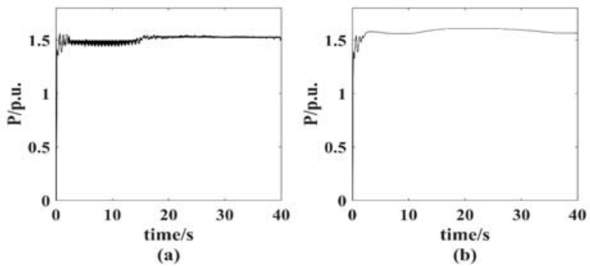

Fig. 10: Active power response of (a) iterate nine times (b) iterate twelve times

The trained reinforcement learning PSS is then implemented in a real-time fashion. It can be seen from Fig. 11 that the rotor angular speed (in p.u.) and active power are sustained even though the wind speed keeps changing.

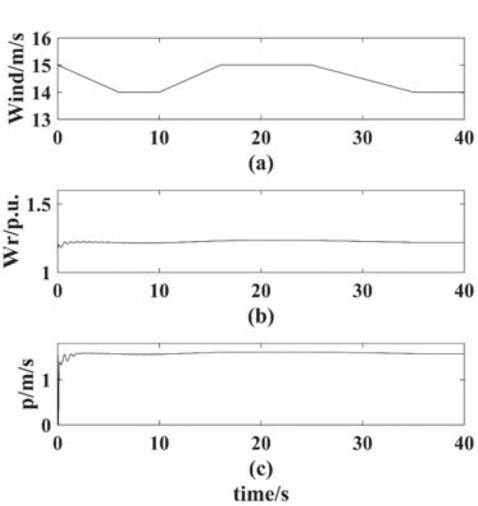

Figure 11 (a) Wind speed change diagram (b) Output speed diagram (c) Output active power diagram

V. CONCLUSIONS

This paper proposed a Q-learning based PSS for a grid-connected DFIG based wind system to suppress the low-frequency oscillations under a time-varying wind speed condition. It is observed from simulation results that the PSS can damp the oscillations under the faulty conditions and as well as normal conditions. With the help of Q-learning agent, the grid-connected system can quickly damp the low frequency oscillation under a continuous variable wind speed signal.

APPENDIX

Wind turbine parameters:

$S_{base} = 1.5; (MVA); H_t = 4.32\ s; K_{tg} = 1.11;$
$V_w = 15\ [m/s];\ V_{dcnominal} = 1150\ [V]$

RSC Parameters:

$K_{p1} = 1.25;\ K_{i1} = 300; K_{p2} = 0.6;\ K_{i2} = 8;$
$K_{p3} = 1.25;\ K_{i3} = 300,\ K_{p4} = 0.6;\ K_{i4} = 8$

GSC Parameters:

$K_{p5} = 8;\ K_{i5} = 400; K_{p6} = 0.83;\ K_{i6} = 5;$
$K_{p7} = 1.25; K_{i7} = 300\ , K_{p8} = 0.83;\ K_{i8} = 5$

PSS with frequency as input (transformation technique):

$K_f = 10;\ T_w = 2;$ And $T_1 = 1.$

Pitch controller:

$K_{pp} = 150; K_{pc} = 3; K_{ic} = 30; \theta_{max} = 27; \theta_{min} = 0$

VI. REFERENCES

[1] Hassan Bevrani, Masayuki Watanabe and Yasunori Mitani, "*Power System Monitoring and Control*". NY, John Wiley & Sons, Inc. 2014.

[2] Gregory S. Vassell, “Notheast Blackout of 1965,” *IEEE Power Engineering Review*, Jan 1991, pp. 4-8.

[3] F. Michael Hughes, Olimpo Anaya-Lara, Nicholas Jenkins, and Goran Strbac, “Control of DFIG-based wind generation for power network support,” *IEEE Trans. Power Syst.,* vol.20, no.4, pp. 1958-1966, Nov 2005.

[4] Katherine Elkington, Mehrdad Ghandhari and Lennart Soder, “Using power system stabilisers in doubly fed induction generators,” *Australasian Universities Power Engineering Conference,* 2008, pp. 1-6.

[5] D. P. Ke, C. Y. Chung and Yusheng Xue, “Controller design for dfig-based wind power generation to damp inter-area oscillation,” *Int. Conf. on Critical Infrastructure,* Sept. 2010, pp. 1-6.

[6] Bing Gong, Dewei Xu and Bin Wu, “Network damping capability of DFIG-based wind farm,” in *IEEE Energy conversion congress and exposition*, Sept.2010, pp. 4083-4090.

[7] Tossaporn Surinkaew and Issarachai Ngamroo, “Coordinated robust control of dfig wind turbine and PSS for stabilization of power oscillations considering system uncertainties,” *IEEE Trans. Sustain. Energy.,* vol.5, no.3, pp. 823-833, July 2014.

[8] Iswadi HR, D. John Morrow, Robert J. Best, “Small signal stability performance of power system during high penetration of wind generation,” *International Universities Power Engineering conference,* Sept. 2014, pp. 1-6.

[9] Y.Zhang, G.P.Chen, O.P.Malik, G.S.Hope, An Artificial Neural Network Based Adaptive Power System Stabilizer, *IEEE Transactions on Energy Conversion*, Vol. 8, No. 1, March 1993

[10] A. Hariri, O.P. Malik, A Fuzzy Logic Based Power System Stabilizer with Learning Ability, *IEEE Transactions on Energy Conversion*, Vol. 11, No. 4, December 1996

[11] R. Sutton, A. Barto, Reinforcement Learning: An Introduction, *IEEE Transactions on Neural Networks* 9 (5) (1998) 1054–1054.

[12] D. Ernst, M. Glavic and L. Wehenkel, "Power systems stability control: reinforcement learning framework," *in IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427-435, Feb. 2004.

[13] R. Hadidi and B. Jeyasurya, "Reinforcement Learning Based Real-Time Wide-Area Stabilizing Control Agents to Enhance Power System Stability," in *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 489-497, March 2013.

[14] R. Hadidi and B. Jeyasurya, “Reinforcement learning approach for controlling power system stabilizers,” *J. Electr. Comput. Eng*., vol. 34, no.3, pp. 99–103, Summer, 2009.

[15] T P Imthias Ahamed, P S Nagendra Rao, P S Sastry. A reinforcement learning approach to automatic generation control. *Electric Power Systems Research,* 2002, 63(1): 9-26.

[16] T. Yu, B. Zhou. A Novel Self-tuning CPS Controller Based on Q-learning method. *In the Proceedings of IEEE PES General Meeting*, July 20-24, 2008: 1-6.

[17] J G Vlachogiannis, N D Hatziargyriou. Reinforcement learning for reactive power control. *IEEE Trans. on Power Systems*, 2004, 19(3): 1317-1325.

[18] Y. Tang, H. He, J. Wen and J. Liu, "Power System Stability Control for a Wind Farm Based on Adaptive Dynamic Programming," *in IEEE Transactions on Smart Grid,* vol. 6, no. 1, pp. 166-177, Jan. 2015.

[19] R. Yousefian, R. Bhattarai and S. Kamalasadan, "Direct intelligent wide-area damping controller for wind integrated power system," *2016 IEEE Power and Energy Society General Meeting (PESGM)*, Boston, MA, 2016, pp. 1-5.

[20] L. P. Kaelbling, M. L. Littman, and A. W. Moore, “Reinforcement learning: A survey,” *J. Artif. Intell. Res*., vol. 4, pp. 237–285, Jan.–June 1996.

[21] T. Yu and W. Zhen, "A reinforcement learning approach to power system stabilizer," *2009 IEEE Power & Energy Society General Meeting*, Calgary, AB, 2009, pp. 1-5.