International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:2, No:1, 2008

Adaptive PID Controller based on Reinforcement Learning for Wind Turbine Control

M. Sedighizadeh, and A. Rezazadeh

***Abstract*—**A self tuning PID control strategy using reinforcement learning is proposed in this paper to deal with the control of wind energy conversion systems (WECS). Actor-Critic learning is used to tune PID parameters in an adaptive way by taking advantage of the model-free and on-line learning properties of reinforcement learning effectively. In order to reduce the demand of storage space and to improve the learning efficiency, a single RBF neural network is used to approximate the policy function of Actor and the value function of Critic simultaneously. The inputs of RBF network are the system error, as well as the first and the second-order differences of error. The Actor can realize the mapping from the system state to PID parameters, while the Critic evaluates the outputs of the Actor and produces TD error. Based on TD error performance index and gradient descent method, the updating rules of RBF kernel function and network weights were given. Simulation results show that the proposed controller is efficient for WECS and it is perfectly adaptable and strongly robust, which is better than that of a conventional PID controller.

***Keywords*—**Wind energy conversion systems, reinforcement learning; Actor-Critic learning; adaptive PID control; RBF network.

I. Introduction

AS a result of increasing environmental concerns, the impact of conventional electricity generation on the environment is being minimized and efforts are made to generate electricity from renewable sources. The main advantages of electricity generation from renewable sources are the absence of harmful emissions and the infinite availability of the prime mover that is converted into electricity. One way of generating electricity from renewable sources is to use wind turbines that convert the energy contained in flowing air into electricity. Various electromechanical schemes for generating electricity from the wind have been suggested, but the main drawback is that the resulting system is highly nonlinear, and thus a nonlinear control strategy is required to place the system in its optimal generation point.

Different intelligent approaches have successfully been applied to identify and nonlinearly control the WECS and other plants. For instance, Kanellos and Hatziargyriou [1], Yong-tong and Cheng-zhi [2] and Zhao-da *et al* [3] have suggested neural networks as powerful building blocks for nonlinear control strategies. The most famous topologies for this purpose are multilayer perceptron (MLP) and radial basis function (RBF) networks [4]. Mayosky and Cancelo [5] proposed a neural-network-based structure for Wind turbine control that consists of two combined control actions, a supervisory control and an RBF network-based adaptive controller. Sedighizadeh *et al* [6,7,8] suggested an adaptive controller using neural network frame Morlet wavelets together with an adaptive PI controller using RASP1 wavenets for Wind turbine control.

In this paper, the reinforcement learning is used to design of controller. This learning method unlike supervised learning of neural network adopts a 'trial and error' mechanism existing in human and animal learning. This method emphasizes that an agent can learn to obtain a goal from interactions with the environment. At first, a reinforcement learning agent exploits the environment actively and then evaluates the exploitation results, based on which controller is modified. It can realize unsupervised on-line learning without a system model [9-10]. Actor-Critic learning proposed by Barto et al is one of the most important reinforcement learning methods, which provides a working method of finding the optimal action and the expected value simultaneously [11]. Actor-Critic learning is widely used in artificial intelligence, robot planning and control, optimization and scheduling fields. Based on this analysis, in this paper a new adaptive PID controller based on reinforcement learning for WECS control is proposed. PID parameters are tuned on-line and adaptively by using the Actor-Critic learning method, which can solve the deficiency of realizing effective control for complex and time-varying systems by conventional PID controllers.

The next section presents details of the wind energy conversion system in this simulation. Section III describes the adaptive network algorithmic implementation. Then, the section IV introduces controller design steps. After that, the section V presents the simulation results and finally, the section VI explains conclusion.

Manuscript received January 5, 2008

The Authors are with Faculty of Electrical and Computer Engineering, Shahid Beheshti University, Tehran, 1983963113, Iran (phone: +98-21-29902290, fax: +98-21-22431804, e-mail: m_sedighi@sbu.ac.ir).

II. Wind Energy Conversion Systems

A. Wind Turbine Characteristics

Before discussing the application of wind turbines for the generation of electrical power, the particular aerodynamic characteristics of windmills need to be analyzed. Here the most common type of wind turbine, that is, the horizontal-axis type, is considered. The output mechanical power available from a wind turbine is [5].

$$P = 0.5\rho C_p (V_\omega)^3 A \quad (1)$$

Where ρ is the air density, A is the area swept by the blades, and V_ω is the wind speed. C_p is called the "power coefficient," and is given as a nonlinear function of the parameter λ

$$\lambda = \omega R / V_\omega \quad (2)$$

Where R is the radius of the turbine and ω is the rotational speed. Usually C_P is approximated as $C_P = \alpha\lambda + \beta\lambda^2 + \gamma\lambda^3$, where α, β and γ are constructive parameters for a given turbine.

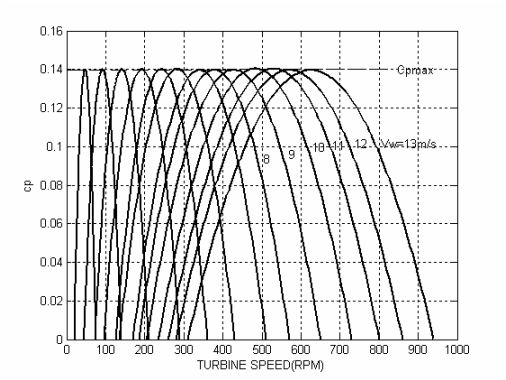

Fig. 1 Power coefficient C_p versus turbine speed [5]

Fig. 1 shows typical C_p versus turbine speed curves, with V_ω as a parameter. It can be seen that $C_{P\max}$, the maximum value for C_p, is a constant for a given turbine. That value, when replaced in (1), gives the maximum output power for a given wind speed. This corresponds to the optimal relationship λ_{opt} between ω and V_ω. The torque developed by the windmill is:

$$T_l = 0.5\rho\left(\frac{C_p}{\lambda}\right)(V_\omega)^2 \pi R^2 \quad (3)$$

Fig. 2 shows the torque/speed curves of a typical wind turbine, with V_ω as a parameter. Note that maximum generated power ($C_{P\max}$) points do not coincide with maximum developed torque points.

Superimposed to those curves is the curve of $C_{P\max}$. It can be seen that the maximum C_p (and thus the maximum generated power), and the maximum torque are not obtained at the same speed. Optimal performance is achieved when the turbine operates at the $C_{P\max}$ condition. This will be the control objective in the present paper.

B. Induction Generators and Slip Power Recovery

As wind technology progresses, an increasing number of variable speed WECS schemes are proposed. An interesting configuration among them is the one that uses grid-connected double-output induction generator (DOIG) with slip energy recovery in rotor, shown in Fig. 3 [8].

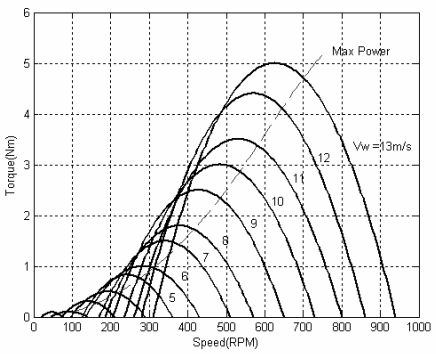

Fig. 2 Torque/speed curves (solid) of a typical wind turbine. The curve of $C_{P\max}$ is also plotted (dotted) [5]

Slip power is injected to the AC line using a combination of rectifier and inverter known as a static Kramer drive [5]. Changes on the firing angle of the inverter can control the operation point of the generator, in order to develop a resistant torque that places the turbine in its optimum (maximum generation) point.

Normally commutated inverter in DOIG's demands some reactive power. In addition, it recovers active slip power to the supply. Consequently, the absorbed reactive power by whole system raises leading to a lower power factor of the drive. Also, a rather large low-order harmonics is injected to the supply. The power factor of a converter can be improved using a forced commutation method. The amplitude of the harmonics can also be reduced [8]. The pulse width modulation (PWM) technique is one of the most effective methods in achieving the above goals. This method of improving the power factor eliminates the low-order harmonics. However, the amplitude of the high-order harmonics is increased, which can be easily filtered. To obtain a convenient performance, a current source type six valve converters from sinusoidal pulse width modulation (SPWM) techniques controls with three-mode switching signals is used [8].

In the SPWM technique, by changing the index modulation (m), the pulse width and the mean value of the inverter voltage are varied, thus the torque generated by DOIG is controlled. The torque developed by the generator/Kramer drive combination is [14]

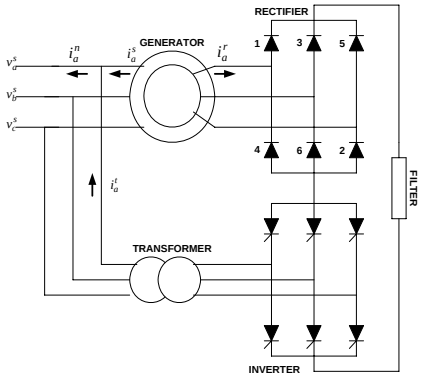

Fig. 3 Basic Power Circuit of a DOIG

$$T_g = \frac{3V^2 sR_{eq}}{\Omega_S[(sR_S + R_{eq})^2 + (s\omega_S L_{ls} + s\omega_S L_{lr})^2]} \quad (4)$$

Where

$$R_{eq} = f(s,m)$$
$$\omega = (1-s)\Omega_S \quad (5)$$

and

R_S : Stator resistance; L_{ls} : Stator dispersion inductance;

L_{lr} : Rotor dispersion inductance;

ω_s: Synchronous pulsation; Ω_s : Synchronous machine rotational speed; m: index modulation (All values referred to the rotor side).

C. Turbine/Generator Model

The dominant dynamics of the whole system (turbine plus generator) are those related to the total moment of inertia. Thus ignoring torsion in the shaft, generator's electric dynamics, and other higher order effects, the approximate system's dynamic model is

$$J\omega^{\bullet} = T_l(\omega, V_\omega) - T_g(\omega, m) \quad (6)$$

Where J is the total moment of inertia. Regarding (3) and (4), system's model becomes

$$\omega^{\bullet} = \frac{1}{J}(0.5\rho(\frac{C_P}{\lambda})(V_\omega)^2 \pi R^2 - T_g(\omega, m) \quad (7)$$

Where R_{eq} depends nonlinearly on the index modulation according to (5). C_P, λ, and V_ω also depend on ω in a nonlinear way (2). Moreover, it is well known that certain generator parameters, such as wound resistance, are strongly dependent on factors such as temperature and aging. Thus a nonlinear adaptive control strategy seems very attractive. Its objective is to place the turbine in its maximum generation point, in despite of wind gusts and generator's parameter changes. Thus the proposed control strategy, which consists of changing m to produce a generator's torque settles the turbine on the ω_{opt}, $T_{l(opt)}$ point [5]. The general form of (7) is $\omega^{\bullet} = h(\omega, m)$, where h is a nonlinear function accounting for the turbine and generator characteristics.

III. Adaptive PID Controller based on Reinforcement Learning

A. Controller Architecture

The structure of an adaptive PID controller based on Actor-Critic learning is illustrated in Fig. 4. It is based on the design idea of the incremental PID controller described by Eq. (8).

$$u(t) = u(t-1) + \Delta u(t) = u(t-1) + K(t)x(t) =$$
$$u(t-1) + k_I(t)x_1(t) + k_P x_2(t) + k_D x_3(t) = \quad (8)$$
$$u(t-1) + k_I(t)e(t) + k_P \Delta e(t) + k_D \Delta^2 e(t)$$

Where $x(t) = [x_1(t), x_2(t), x_3(t)] = [e(t), \Delta e(t), \Delta^2 e(t)]^T$; $e(t) = y_d(t) - y(t)$, $\Delta e(t) = e(t) - e(t-1)$ and $\Delta^2 e(t) = e(t) - 2e(t-1) + e(t-2)$ represent the system output error, the first-order difference of error and the second-order difference of error respectively; $K(t) = [k_I(t), k_P(t), k_D(t)]$ is a vector of PID parameters.

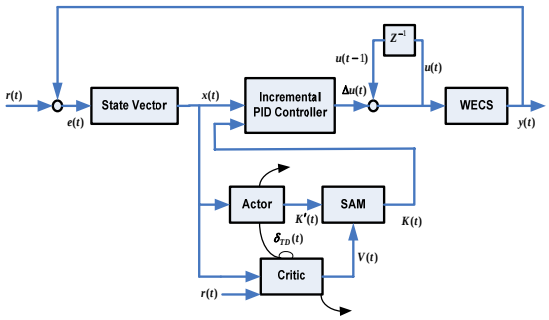

Fig. 4 Self-adaptive PID controller based on reinforcement learning

In Fig. 4, $y(t)$ and $y_d(t)$ are the desired and the actual system outputs respectively. The error $e(t)$ is converted into a system state vector $x(t)$ by a state converter, which is needed by the Actor-Critic learning part. There are three essential components of an Actor-Critic learning architecture, including an Actor, a Critic and a stochastic action modifier (SAM). The Actor is used to estimate a policy function and realizes the mapping from the current system state vector to the recommended PID parameters $K'(t) = [k'_I(t), k'_P(t), k'_D(t)]$ that will not participate in the design of the PID controller directly. The SAM is used to generate stochastically the actual PID parameters $K(t)$ according to the recommended PID parameters $K'(t)$ suggested by the Actor and the estimated signal $V(t)$ from the Critic. The Critic receives a system state vector and an external reinforcement signal (i.e., immediate reward) $r(t)$ from the environment and produces a TD error

(i.e., internal reinforcement signal) $\delta_{TD}(t)$ and an estimated value function $V(t)$. $\delta_{TD}(t)$ is provided for the Actor and the Critic directly and is viewed as an important basis for updating parameters of the Actor and the Critic. $V(t)$ is send to the SAM and is used to modify the output of the Actor. The effect of the system error and the change rate of error on control performance must be considered simultaneously during the design of the external reinforcement signal $r(t)$. Therefore, $r(t)$ is defined as

$$r(t) = \alpha r_e(t) + \beta r_{ec}(t) \tag{9}$$

Where α and β are weighted coefficients,

$$r_e(t) = \begin{cases} 0 & |e(t)| \le \varepsilon \\ -0.5 & otherwise \end{cases}$$

$$r_{ec}(t) = \begin{cases} 0 & |e(t)| \le |e(t-1)| \\ -0.5 & otherwise \end{cases}$$

and ε is a tolerant error band.

B. Actor-Critic Learning based on RBF Network

The RBF network is a kind of multi-layer feed forward neural network. It has the characteristics of a simple structure, strong global approximation ability and a quick and easy training algorithm [12]. On the other hand, the inputs of the Actor and the Critic are both the same state vector derived from the environment and their small difference is the difference in their outputs. Therefore, there is only one RBF network, as shown in Fig. 5. It is used to implement the policy function learning of the Actor and the value function learning of the Critic simultaneously. That is, the Actor and the Critic can share the input and the hidden layers of the RBF network. This working manner can decrease the demand for storage space and avoid the repeated computation for the outputs of the hidden units in order to improve the learning efficiency. The definite meaning of each layer is described as follows:

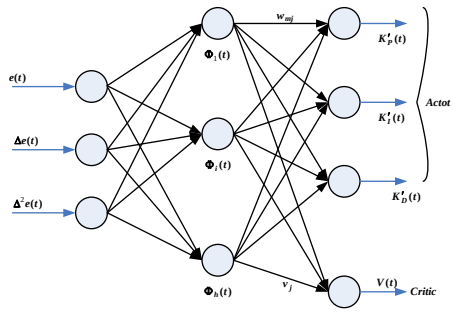

Fig. 5 Actor-Critic learning based on RBF network

Layer 1: input layer. Each unit in this layer denotes a system state variable x_i where i is an input variable index. Input vector $x(t) \in R^3$ is transmitted to the next layer directly.

Layer 2: hidden layer. The kernel function of the hidden unit of RBF network is adopted as a Gaussian function. The output of the j th hidden unit is

$$\Phi_j(t) = \exp\left(-\frac{\left\|x(t) - \mu_j(t)\right\|^2}{2\sigma_j^2(t)}\right),\ j = 1,2,...,h \tag{10}$$

where $\mu_j = [\mu_{1j}, \mu_{2j}, \mu_{3j}]^T$ and σ_j are the center vector and the width scalar of the j th hidden unit respectively, h the number of hidden units.

Layer 3: output layer. The layer is made up of an Actor part and a Critic part. The m th output of the Actor part, $K'_m(t)$ and the value function of the Critic part, $V(t)$ are calculated as

$$K'_m(t) = \sum_{j=1}^{h} w_{mj}(t)\Phi_j(t), \quad m = 1,2,3 \tag{11}$$

$$V(t) = \sum_{j=1}^{h} v_j(t)\Phi_j(t) \tag{12}$$

where w_{mj} denotes the weight between the j th hidden unit and the m th Actor unit, and v_j denotes the weight between the j th hidden unit and the single Critic unit.

In order to solve the dilemma of 'exploration' and 'exploitation', the output of the Actor part does not pass to the PID controller directly. A Gaussian noise term η_k is added to the recommended PID parameters $K'(t)$ coming from the Actor [9], consequently the actual PID parameters $K(t)$ are modified as Eq. (13). The magnitude of the Gaussian noise depends on $V(t)$. If $V(t)$ is large, η_k is small, and vice versa.

$$K(t) = K'(T) + \eta_k(0, \sigma_V(t)) \tag{13}$$

Where $\sigma_V(t) = \dfrac{1}{1 + \exp(2V(t))}$

The feature of Actor-Critic learning is that the Actor learns the policy function and the Critic learns the value function using the TD method simultaneously [12]. The TD error $\delta_{TD}(t)$ is calculated by the temporal difference of the value function between successive states in the state transition.

$$\delta_{TD}(t) = r(t) + \gamma V(t+1) - V(t) \tag{14}$$

Where $r(t)$ is the external reinforcement reward signal, $0 < \gamma < 1$ denotes the discount factor that is used to determine the proportion of the delay to the future rewards. The TD error indicates, in fact, the goodness of the actual action. Therefore, the performance index function of system learning can be defined as follows.

$$E(t) = \frac{1}{2}\delta_{TD}^2(t) \tag{15}$$

Based on the TD error performance index, the weights of Actor and Critic are updated according to the following equations through a gradient descent method and a chain rule.

$$w_{mj}(t+1) = w_{mj}(t) + \alpha_A \delta_{TD}(t) \frac{K_m(t) - K'_m(t)}{\sigma_V(t)} \Phi_j(t) \tag{16}$$

$$v_j(t+1) = v_j(t) + \alpha_C \delta_{TD}(t) \Phi_j(t) \tag{17}$$

Where α_A and α_C are learning rates of Actor and Critic respectively.

Because the Actor and the Critic share the input and the hidden layers of RBF network, the centers and the widths of hidden units need to be updated only once according to the following rules.

$$\mu_{ij}(t+1) = \mu_{ij}(t) + \eta_{\mu}\delta_{TD}(t)v_j(t)\Phi_j(t)\frac{x_i(t)-\mu_{ij}(t)}{\sigma_j^2(t)} \quad (18)$$

$$\sigma_j(t+1) = \sigma_j(t) + \eta_{\sigma}\delta_{TD}(t)v_j(t)\Phi_j(t)\frac{\|x(t)-\mu_j(t)\|^2}{\sigma_j^3(t)} \quad (19)$$

Where μ_{μ} and μ_{σ} are learning rates of center and width respectively.

IV. Controller Design Steps

The overall block diagram of controller is illustrated in fig. 6. The whole design steps of the proposed adaptive PID controller can be described as follows.

Step 1. Initializing parameters of Actor-Critic learning controller, including $w_{mj}(0)$, $v_j(0)$, $\mu_{ij}(0)$, $\sigma_j(0)$, μ_{μ}, μ_{σ}, α_C, α_A, γ, ε, α and β.

Step2. Detecting the actual system output $y(t)$, calculating the system error $e(t)$, constituting system state variables $e(t)$, $\Delta e(t)$ and $\Delta^2 e(t)$.

Step3. Receiving an immediate reward $r(t)$ from Eq.(9).

Step4. Calculating the Actor output $K'(t)$ and the Critic value function $V(t)$ from Eq. (11) and Eq.(12) at time t respectively.

Step5. Calculating the actual PID parameters $K(t)$ from Eq. (13) and consequently calculating the control output of PID controller $u(t)$ from Eq. (8).

Step 6. Applying $u(t)$ to the controlled plant and observing the system output $y(t+1)$ and the immediate reward $r(t+1)$ at the next sampling time.

Step 7. Calculating the Actor output $K'(t+1)$ and the Critic value function $V(t+1)$ from Eq. (11) and Eq. (12) at time respectively.

Step 8. Calculating the TD error $\delta_{TD}(t)$ from Eq. (8).

Step9. Updating the weights of the Actor and the Critic from Eq. (16) and Eq. (17) respectively.

Step10. Updating the centers and the widths of RBF kernel functions according to Eq. (18) and Eq. (19) respectively.

Step11. Judging whether the control process is finished or not. If not, then $t \leftarrow (t+1)$ and turn to Step2.

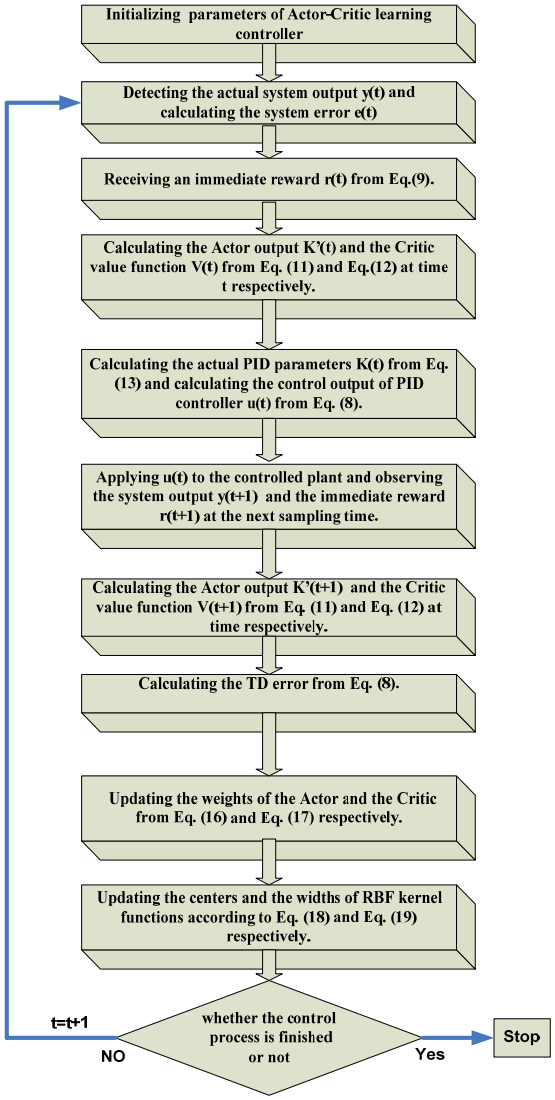

Fig. 6 Overall Controller block diagram

V. Simulation Results

Fig. 4 depicts the block diagram of the adaptive PID Controller Based on Reinforcement Learning for WECS control, while the dynamic of WECS is described by Eq. (7). For this case study, the desired signal $y_d(t)$ is optimal rotor speed ω_{opt}, actual output $y(t)$ is rotor speed ω and control signal $u(t)$ is index modulation (m). The optimum shaft rotational speed ω_{opt} is obtained, for each wind speed V_{ω}, and used as a reference for the closed loop. Note that wind speed acts also as a perturbation on the turbine's model. We applied the proposed adaptive PID controller and the conventional PID controller to track the optimal rotor speed signal. Sampling period $T_s = 0.0015s$ during the simulation. PID parameters of the conventional PID controller are set off-line as $k_P = 15$, $k_I = 35$ and $k_D = 10$ through the use of the Ziegler-Nichols tuning rule. The corresponding parameters for

the adaptive PID controller are set as follows, $\alpha = 0.67$, $\beta = 0.47$, $\varepsilon = 0.014$, $\gamma = 0.92$, $\alpha_A = 0.017$, $\alpha_C = 0.014$, $\eta_\mu = 0.032$ and $\eta_\sigma = 0.018$. The detailed simulation results are shown in Fig. 7. The Simulation results indicate that the proposed adaptive PID controller exhibits perfect control performance and adapts to the changes of parameters of the WECS. Therefore, it has the characteristics of being strongly robust and adaptable.

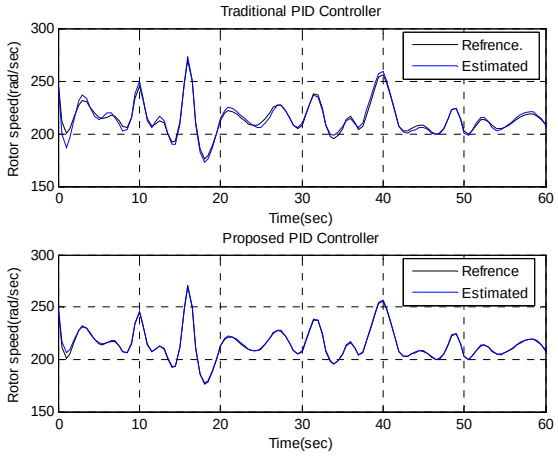

Fig. 7 Simulation Results

VI. Conclusion

Simulation results indicate that the proposed adaptive PID controller can realize stable tracking control for WECS. It is strongly robust for system disturbances, which is better than that of a conventional PID controller.

References

[1] Kanellos, F.D., Hatziargyriou, N.D., 2002. A new control scheme for variable speed wind turbine using neural networks. IEEE Power Engineering Society Winter Meeting, 1:

[2] You-tong, F., Cheng-zhi, F., 2007. Single neuron network PI control of high reliability linear induction motor for Maglev. Journal of Zhejiang University SCIENCE A, 2007, 8(3):408-411.

[3] Zhao-da, Y., Chong-guang, Z., Shi-chuan, S., Zhen-tao, L., Xi-zhen, W., 2003. Application of neural network in the study of combustion rate of natural gas/diesel dual fuel engine. Journal of Zhejiang University SCIENCE A, 2003, 4(2):170-174

[4] Haykin, S., 1994. Neural Networks, A Comprehensive Foundation. New York: Macmillan, 1994.

[5] Mayosky, M. A., Cancelo, G. I. E., 1999. Direct adaptive control of wind energy conversion systems using gaussian networks. IEEE Transactions on neural networks, 10(4): 898-906.

[6] Kalantar, M., Sedighizadeh, M., 2004. Adaptive Self Tuning Control of Wind Energy Conversion Systems Using Morlet Mother Wavelet Basis Functions Networks. *12th Mediterranean IEEE Conference on Control and Automation MED'04* , Kusadasi, Turkey.

[7] Sedighizadeh, M., Kalantar, M., 2004. Adaptive PID Control of Wind Energy Conversion Systems Using RASP1 Mother Wavelet Basis Function Networks. *IEEE TENCON 2004*, Chiang Mai, Thailand.

[8] Sedighizadeh, M., et al, 2005. Nonlinear Model Identification and Control of Wind Turbine Using Wavenets. Proceedings of the 2005 IEEE Conference on Control Applications Toronto, Canada, PP.1057-1062.

[9] WANG Xue-song, CHENG Yu-hu, SUN Wei. A Proposal of Adaptive PID Controller Based on Reinforcement LearningJ China Univ Mining & Technol 2007, 17(1): 0040–0044.

[10] Wang X S, Cheng Y H, Sun W. Q learning based on self-organizing fuzzy radial basis function network. *Lecture Notes inComputer Science*, 2006, 3971: 607–615.

[11] Barto A G, Sutton R S, Anderson C W. Neuronlike adaptive elements that can solve difficult learning control problems. *IEEETransactions on Systems, Man and Cybernetics*, 1983, 13(5): 834–846.

M. Sedighizadeh received the B.S. degree in Electrical Engineering from the Shahid Chamran University of Ahvaz, Iran and M.S. and Ph.D. degrees in Electrical Engineering from the Iran University of Science and Technology, Tehran, Iran, in 1996, 1998 and 2004, respectively. From 2000 to 2007 he was with power system studies group of Moshanir Company, Tehran, Iran. Currently, he is an Assistant Professor in the Faculty of Electrical and Computer Engineering, Shahid Beheshti University, Tehran, Iran. His research interests are Power system control and modeling, FACTS devices and Distributed Generation.

A. Rezazade was born in Tehran, Iran in 1969. He received his B.Sc and M.Sc. degrees and Ph.D. from Tehran University in 1991, 1993, and 2000, respectively, all in electrical engineering. He has two years of research in Electrical Machines and Drives laboratory of Wuppertal University, Germany, with the DAAD scholarship during his Ph.D. and Since 2000 he was the head of CNC EDM Wirecut machine research and manufacturing center in Pishraneh company. His research interests include application of computer controlled AC motors and EDM CNC machines and computer controlled switching power supplies. Dr. Rezazade currently is an assistant professor in the Power Engineering Faculty of Shahid Beheshti University. His research interests are Power system control and modeling, Industrial Control and Drives.