SHIRAZ UNIVERSITY

ICCIA 2011

Multi-agent Reinforcement Learning Design of Load-Frequency Control with Frequency Bias Estimation

F. Daneshfar, F. Mansoori and H. Bevrani, *Senior Member, IEEE*

***Abstract*—Conventional load-frequency control (LFC) systems use proportional-integral (PI) controllers. These controllers are designed based on a linear model and the nonlinearities of the system are not accounted for. Then they are incapable to gain good dynamical performance for a wide range of operating conditions. A control strategy for solving this problem in a multi-area power system is presented by using a multi-agent reinforcement learning (MARL) approach based on the frequency bias (*β*) estimation that genetic algorithm (GA) optimization is used to tune its parameters. This approach contains two agents in each control area, estimator agent and controller agent that communicate with each other.**

The proposed method does not depend on any knowledge of the system and finding area control error (ACE) signal based on the frequency biased estimation, improves the LFC performance. To demonstrate the capability of the proposed control structure, a three-control area power system simulation with two different scenarios is presented.

***Index Terms*— Multi-agent reinforcement learning; Load-frequency control; *β* estimation**

I. Introduction

FREQUENCY changes in large scale power systems are a direct result of the imbalance between the electrical load and the power supplied by system connected generators [1]. Therefore load-frequency control is one of the important power system control problems which there have been considerable research works for it [2-5].

However the conventional controllers are designed for a specific disturbance and if the nature of the disturbance varies, they may not perform as expected. Also most of them assume all model parameters are defined and measurable (fix) too, that in a real power system some parameters like β change with environment conditions and don't have constant values. Therefore, design of intelligent controllers that are more adaptive than conventional controllers is become an appealing approach [6-8].

Multi-agent reinforcement learning (MARL) is one of the adaptive and intelligent control techniques [9-13] that has found little attentions in the LFC design [9-11]. As it is based on learning, it can learn each kind of environment disturbances, and can easily scalable for large scale systems.

In this paper, a multi-agent reinforcement learning controller with β estimation is proposed. It has two agents in each control area that communicate with each other to control the whole system. The first agent (estimator agent) provides the ACE signal based on β parameter estimation and the second agent (controller agent) provides ΔP_c according to ACE signal received from estimator agent, using reinforcement learning, then it is distributed among the different units under control using fixed participation factors.

The above technique has been applied to the LFC problem in a three-control area power system as a case study. In the new environment, each control area consists of a number of generating companies (Gencos) and it is responsible for tracking its own load and performing the LFC task.

The organization of the rest of the paper is as follows. In Section 2, a brief introduction to multi-agent RL and LFC problem is given. In section 3, an explanation on how a load-frequency controller can work within this formulation is provided. In Section 4, a case study of three-control area power system and simulation results is discussed, finally the paper is concluded in Section 5.

II. Backgrounds

A. Multi-agent Reinforcement Learning

Reinforcement learning methods learn to solve a problem by interacting with a system and multi-agent reinforcement learning (MARL) is learning how a multi-agent system maps situations (x) to actions (a) so as to maximize a numerical reward signal (r) [14] while following policy $\pi(x,a)$. Policy is the way the agent maps the states to the actions [14]. In most RL methods, another term known as $Q^{\pi}(x,a)$ is defined which is the expected discounted reward while starting at state x and taking action a.

In fact the presentation of an MARL is as a tuple $< X, A_1, \dots, A_n, p, r_1, \dots, r_n >$ where n is the number of agents, X is the discrete set of environment states, A_i, $i = 1, \dots, n$ are the discrete sets of actions available to the agents, yielding the joint action set $A = A_1 \times \cdots \times A_n, p{:}\, X \times A \times X \to [0,1]$ is the state transition probability function, and $r_i : X \times A \times X \to R, i = 1, \dots, n$ are the reward functions of the agents. In the multi-agent case, the state transitions are the result of the joint action of all the agents,

F. Daneshfar, F. Mansoori, and H. Bevrani are with the Department of Electrical and Computer Engineering of University of Kurdistan. (Corresponding author e-mail: daneshfar@ieee.org).

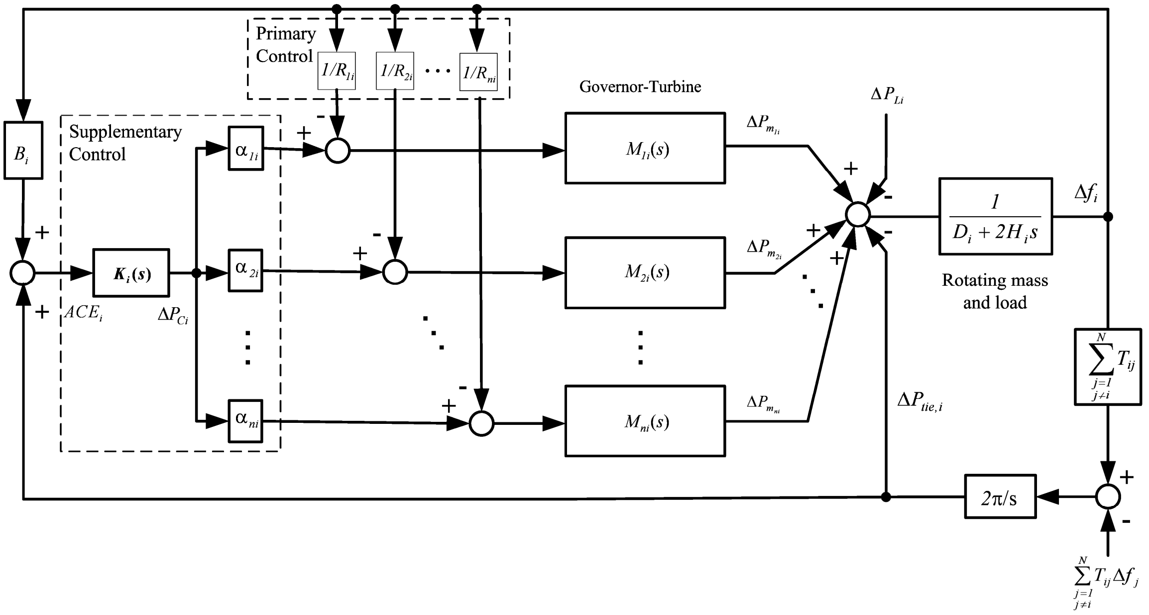


Fig. 1 LFC system with different generation units and participation factors in area *i* [16]

$a_k = [a_{1,k}^T, \dots, a_{n,k}^T], a_k \in A, a_{i,k} \in A_i$ (T denotes vector transpose). As a result, the rewards $r_{i,k+1}$ and the returns $R_{i,k}$ also depend on the joint action. The policies $h_i : X \times A_i \rightarrow [0, 1]$ form together the joint policy h. The Q-function of each agent depends on the joint action and is conditioned on the joint policy, $Q_i^h : X \times A \rightarrow R$ [15].

Agents will discover which joint action should be taken by interacting with the system and trying the different joint actions which may lead to the highest reward.

B. LFC Design Model

As mentioned, a three-control area power system used to examine the applicability of the proposed intelligent controller. The block diagram of a control area-*i* of the test system which includes *n* Gencos, is shown in Fig. 1[16]. Within a control area, following a load disturbance, the frequency of that area experiences a transient change, the feedback mechanism comes into play and generates appropriate rise/lower signal to the participating Gencos according to their participation factors (α_{ji}) to make generation follow the load. In the steady state, the generation is matched with the load, driving the tie-line power and frequency deviations to zero. Therefore the ACE for each control area can be expressed as a linear combination of tie-line power change and frequency deviation (according to (1)) [16].

$$ACE_i = \beta_i \Delta f_i + \Delta P_{tie-i} \tag{1}$$

III. Proposed Intelligent Control Design

In this section, an intelligent control design algorithm using MARL technique for a PI controller is presented. The design objective of the proposed method is to regulate the frequency in power system with various load disturbances and achieve a desirable control performance.

Fig. 2 shows the proposed model for area *i*. two kinds of intelligent agent have been used in this structure, controller agent and estimator agent. The estimator agent is responsible to estimate frequency bias parameter (β) and calculate ACE signal, however controller agent is responsible to find ΔP_c according to ACE signal using RL and GA.

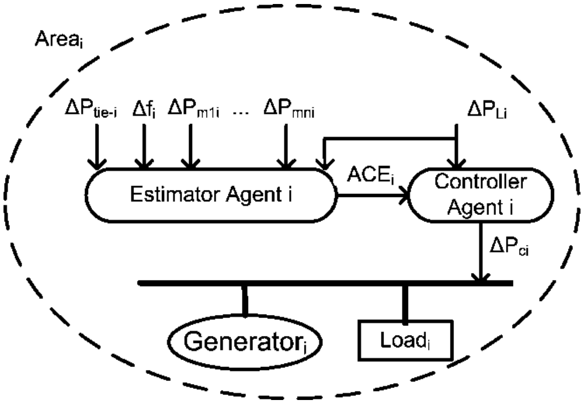


Fig. 2 The proposed model for area *i*

A. MARL Controller Agent

In this algorithm the average of ACE signal instances and ΔP_L instances over the LFC execution period are used as the state vector. The state vector consists of some quantities, which are normally available to the controller agent. For the algorithm presented in this paper, it is assumed that the set of all possible states X and actions A, is finite. Therefore the values of various quantities that constitute the state and action information should be quantized. The possible actions of the controller agent are the various values of ΔP_c, that can be demanded in the generation level within a LFC interval. ΔP_c is also discretised to some finite number of levels. Now, since both X and A are finite sets, a model for this dynamic system can be specified through a set of probabilities.

At each time step (as determined by the sampling time for

LFC action) the state input vector, x, to the LFC is determined then an action in that state is selected and applied on the model, the model is integrated for a time interval equal to the sampling time of LFC to obtain the state vector $\acute{x}$ at the next time step.

In this paper an algorithm same as the presented algorithm in [17] is used as follow.

At each instant (on a discrete time scale k), $k = 1, 2, \ldots$, the controller agent observes the current state of the system, x_k, and takes an action, a_k. If a sequence of samples is like, (x_k, x_{k+1}, a_k, r), k=1, 2 ... (k is the LFC execution period). Each sample is such that x_{k+1} is the (random) state that resulted when action a_k is performed in state x_k and $r_k = g(x_k, x_{k+1}, a_k)$ is the consequent immediate reinforcement. This sequence of samples (called training set) can be used to estimate Q^*. The specific algorithm that is used is as following. Suppose Q^k is the estimate of Q^* at kth iteration. Let the next sample be (x_k, x_{k+1}, a_k, r). Then Q^{k+1} is obtained as [17]:

$$Q^{k+1}(x_k, a_k) = Q^k(x_k, a_k) + \alpha[g(x_k, x_{k+1}, a_k) +$$

$$\gamma \max_{a\acute{\in}A} Q^k(x_{k+1}, \acute{a}) - Q^k(x_k, a_k)] \qquad (2)$$

where $0 < \alpha < 1$ is a constant called the step size of learning algorithm.

Here an exploration policy for choosing actions in different states is used. In this algorithm for each state x, actions are chosen based on a probability distribution over the action space. Initially a uniform probability distribution is chose (3). Let P_x^k denote the probability distribution over the action set for state vector x at the kth iteration of learning. That is, P_x^k (a) is the probability of choosing action a in state x at iteration k [17].

$$P_x^0(a) = \frac{1}{|A|} \forall a \in A \; \forall x \in X \qquad (3)$$

Using our simulation model the system is integrated for the next time interval and Q^k is updated to Q^{k+1} using (2) also at iteration k the probability of choosing the greedy action a_g in state x is slightly increased and the probabilities of choosing all other actions in state x are proportionally decreased like as follows [17],

$$P_x^{k+1}(a_g) = P_x^k(a_g) + \beta\left(1 - P_x^k(a_g)\right)$$

$$P_x^{k+1}(a) = P_x^k(a)(1-\beta) \qquad \forall a \in A, a \neq a_g \qquad (4)$$

$$P_y^{k+1}(a) = P_y^k(a) \;\; \forall y \in X, y \neq x$$

Where $0 < \beta < 1$ is a constant.

Here each state vector consists of two state variables: the average value of the ACE (the first state variable, x^1) and the ΔP_L (the second state variable, x^2) and the control action is the set point, ΔP_c. Since, The RL algorithms are based on finite number of states and actions, state and action variables will be discretised to finite levels using genetic algorithms optimization.

The next step is to choose an immediate reinforcement function, g. The reward matrix initially is full of zero, at each LFC execution period the average value of ΔP_L and average value of ACE signal are obtained, then according to the discritised values gained from GA, determine the state of the system, whenever the state is desirable (i.e. $|ACE|$ is less than ε) then reward function $g(x_k, x_{k+1}, a_k)$ is assigned a value zero. When it is undesirable (i.e. $|ACE| > \varepsilon$) then $g(x_k, x_{k+1}, a_k)$ is assigned a value -$|ACE|$ (all actions which cause to go to an undesirable state are penalized with a negative value) [17].

B. *Discretized Actions and States Using GA*

To quantize the state range and action rang using GA, each individual is a double vector (population type) that is quantized values of states and actions, with 406 variables between $[0:1]$ that consists of 400 variables for ACE signal and 6 variables for ΔP_c signal.

The start population size is equal to 30 individuals and it was run for 100 generations.

To find eligibility (fitness) of individuals, 6 variables are randomly chosen as discretized values of actions from each individual, then the model is run with these properties, and the individual's fitness is obtained from below:

$$Individual\ Fitness = \sum |ACE| / (simulation\ time)$$

Each individual that has the smallest fitness is the best one.

C. *Estimator Agent*

Finding ACE from equation (1) required to know frequency bias factor (β). Getting a good estimate of the area's β to improve the load-frequency control performance is a motivation for estimator agent to estimate β parameter and find ACE signal based on it. The conventional approaches in tie-line bias control use the frequency bias coefficient *-10B*, to offset the area's frequency response characteristic, β. But it is related to many factors and with *-10B* = β, the area control error (ACE) would only react to internal disturbances. To do that in each time, the estimator agent gets $\Delta P_{tie}, \Delta f, \Delta P_m, \Delta P_L$ as inputs, then calculates the β parameter and finds ACE signal according to that.

Equation (5) shows the per unit equation of the electromechanical power balance for the local control area.

$$\sum_{j=1}^{n} \Delta P_{mji}(t) - \Delta P_{Li}(t) - \Delta P_{tie,i}(t) = 2H_i \Delta \dot{f}_i(t) + D_i \Delta f_i(t) \qquad (5)$$

Also the below equation is concluded from (1):

$$\Delta P_{tie,i}(t) = ACE_i(t) - \beta_i \Delta f_i(t) \qquad (6)$$

Then according to (5), (6),

$$\sum_{j=1}^{n}\Delta P_{mji}(t)-\Delta P_{Li}(t)+\beta_i\Delta f_i(t)-ACE_i(t)=2H_i\Delta\dot{f_i}(t)+D_i\Delta f_i(t) \tag{7}$$

And,

$$ACE_i(t)=\sum_{j=1}^{n}\Delta P_{mji}(t)-\Delta P_{Li}(t)+(\beta_i-D_i)\Delta f_i(t)-2H_i\Delta\dot{f_i}(t) \tag{8}$$

The following equation is obtained for a moving average over a T-second interval to (8):

$$\overline{ACE_T}=\frac{1}{T}\{\sum_T\sum_{j=1}^{n}\Delta P_{mji}-\sum_T\Delta P_{Li}+(\beta_i-D_i)\sum_T\Delta f_i-2H_i(\Delta f(t_f)-\Delta f(t_i))\} \tag{9}$$

Since the values of β vary with system conditions, these model parameters would have to be updated regularly using a recursive least square (RLS) algorithm [18].

IV. Simulation Results

To demonstrate the effectiveness of the proposed control design, some simulations are carried out. In these simulations, the proposed controllers are applied to the three-control area power system model described in Section II and III (see Fig. 3), and will be tested for the various possible load disturbances. Also the results have been compared to a Bayesian Network (BN) based controller described in [13]. It is assumed that each control area includes three Gencos and its parameters are given in [19].

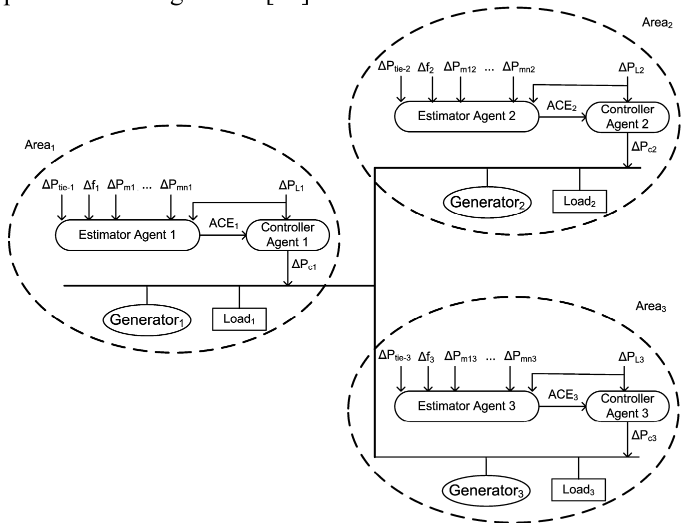


Fig. 3 The proposed model for the three-control area

Scenario 1: As the first test case, the following large load disturbances (step increase in demand) are applied to three areas:

$\Delta Pd_1=100MW;\ \Delta Pd_2=80MW;\ \Delta Pd_3=50MW;$

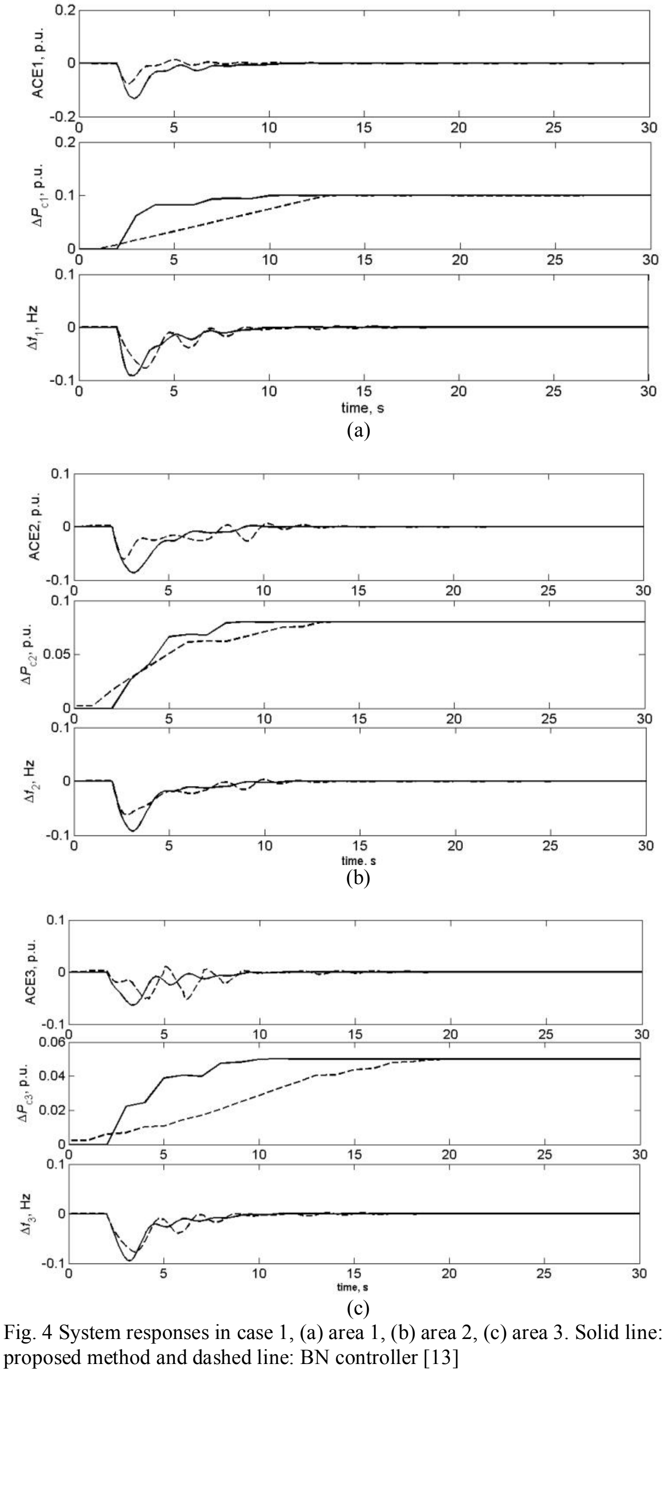
The frequency deviation (Δf) area control error (*ACE*) and control action (ΔP_c) signals of the closed-loop system are shown in "Fig. 4".

Scenario 2: Consider larger demands by areas 2 and 3, i.e.

$\Delta Pd_1=100MW;\ \Delta Pd_2=100MW;\ \Delta Pd_3=100MW;$

The closed-loop responses for each control area are shown in "Fig. 5".

The simulation results for the test system illustrate the effectiveness and capability of the proposed MARL based LFC scheme. It causes the ACE and frequency deviation of all areas are properly driven back to zero, also the generation control signal deviation (ΔP_c) change is low and it smoothly goes to the steady state and satisfies the system physical conditions well.

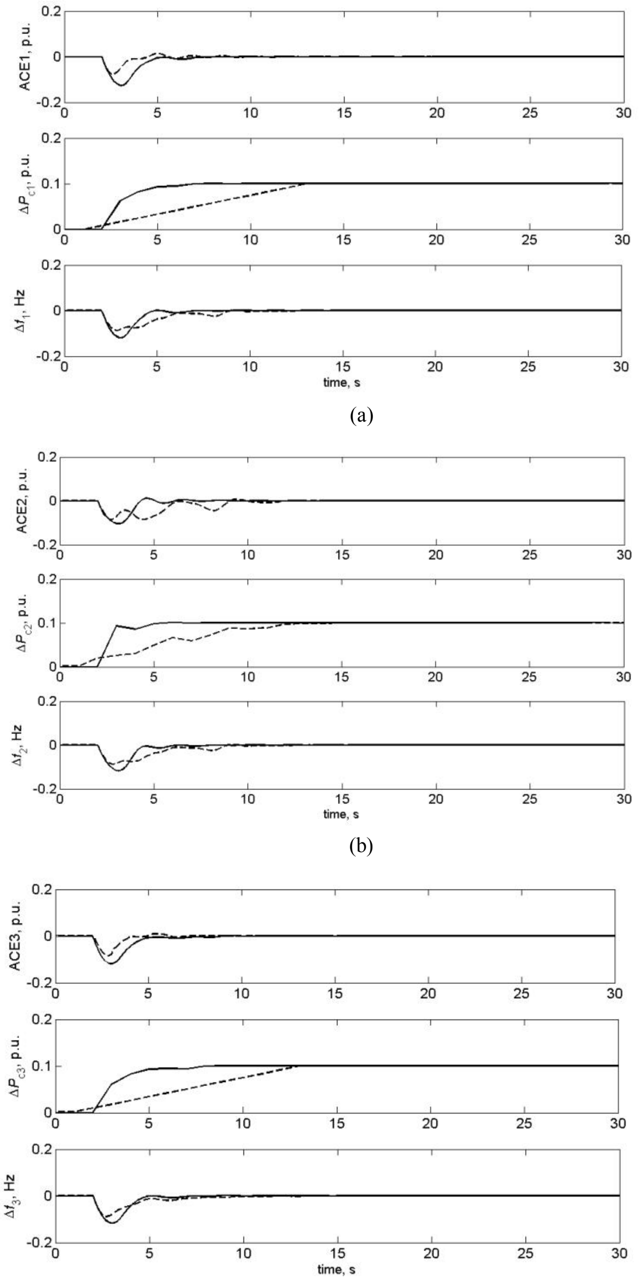


(a)

(b)

(c)

Fig. 4 System responses in case 1, (a) area 1, (b) area 2, (c) area 3. Solid line: proposed method and dashed line: BN controller [13]



(a)

(b)

(c)

Fig. 5 System responses in case 2, (a) area 1, (b) area 2, (c) area 3. Solid line: proposed method and dashed line: BN controller [13]

V. CONCLUSION

A new method for load-frequency control design using MARL has been proposed for a three-control area power system. The results show that the new algorithm presents a desirable performance. Two important features of the new approach, i.e. model independence from power system parameters and flexibility in specifying the control objectives, make it very attractive for frequency control practices.

REFERENCES

[1] N. Jaleeli, D. N., Ewart, and L.H. Fink, “Understanding automatic generation control,” *IEEE Trans. Power Syst*., vol. 7, 1992, pp. 1106-1112

[2] P., Ibraheem, Kumar, and P. Kothari, “Recent philosophies of automatic generation control strategies in power systems,” *IEEE Trans. Power Syst*., vol. 20, 2005, pp. 346-357

[3] T. M. Athay, “Generation scheduling and control,” *Proc. of the IEEE*, vol. 75, 1987, pp. 1592-1606

[4] M.L., Kothari, J. Nanda, D.P. Kothari, and D. Das, “Discrete mode AGC of a two area reheat thermal system with new ACE,” *IEEE Trans. Power Syst*., vol. 4, 1989, pp. 730-738

[5] R.R., Shoults, and J.A. Jativa, “Multi area adaptive LFC developed for a comprehensive AGC simulator,” *IEEE Trans. Power Syst*., vol. 8, 1991, pp. 541-547

[6] Y.L. Karnavas, and D.P. Papadopoulos, “AGC for autonomous power system using combined intelligent techniques,” *Electric power systems research*, vol. 62, 2002, pp. 225-239

[7] A. Demiroren, H.L., Zeynelgil, and N.S. Sengor, “Automatic generation control for power system with SMES by using neural network controller,” *Electr. Power Comp Syst.,* vol. 31, 2003, pp. 1-25

[8] Du. Xiuxia, and Li. Pingkang, “Fuzzy logic control optimal realization using GA for multi-area AGC systems,” *International Journal of Information Technology,* vol. 12, 2006, pp. 63-72

[9] F. Daneshfar, and H. Bevrani, “'Load-Frequency Control: A GA-based Multi-agent Reinforcement Learning'', *IET Generation, Transmission & Distribution*, vol. 4, no. 1, pp. 13-26, 2010.

[10] H. Bevrani., F. Daneshfar, and P. Daneshmand, ‘Intelligent Power System Frequency Regulations Concerning the Integration of Wind Power Units’ Chapter Book of “Wind Power Systems: Applications of Computational Intelligence” L. Wang et al. (Eds): Wind Power Systems, Green Energy and Technology, pp. 407–437, 2010

[11] H. Bevrani, F. Daneshfar, P. R. Daneshmand, T. Hiyama, “Reinforcement learning based multi-agent LFC design concerning the integration of wind farms”, In: *International Conference on Control Applications*, CD-ROM, IEEE, Yokohama, 2010

[12] H. Bevrani, F. Daneshfar, P. R. Daneshmand, “Intelligent Automatic Generation Control: Multi-agent Bayesian Networks Approach,” *International Conference on Control Applications*, CD-ROM, IEEE, Yokohama, 2010

[13] F. Daneshfar, H. Bevrani, and F. Mansoori, “Load-Frequency Control: a GA based Bayesian Networks Multi-agent System”, *Iranian Journal of Electrical & Electronic Engineering*, Vol. 7, No. 2, June 2011

[14] R. S. Sutton, and A.G. Barto, “Reinforcement learning: an introduction” (MA: MIT Press, 1998, Cambridge)

[15] L. Busoniu, R. Babuska, and B. De Schutter, “A comprehensive survey of multi-agent reinforcement learning,” *IEEE Trans Syst., Man., Cyber., Part C: Applications and Reviews*, vol. 38, 2008, pp. 156-172

[16] H. Bevrani, “Real power compensation and frequency Control”; *in: Robust power system frequency control,* 1st edn, Springer Press, 2009; 15-41

[17] T.P.I., Ahamed, P.S.N. Rao, and P.S., Sastry, “A reinforcement learning approach to automatic generation control,” *Electric Power Systems Research,* vol. 63, 2002, pp. 9–26

[18] E.K.P. Chong, and S.H. Zak, “An introduction to optimization”, (John Wiley & Sons Press, New York, 1996)

[19] H. Bevrani, Y. Mitani, and K. Tsuji, “Robust decentralised load-frequency control using an iterative linear matrix inequalities algorithm,” *IEE Proc. Gener. Transm. Distrib.*, vol. 3, 2004, pp. 347-354