# A Reinforcement Learning Model to Assess Market Power Under Auction-Based Energy Pricing

Vishnuteja Nanduri, *Student Member, IEEE*, and Tapas K. Das, *Member, IEEE*

***Abstract*—Auctions serve as a primary pricing mechanism in various market segments of a deregulated power industry. In day-ahead (DA) energy markets, strategies such as uniform price, discriminatory, and second-price uniform auctions result in different price settlements and thus offer different levels of market power. In this paper, we present a nonzero sum stochastic game theoretic model and a reinforcement learning (RL)-based solution framework that allow assessment of market power in DA markets. Since there are no available methods to obtain exact analytical solutions of stochastic games, an RL-based approach is utilized, which offers a computationally viable tool to obtain approximate solutions. These solutions provide effective bidding strategies for the DA market participants. The market powers associated with the bidding strategies are calculated using well-known indexes like Herfindahl–Hirschmann index and Lerner index and two new indices, quantity modulated price index (QMPI) and revenue-based market power index (RMPI), which are developed in this paper. The proposed RL-based methodology is tested on a sample network.**

***Index Terms*—Auctions, average reward stochastic games, competitive Markov decision processes (CMDPs), deregulated electricity markets, market power, reinforcement learning (RL).**

## I. Introduction

MARKET power is defined as the ability of a seller to maintain prices above competitive levels for a significant period of time. Market power (MP) of the participants is among the chief concerns of the designers of deregulated electric power markets. The participants could derive MP from different sources that are either inherent in the market design or are manifested through operational parameters. The design parameters that could potentially yield MP include market rules (e.g., capacity withholding, price caps, arbitrage), pricing and settlement mechanisms (e.g., LMP, types of auctions, transmission rights), and demand side bidding. Examples of operational parameters include types and sizes of available generation technologies (nuclear, coal, gas, hydro), transmission constraints, existing forward contracts, and load distribution in the network. The objective of this research was to develop a modeling approach and its solution strategy that would allow assessment of the impact of auction-based pricing strategies on MP.

Common forms of multiunit electricity auctions are uniform price auction, discriminatory auction, and second-price uniform auction. Cramton [1] notes that, under a uniform price auction, the generator supplying the last MWh meeting the market demand sets the market-clearing price. This gives the clearing generator an incentive to overstate costs. Such a tendency increases with the quantity supplied. Under a discriminatory auction, the bidders' incentive is to bid as close to the clearing price as possible since this auction rewards those that can best guess the clearing price. Typically, this favors larger companies that can spend more on forecasting and are more likely to set the clearing price as a result of their size. In sharp contrast, uniform pricing favors the smaller companies (or those with small unhedged positions going into the market). The small generators are able to free ride on the exercise of market power by the larger generators. The fundamental insight of the second-price uniform auction is that, since the price a generator receives is independent of his/her own offer price, marginal cost bidding can be induced as a weakly dominant strategy [2], [3]. Other important literature on electricity auctions include [4]–[8] and [9].

In POOLCO-type markets, bidding tends to be strategic in nature, where bidders seek opportunities to exercise market power. The extent of MP could vary under different auction mechanisms as well as network conditions, including transmission constraints [10], [11]. Other papers that address the issue of market power are Stoft [12], Mount [13], Nicolaisien *et al.* [7], Spear [14], Borenstein *et al.* [15], Bunn and Oliviera [16], and Hogan and Harvey [17].

The daily operation of a day-ahead (DA) electricity market with competing market participants and random demand realizations can be studied within the framework of nonzero sum stochastic games. Such games with Markovian probability structure lend themselves to be modeled as competitive Markov decision processes (CMDPs). In this paper, we present a CMDP model for a DA energy market in which prices and quantity settlements are accomplished through multiunit auctions. Since there are no available methods to obtain exact analytical solutions for CMDPs, a reinforcement learning (RL)-based approach is utilized, which offers a computationally viable tool to obtain approximate solutions. It is shown through detailed study of a sample network how the CMDP model and the RL-based solution approach can be used to examine market powers under various auction-based pricing strategies.

A critical aspect in the study of stochastic games is the reward mechanism, common forms of which are discounted reward, average reward, and total reward. In the repeated game environment of a DA market, the rewards from the bids are realized within a day, and hence, average reward appears to be the most appropriate reward criterion. The reward criterion significantly impacts the existence of Nash equilibria of stochastic

Manuscript received April 6, 2006; revised August 17, 2006. This work was supported in part by the National Science Foundation under Grant #ECS 0400268. Paper no. TPWRS-00203-2006.

The authors are with the University of South Florida, Tampa, FL 33620 USA.

Color versions of Figs. 2 and 3 are available at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TPWRS.2006.888977

0885-8950/$25.00 © 2007 IEEE

games. For example, discounted reward nonzero sum CMDPs are guaranteed to have at least one mixed strategy Nash equilibrium. However, the question whether, with respect to the average reward criterion, there always exist equilibrium solutions for nonzero sum games is still open [18]. However, the existence of equilibria under average reward is known for some special cases, including irreducible games [18]. We note that the Markov chain underlying the DA energy market, as described in Section II-A, is irreducible, since for any combination of bidding strategies, all states of the DA market form an ergodic class. Hence, the existence of an equilibrium for the DA market is guaranteed, though its uniqueness is not.

As mentioned earlier, no exact computational method exists for obtaining the Nash equilibria of a nonzero sum average reward stochastic game. The difficulty of computation arises from the complex nature of interactions among the competing decisions of the participants, probabilities of state transitions, and the reward structure. For stochastic games with a single state, which essentially becomes a matrix game, efficient solution methods exist [19]–[21]. In the recent years, algorithms based on a stochastic approximation method (known as reinforcement learning) have been presented to the literature to solve stochastic games. An approach can be found in [22], where it is shown that for each stage of a $\beta$-discounted nonzero sum stochastic game, there exists an equivalent matrix game. A reinforcement learning algorithm is presented in [22] that learns the elements of the matrices via a long simulation of the system. For a discount factor $0 < \beta < 1$, the convergence and optimality of the algorithm under the critical assumption of uniqueness of equilibrium are established. Such results for average reward games, however, are much more difficult to obtain in the absence of helpful convergence properties of the discounting factor. An approach similar to [22] was adopted in [23] in establishing equivalent matrix games for average reward irreducible nonzero sum stochastic games. A learning-based algorithm was also presented in [23] that incrementally constructs the equivalent reward matrices at each time stage of a game. In this paper, we use a similar approach, where a reward vector is constructed for each game participant for all state-action combinations via a learning process.

Energy markets have been widely modeled as static (one shot) games ([24]–[27]). A large body of literature addresses obtaining equilibrium of such static games via complementarity-based approach. A good exposition of this approach can be found in Hobbs and Helman [28] and Daxhelet and Smeers [29], which are also excellent sources for other references on this topic. Matrix game approaches to solving two- and three-player games arising in energy markets can be found in Lee and Baldick [30], [31]. An earlier version of the stochastic game model presented in this paper can be found in Ragupathi and Das [32]. Some of the learning-based approaches to power market games include [7], [16], and [33].

The rest of this paper is organized as follows. Section II presents our stochastic game theoretic model. A novel solution methodology using RL is presented in Section III. Section IV consists of a numerical study for assessment of market power. Concluding remarks appear in Section V.

## II. Game Theoretic Model for Day-Ahead Energy Market

In a DA energy market, multiple generators compete by bidding to supply power in a network. The market operates as follows. At the $t$th day, the generators submit their bid for the $(t+1)$th day. Actual loads and prices that comprise the system state of the $(t-1)$th day are used as the forecasted loads and prices for preparing the bids of the $(t+1)$th day. Using the price and quantity bids and the forecasted load conditions, the system operator solves an optimal power flow (OPF) problem to determine the least cost dispatch quantities and the bus prices for the $(t+1)$th day. The auction strategy in use provides a critical input in determining the prices for the buses with more than one generator. The actual realizations of random load on the $(t+1)$th day are used in settling the physical dispatch and spot prices. Thus, the state of the $(t+1)$th day is determined by the actual demand realizations on that day and the DA bids that were submitted on the $t$th day. Clearly, the knowledge of the system state on the $t$th day allows us to predict the state of the $(t+1)$th day. This process of system state transitions is shown to be a Markov chain. Also, since the generators bid noncooperatively with an aim of maximizing their individual market powers, the DA market is modeled as a nonzero sum stochastic game. It is well known that a stochastic game, for which the underlying process of state transition is a Markov chain, can be modeled as a CMDP [18]. The Markov chain and the resulting CMDP model for the DA market are presented next.

### A. Markov Chain and CMDP Model

In this section, we first establish the notation for the network characteristics and the system state. Thereafter, we define the stochastic processes and show how they can be modeled as a Markov chain and CMDP. Let $\mathcal{B}$ denote the set of buses in the network, and $\mathcal{B}_s \subset \mathcal{B}$ denotes the subset of supply buses (nodes). Let the number of generators at a supply bus $i \in \mathcal{B}_s$ be denoted by $N_i$, and $M$ denotes the number of loads in the network. Let $\mathcal{G}_i = \{1, 2, \ldots, N_i\}$ and $\mathcal{L} = \{1, 2, \ldots, M\}$ denote the set of generators at a supply bus $i$ and the set of loads in the network, respectively. Let $N = \sum N_i$, and $\mathcal{G} = \cup \mathcal{G}_i$. It is assumed that the DA energy market bids are submitted at the end of every day (after 12 A.M.) when the system state for the just completed day is known. Also, to keep the model exposition simple, we assume that only the generators bid in the DA market. The model can be extended to allow retailers to submit price-sensitive demand bids.

We define the system state for the $t$th day $X^t$ as the vector of realized loads $q^t$ and prices $p^t$ of the most recently completed day. Hence, $X^t = \{q^t, p^t\}$, where $q^t = (q_1^t, q_2^t, \ldots, q_{|\mathcal{B}|}^t)$ and $q_s^t$ denotes the realized hourly load quantity vector at the $s$th bus, $s \in \mathcal{B}$. Also, $p^t = (p_1^t, p_2^t, \ldots, p_{|\mathcal{B}|}^t)$, where $p_s^t$ represents the realized hourly price vector at bus $s \in \mathcal{B}$. Since both load $(q^t)$ and price $(p^t)$ that constitute the system state are continuous random variables, it is necessary to discretize them to allow formulation of a discrete stochastic model, as presented here. Let the range of possible values for both loads and prices be discretized in $Q$ and $P$ steps, respectively. Then the cardinality of the system state space $(\mathcal{E})$ is given as $|\mathcal{E}| = (Q^{24*M} \times P^{24*|\mathcal{B}|})$, where 24 accounts for the hours of a day. Note that our model

allows the level of discretization to be as refined as necessary for the desired modeling accuracy. That is, the values of $Q$ and $P$ can be chosen to be very large. However, the choice of finer discretization leads to a significantly larger state space and increased computational burden. Hence, this choice must be made keeping in mind a balance between accuracy and computation needs.

Now we can define the stochastic process for the state transition of the DA market as $\mathcal{X} = \{X^t : t \in \mathbf{Z}\}$, where $\mathbf{Z}$ is the set of integers. As discussed earlier in this section, the value of $X^t$ along with the bid submitted on the $t$th day dictate the system state of $(t+1)$th day, $X^{t+1}$. Clearly, the $\mathcal{X}$ process satisfies the Markov property. This along with other characteristics such as discrete and finite system states and time homogeneity assumption for the load realization process make the $\mathcal{X}$ process a Markov chain. The time homogeneity assumption ensures that the probability distribution governing the load realization process remains unchanged during the demand season being modeled. We note that obtaining a closed-form analytical characterization of the transition probabilities of the Markov chain is very difficult due to: 1) size of the system state vector and the state space, 2) size of the decision vector space, 3) OPF and related system constraints, and 4) auction-based price settlement. However, as presented later, our machine learning-based solution approach uses a simulation model for the DA market instead of an analytical model requiring the transition probability matrices. This simulation-based approach easily integrates solution methods for OPF-based dispatch and auction-based pricing. Thus, we are able to avoid the complexity of analytical estimation of the transition probabilities. This reduction of analytical complexity is a well-established advantage of the simulation-based machine learning approach for solving stochastic decision-making problems [22], [34]–[36]. The notation for decision vector space and the CMDP model is developed next.

Let the bid decision vector at the $t$th day be given by $D^t = \{\mathcal{D}_l^t : l \in \mathcal{G}\}$, where $\mathcal{D}_l^t$ is the decision vector of generator $l$ and given as $\mathcal{D}_l^t = (S_l^t)$. The element $S_l^t$ denotes the vector of bid parameters for all 24 h. To limit the cardinality of the decision space to a finite value, the bid parameters are also suitably discretized. The stochastic bidding process involves daily selection of bid parameters by the generators. We refer to this stochastic process as the *decision process*, denoted by $\mathcal{D} = \{D^t : t \in \mathbf{Z}\}$, where $D^t$ is the decision vector chosen on the $t$th day. Since the decision vectors $D_l^t$ are chosen by the generators in a noncooperative manner, the bidding scenario characterized by the joint process $\mathcal{X}$ and $\mathcal{D}$ is modeled as a stochastic game. This stochastic game is a CMDP [18].

Solution of the CMDP requires calculation of the rewards resulting from the bids submitted by the game participants. The rewards for the bidding decisions made on the $t$th day are determined from the price and quantity settlement of the $(t+1)$th day. This settlement is a function of the following: submitted bids, actual load realizations, OPF problem considerations, and auction strategy. Reward calculation is discussed next.

### *B. Calculation of Rewards by Solving Optimal Power Flow Problem*

Rewards are defined as the revenue (price $\times$ quantity supplied) obtained by the generators from the market. The prices and the quantities supplied at the buses are determined by the OPF solution of the network for a given load realization. For networks with buses having more than one generator, the prices are influenced by both the auction strategy and the supply quantities. The iterative solution process of the nonlinear OPF model, which is described next, determines the supply allocation to the buses that minimizes the total cost. Such least cost supply allocations vary for different auction strategies, since the prices paid to the generators at a bus for any given allocation may vary with the auction strategy in use. This indicates that the solutions of the OPF and the auction problems are intertwined and thus should be considered together.

OPF problem formulations generally maximize social welfare in the presence of demand-side bidding (using consumer benefit functions), which is not considered here. Also, since our model examines DA market with bidding for active power only, we use an AC-OPF formulation that minimizes the total cost of meeting the active power demand while considering the system constraints, including that of reactive power. The mathematical formulation given below is similar to that in [37]. The formulation is applicable under both uniform and second price uniform auctions. Modifications necessary for discriminatory auctions are discussed later

$$\min \sum_{j \in \mathcal{B}_s} f_1^j(P^j)$$

$$\text{subject to} \sum_{j \in \mathcal{B}_s} P^j - l - l(V,\theta) = 0 \tag{1}$$

$$\sum_{j \in \mathcal{B}_s} Q^j - \tilde{l} - \tilde{l}(V,\theta) = 0 \tag{2}$$

$$S_{y,z} \le S_{y,z}^{\max}, \quad \forall y \ne z \in \{\mathcal{B}\} \tag{3}$$

$$V_w^{\min} \le V_w \le V_w^{\max}$$

$$\forall w \in \{\mathcal{B}\}, \quad \mathcal{B} = \{\text{set of buses}\} \tag{4}$$

$$P_{\min}^j \le P^j \le P_{\max}^j, \quad \forall j \in \{\mathcal{B}_s\} \tag{5}$$

$$Q_{\min}^j \le Q^j \le Q_{\max}^j, \quad \forall j \in \{\mathcal{B}_s\}. \tag{6}$$

In the objective function equation, $f_1^j$ denotes the auction-based clearing price of active power at bus $j$, which serves as an input parameter to the optimization problem. This clearing price, which is received by all the generators at bus $j$, is determined as follows. $P^j$, the supply quantity allocated to bus $j$, is subdivided to the set of generators at the bus based on their bid prices. For uniform auction policy, the generator that supplies the last MW sets the clearing price $f_1^j$ for the bus. For second price uniform auction, the clearing price for the bus is set by the generator whose price is immediately below the clearing generator.

Under discriminatory auction, since every generator gets his/her bid price, the OPF objective function is modified to

$$\min \sum_{g \in \mathcal{G}} f_1^g(P^g)$$

where $\mathcal{G}$ is the set of all generators in the network, and $f_1^g$ denotes the bid price (equal to price received) corresponding to the active power supply allocation $P^g$ to generator $g \in \mathcal{G}$. The price $f_1^g$, which serves as an input parameter to the optimization

problem, is obtained as follows. Let the active power supply allocation to bus $j$ be $P^j$. This quantity is subdivided to the set of generators $\mathcal{G}_j$ at the bus (where $\cup_j \mathcal{G}_j = \mathcal{G}$) based on their bid prices, such that $\sum_{g\in\mathcal{G}_j} P^g = P^j$. Then, as per discriminatory auction, each generator receives a price $f_1^g$ corresponding to $P^g$ on his/her supply curve.

Constraint (1) in the OPF model ensures that all the active demand $(l)$ and the active transmission losses $(l(V,\theta))$ are met by the generators selected for dispatch at any given time (active power balance equation). The constraint (2) ensures that all the reactive demand $(\tilde{l})$ and the reactive transmission losses $(\tilde{l}(V,\theta))$ are met by generators selected for dispatch (reactive power balance equation). The term $S_{y,z}$ in (3) denotes the flow limit for the power transmitted from Bus $y$ to Bus $z$. Constraint (3) ensures that the maximum flow limit constraints in both directions are not violated. The constraint (4) is used to maintain the voltage limits for each Bus. Constraints (5) and (6) are used to maintain active and reactive power generation limits.

## III. CMDP Model Solution Strategy

In a power network with numerous buses and generators, the number of possible system states (represented by discretized values of load, $q^t$, and price, $p^t$) is very large. For example, in a 12-bus network, with loads and prices discretized in three steps, the number of system states is approximately $5 \times 10^5$. As a result, obtaining the matrices for transition probabilities and rewards associated with the CMDP model are very difficult. Since the RL approach uses a simulation model to capture system dynamics instead of transition probability matrices, it can handle very complex system structures. Note that, even for very complex systems, simulation models can be developed with relative ease knowing the probability distributions of the random variables that govern the system behavior. The dimensionality problem in RL, though much less pronounced compared to analytical approaches, still persists. However, new tools such as diffusion wavelet-based function approximation [35] are being developed by RL researchers to improve scalability of RL. For those who are unfamiliar with the topic of RL, a brief overview of how the competing agents learn their bidding strategies is given in the following subsection.

### A. Overview of Reinforcement Learning Approach

The theory of RL is founded on two important principles: Bellman's equation and the theory of stochastic approximation [38], [39]. Any learning model contains four basic elements:

1) system environment (*simulation model*);
2) learning agents (*market participants*);
3) set of actions for each agent (*action spaces*);
4) system response (*participant rewards*).

Consider a system with three competing market participants. At a decision-making epoch when the system is in state $s$, the three learning agents that mimic the market participants select an action vector $(a = (a^1, a^2, a^3)\epsilon\mathcal{A})$. These actions and the system environment (model) collectively lead the system to the next decision-making state (say, $s'$). As a consequence of the action vector $(a)$ and the resulting state transition from $s$ to $s'$, the agents get their rewards $(r^1(s,a,s'), r^2(s,a,s\prime))$, and $r^3(s,a,s'))$ from the system environment. Using these rewards, the learning agents update their knowledge base (R-values, also called reinforcement value) for the most recent state-action combination encountered $(s,a)$. The updating of the R-values is carried out slowly using a small value for the *learning rate*. This completes a learning step. At this time, the agents select their next actions based on the R-values for the current state $s'$ and the corresponding action choices. The policy of selecting an action based on the R-values is often violated by adopting a random choice, which is known as *exploration*, since this allows the agents to explore other possibilities. The probability of taking an exploratory action is called the *exploration rate*. Both learning and exploration rates are decayed during the iterative learning process. This process repeats and the agent performances continue to improve. For stochastic games with average reward, a sample greedy updating scheme for the R-value for the player $i$ after the visit to the state-action combination $(s,a)$ at the $n$th step (denoted as $R^i_{n+1}(s,a)$) can be given as

$$R^i_{n+1}(s,a) = (1-\alpha_n)R^i_n(s,a) + \alpha_n\left[(r^i(s,a,s') - \rho^i_n + \max_{b\in\mathcal{A}} R^i_n(s',b)\right] \quad (7)$$

where $\alpha_n$ is the learning rate at step $n$ (which is decayed after every step), and $\rho^i_n$ is a scalar for which current average reward of agent $i$ can be used. The current average reward can be obtained by dividing the running sum of rewards for agent $i$ by the number of decision steps encountered. In the RL algorithm used in this paper, the current average reward values are also learned to avoid large fluctuations. After continuing learning for a large number of steps, if the R-values for all state-action combinations converge, the learning process is said to be complete. The converged R-values are then used to find a stable bidding policy for each of the agents. A rationale for the above R-value updating scheme can be found in the reinforcement learning literature (Gosavi [40], Abounadi *et al.* [41]).

*1) Scalability of RL Approach:* Implementation of RL algorithms for networks with large state-action spaces suffers from computational difficulties in storing and updating of the $R$-values. This problem of scalability can be addressed by the use of a function approximation scheme via artificial neural networks (ANNs). Large state-action spaces can be suitably subdivided into smaller subsets, where every state-action combination in a subset can share a single $R$-value. The $R$-value is represented by a linear *neuron* with two layers (simplest possible ANN), one for input and the other for output. This is equivalent to piecewise linear approximation of a nonlinear function. That is, instead of directly learning a separate $R$-value for each state-action combination, the ANN weights are learned for each neuron representing a set of state-action combinations. An example of such an implementation can be found in [42]. A recent paper by Manfredi and Mahadevan [35] discusses a diffusion wavelet-based function approximation approach for R-values. Such efforts are likely to further improve the scalability of RL approaches.

### B. Average Reward RL Algorithm for CMDP Model

A schematic diagram of the steps of the RL algorithm is shown in Fig. 1. It has two nested layers. The outer layer implements the learning, while the inner layer implements the auction

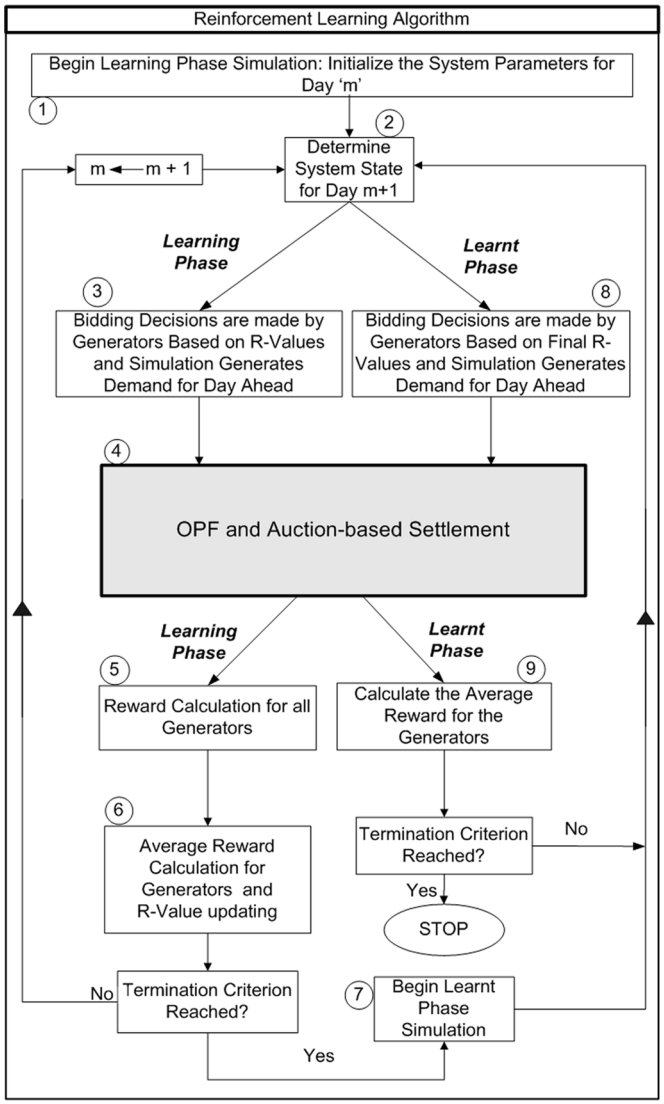


Fig. 1. Schematic for implementation of the RL algorithm.

and AC-OPF procedures. The outer layer (RL) is coded in C. For the inner layer, MATLAB code available from MATPOWER for solving AC-OPF is adopted. The learning phase (Box 1) begins by initializing the parameters for simulation of the DA energy market. System state for the DA is forecasted based on the realized demand and prices of the previous day (Box 2). For this forecasted system state, the bid decisions are made by the generators based on the existing R-values (Box 3). After the bid decisions are made, the actual DA demand is simulated, which along with the bid decisions are sent to the combined auction-OPF program (Box 4). The prices and the quantities corresponding to the least cost dispatch obtained from the OPF solution are then sent for generator reward calculations (Box 5). These rewards are then used to update the R-values (Box 6). If the convergence criterion for the R-values is not met, the program returns to Box 2; else, the learned phase of the algorithm begins at Box 7. In this phase, a learned bidding strategy based on the final R-values is employed and no further updating of the R-values are done. Completion of the learned phase implementation yields the average rewards for each generator.

Before presenting a mathematical statement of the complete RL algorithm, we next provide a brief explanation of the purpose of each step of the algorithm. The algorithm is derived from the well-established literature on single agent reinforcement learning approaches to stochastic dynamic programming problems (Gosavi [40], Kaelbling [43], Abounadi *et al.* [41]). A recent extension of single agent RL literature to multi-agent discounted reward games is presented by Hu and Wellman [22]. Average reward RL algorithms for stochastic games, similar in structure to the one presented here, have also appeared recently (Ragupathi and Das [32], Ravulapati *et al.* [44]). This stepwise explanation corresponds exactly to the steps of the algorithm.

1) Initialize the following variables:
   — R-values for each generator and all state-action combinations;
   — current average reward values for the generators;
   — input parameters for the two different learning rates and an exploration rate.
2) Assume that the iteration count $\mathrm{m} < \mathrm{MaxSteps}$, the system state is $l$. MaxSteps denotes the length of simulation run over which the generators learn; it is a termination criterion.
   a) Each generator chooses the bid (action) that has the highest reinforcement value $(R^{,,})$ with a probability of one minus the current exploration rate. All of the remaining actions are possible choices for an exploratory action and are assigned equal proportion of the exploration probability. A uniform (0,1) random variable is used for each generator to select an action according to the above probabilities.
   b) The system simulation is initiated to generate the DA demand. This demand information along with bids chosen by the generators are sent to the OPF program.
   c) The nonlinear OPF program, solved in conjunction with the auction rule, provides the optimal quantity allocation and bus prices.
   d) Rewards of the generators resulting from the bids are computed based on OPF solution outcomes. Also, the system state for the next day is determined from the information on demand and bus prices for the day.
   e) For each generator, the R-value for the most recent state-action combination is updated. Also updated are the generator's average reward values.
   f) Update the current system state and the iteration count.
   g) The learning and exploration parameters are updated. A well-known decay scheme proposed by Darken *et al.* [45] is used.
   h) If the current iteration count is less than MaxSteps, the algorithm continues at Step 2a); else, it moves to Step 3. (MaxSteps is the maximum number of iterations by which the R-values are expected to converge.)
3) Use the final R-values obtained from the learning phase to obtain the stable bidding strategies for all generators. Simulate the DA market with these strategies to assess average rewards for the generators.

*Algorithm:*

1) Let iteration or decision epoch count $m = 1$. Initialize for all generators $k \in \mathcal{G}$, the $R$-values as $R^k(s) = 0$ for all states $s \in \mathcal{E}$, and the average reward values as $\rho^k = 0$. Set the count for the number of visits to each state-action combination $(s, a^k)$ as $n(s, a^k) = 0$, where $a^k$ is an element of the set of all actions $A^k(s)$ available to generator $k$ in state $s$. It is assumed that the generators do not have knowledge of their competitors' actions. Also initialize the input parameters for two different learning rates $(\alpha_m, \beta_m)$ and the exploration rate $(\gamma_m)$.
2) Assume that at the iteration count $m < \text{MaxSteps}$, the system state is $s$. MaxSteps is the termination criterion.
   a) For each player $k \in \mathcal{N}$, with probability $(1 - \gamma_m)$, choose an action $a^k \in A^k(s)$ for which $R^k(s, a^k)$ is maximum. With a probability of $\gamma_m$, choose a random (exploratory) action from the set $A^k(s) \backslash a^k$. At $m = 1$ (i.e., in the first step), choose an action randomly since all the R-values are zeros.
   b) Start the system simulation by generating the demand for the day. Send the information on demand and generator bids to the OPF program.
   c) Solve the OPF problem for the network using the auction rule to obtain the optimal price and quantity allocations while satisfying all the system-related constraints. The constraints satisfied are demand and supply constraints, voltage constraints, thermal limit constraints, and the constraints of power flow, as explained in Section II-B.
   d) Use the optimal price and quantity allocations obtained by the OPF to determine the system state for the next decision epoch. Let the system state at that epoch be $s'$. Calculate $r^k(s, s', a^k)$, the reward for $k$th generator resulting from the actions $(a^1, \ldots, a^N)$ chosen by the generators 1 through $N$ in state $s$.
   e) Update $\forall k \in \mathcal{G}$ the R-values $(R^k(s, a^k))$ and the average reward $(\rho^k)$ as follows:

$$R^k_{\text{new}}(s, a^k) \leftarrow (1 - \alpha_m) R^k_{\text{old}}(s, a^k) + \alpha_m \left( r^k(s, s', a^k) - \rho^k_m + R^k_{\text{exp}}(s') \right) \quad (8)$$

where

$$R^k_{\text{exp}}(s') = \sum_{a^k \in A^k(s')} p^k(s', a^k) R^k(s', a^k)$$

and

$$p^k(s', a^k) = \begin{cases} (1 - \gamma_m), & \text{for } a^k = \text{greedy action} \\ \frac{\gamma_m}{A^k(s') - 1}, & \text{for other actions.} \end{cases}$$

Also

$$\rho^k_{m+1} = (1 - \beta_m) \rho^k_m + \beta_m \times \left[ \frac{m \rho^k_m + r^k(s, s', a^k)}{(m + 1)} \right]. \quad (9)$$

   f) Set $s \leftarrow s'$ and $m \leftarrow m + 1$.

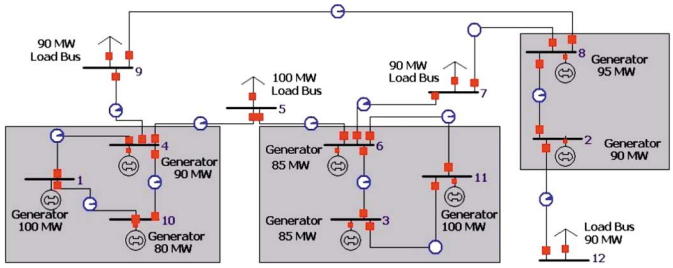


Fig. 2. One-line diagram of the sample power network.

   g) Update the learning parameters $\alpha_m$ and $\beta_m$ and exploration parameter $\gamma_m$ following the DCM [45] scheme given below:

$$\Theta_m = \left( \frac{\Theta_0}{1 + u} \right), \quad \text{where } u = \left( \frac{m^2}{\Theta_\tau + m} \right) \quad (10)$$

   where $\Theta_0$ denotes the initial value of a learning/exploration rate, and $\Theta_\tau$ is a large value (e.g., $10^4$) chosen to obtain a suitable decay rate for the learning/exploration parameters. Exploration rate generally has a large starting value (e.g., 0.4) and a quicker decay, whereas learning rates have small starting value (e.g., 0.01) and very slow decay rate. Exact choice of these values depends on the application (see [34] and [42]).
   h) If $m < \text{MaxSteps}$, go to Step 2a; else, go to Step 3.
3) Simulate the system with the final R-values, $\{R^k(s, a^k) : \forall a^k \in A^k(s), k \in \mathcal{G}, s \in \mathcal{E}\}$, and estimate the average reward for each generator. These are assumed to be stable rewards realized by the generators in the DA energy market.

## IV. Numerical Study

In this section, we present a detailed description of the 12-bus electric power network, which was used as a vehicle to implement the RL algorithm and perform numerical analysis. Also presented are the generator rewards and the corresponding market concentration and market power index values for each of the auction strategies. Results obtained from a designed statistical experiment, performed to analyze the significance of auction strategies, load, and congestion on market performance, are also presented.

### A. Sample Network

The sample network is adopted from networks available in MATPOWER 2.0 software [46]. A one-line representation of the network, developed using POWERWORLD software [47], is shown in Fig. 2. Some of the key features of the network (resistance, reactance, and long-term line ratings) are provided in Table I. Since the adopted network has single generator in each of the supply buses, in order to implement auction-based pricing that requires multiple generators at a bus, we made the following modifications. Subsets of the supply buses were connected by zero resistance lines in order to simulate the multiple generator bus scenario. In particular, we connected 1) Buses 1, 4, and 10; 2) Buses 3, 6, and 11; and 3) Buses 2 and 8, as depicted in the shaded portions of the one-line diagram. The modified network, which now effectively has a total of seven

TABLE I
KEY NETWORK FEATURES

| From Bus | To Bus | Resistance p.u | Reactance p.u | Long Term Rating (MW) |
|---|---|---|---|---|
| 1 | 4 | 0.0000 | 0.0006 | 250.0 |
| 4 | 5 | 0.0170 | 0.0920 | 250.0 |
| 4 | 10 | 0.0000 | 0.0006 | 250.0 |
| 1 | 10 | 0.0390 | 0.1700 | 150.0 |
| 5 | 6 | 0.0390 | 0.1700 | 150.0 |
| 3 | 6 | 0.0000 | 0.0006 | 300.0 |
| 3 | 11 | 0.0000 | 0.0006 | 300.0 |
| 6 | 11 | 0.0000 | 0.0006 | 300.0 |
| 6 | 7 | 0.0119 | 0.1008 | 150.0 |
| 7 | 8 | 0.0085 | 0.0720 | 250.0 |
| 8 | 2 | 0.0000 | 0.0005 | 250.0 |
| 12 | 2 | 0.0320 | 0.1610 | 250.0 |
| 8 | 9 | 0.0320 | 0.1610 | 250.0 |
| 9 | 4 | 0.0100 | 0.0850 | 250.0 |

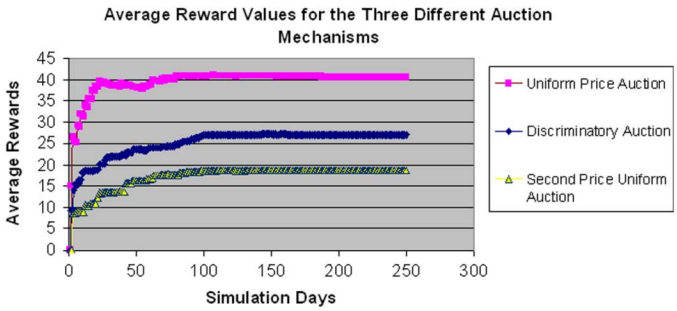


Fig. 3. Learning curves for Generator 1.

buses (three supply buses and four load buses), was validated using POWERWORLD. Fig. 3 shows the improvement of the rewards of Generator 1 during the learning phase under all three auction rules. It can be seen from the figure that the average rewards reached stable values after a relatively small number of simulation runs (approximately 250 days), which took about 2 h on a Pentium IV PC with a 2-GHz processor. It appears that the choices of the numerical values of the network and market parameters in the numerical example (e.g., transmission constraints, marginal costs, generator capacities, demand, bid markups) were such that a few state-action combinations were visited more frequently, resulting in a quick stabilization of the average reward. For networks with large number of buses and closely competing generators, the number of state-action combinations for which R-values are to be learned could be very large. This would take the learning process a much longer simulation run (in number of days simulated) to converge, requiring extended computing time. This can be tackled through use of cluster computing.

### B. Results From the Numerical Study

The primary objective of the numerical study was to demonstrate the ability of the RL-based methodology to assess market power of the participants. We organized the numerical study in three major stages. First, we obtained the prices and quantity allocations to the buses of the sample network under all three auction types. This was carried out for two separate (high and low) load scenarios (see Table II). In the second stage, we considered two levels of congestion (high and low) in addition to the two

TABLE II
LOAD (MWH) PARAMETERS

| Load Bus | High Load/Variance | Low Load/Variance |
|---|---|---|
| 5 | 100/20 | 80/20 |
| 7 | 90/20 | 60/20 |
| 9 | 90/20 | 60/20 |
| 12 | 90/20 | 60/20 |

TABLE III
AVERAGE PRICES ($/MWH) AND QUANTITIES (MWH) IN HIGH AND LOW LOAD SCENARIOS

| | | Generator 1 | | Generator 4 | | Generator 10 | |
|---|---|---|---|---|---|---|---|
| auction | case | price | quantity | price | quantity | price | quantity |
| Uniform | High Load | 13.1 | 19.6 | 13.1 | 84.0 | 13.1 | 22.9 |
| | Low Load | 13.1 | 5.5 | 13.1 | 64.9 | 13.1 | 13.9 |
| Discriminatory | High Load | 13.3 | 18.9 | 13.5 | 69.2 | 13.3 | 46.1 |
| | Low Load | 12.0 | 7.5 | 12.1 | 64.2 | 12.0 | 24.9 |
| Second Price | High Load | 12.9 | 16.7 | 12.9 | 81.2 | 12.9 | 30.3 |
| | Low Load | 12.6 | 3.4 | 12.6 | 70.5 | 12.6 | 21.2 |
| Marginal Cost ($/MW) & Capacity (MW) | | 12.9 | 100 | 10.6 | 90 | 12.0 | 80 |

| | | Generator 2 | | Generator 8 | |
|---|---|---|---|---|---|
| auction | case | price | quantity | price | quantity |
| Uniform | High Load | 13.1 | 65.9 | 13.1 | 67.9 |
| | Low Load | 13.1 | 63.9 | 13.1 | 40.4 |
| Discriminatory | High Load | 13.9 | 42.8 | 13.5 | 80.4 |
| | Low Load | 12.4 | 16.3 | 12.1 | 75.1 |
| Second Price | High Load | 14.0 | 74.7 | 14.0 | 72.9 |
| | Low Load | 12.7 | 61.2 | 12.3 | 56.8 |
| Marginal Cost ($/MW) & Capacity (MW) | | 10.6 | 90 | 11.3 | 95 |

| | | Generator 3 | | Generator 6 | | Generator 11 | |
|---|---|---|---|---|---|---|---|
| auction | case | price | quantity | price | quantity | price | quantity |
| Uniform | High Load | 13.0 | 25.5 | 13.0 | 12.7 | 13.0 | 72.8 |
| | Low Load | 13.1 | 14.0 | 13.1 | 5.4 | 13.1 | 51.3 |
| Discriminatory | High Load | 13.2 | 13.5 | 13.3 | 53.6 | 13.2 | 48.9 |
| | Low Load | 12.0 | 7.0 | 12.1 | 35.6 | 11.3 | 29.7 |
| Second Price | High Load | 13.0 | 16.7 | 13.0 | 13.8 | 13.0 | 65.4 |
| | Low Load | 12.7 | 5.0 | 12.7 | 3.3 | 12.7 | 38.0 |
| Marginal Cost ($/MW) & Capacity (MW) | | 12.8 | 85 | 12.9 | 85 | 11.3 | 100 |

levels of load. For all congestion-load combinations, two commonly used MP indexes along with a new index, introduced in this paper, were calculated for all auction types. The third stage involved examining the sensitivity of MP to factors such as load, congestion, and auction type using a new revenue-based market power index developed in this research. The sensitivity analysis was performed via a designed statistical (factorial) experiment.

*1) Stage 1—Average Price and Quantity Allocations:* Table III presents the average prices received and average quantities supplied by all the eight generators in the network for all six auction-load combinations. The loads were assumed to be normally distributed. The normality assumption was motivated by a statistical test of seasonal load data from the PJM market [32]. Table III is broken into three parts (for formatting reasons), where each part includes a subset of competing generators. The following observations can be made from the table—1) in most of the cases, discriminatory auction offered highest average prices to the generators followed by the uniform and second price uniform auctions, 2) generators in a discriminatory auction tend to bid higher as the network load increases, whereas in the uniform and second price uniform auctions, the bids do not change appreciably, and 3) the allocation of supply quantity among the generators supplying to a bus varies considerably with the auction rule. The table also contains, in the last row,

TABLE IV
AVERAGE QUANTITY WEIGHTED PRICES ($) IN HIGH AND LOW LOAD SCENARIOS

| Load | Auction | Average Quantity Weighted Price of Generators | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 4 | 10 | 2 | 8 | 3 | 6 | 11 |
| High Load | Uniform | 13.1 | 13.3 | 13.2 | 13.3 | 13.2 | 13.4 | 13.4 | 13.2 |
| | Discriminatory | 14.3 | 13.1 | 13.2 | 13.9 | 13.3 | 14.4 | 12.8 | 13.2 |
| | Second price | 12.9 | 12.7 | 12.6 | 12.7 | 12.8 | 12.9 | 12.8 | 12.7 |
| Low Load | Uniform | 13.1 | 13.3 | 13.2 | 13.3 | 13.2 | 13.4 | 13.4 | 13.2 |
| | Discriminatory | 12.5 | 11.6 | 12.9 | 13.3 | 11.7 | 12.5 | 11.7 | 12.8 |
| | Second price | 12.5 | 12.2 | 12.5 | 12.4 | 12.1 | 12.9 | 12.7 | 12.5 |
| | Marginal Cost | 12.9 | 10.6 | 12 | 10.6 | 11.3 | 12.8 | 12.9 | 11.3 |

the marginal costs of each generator. This provides a visual estimate of the extent of MP in various auction-load situations. A detailed study of different MP indexes is presented in Stage 2 of the numerical study.

We note that using average price as a measure of performance of a generator could be misleading, since in an auction-based market settlement, it is possible for a generator to receive a high price at the cost of a limited supply quantity. In averaging price over a large number of days of operation in the simulation run, several such high entries for price could raise the average price value. Use of such an average to assess MP could lead to wrong conclusions since it may not be appropriately reflected in the revenue earned by the generators. Hence, we used *average quantity weighted price* (AQWP) as a substitute for average price, which we defined as

$$\text{AQWP of Generator } j = \sum_{i=1}^{D} f_1^j(i) \frac{P^j(i)}{\sum_{i=1}^{D} P^j(i)}$$

where $D$ is the total number of days simulated, and $f_1^j(i)$ and $P^j(i)$ denote the price received and active power quantity supplied, respectively, by generator $j$ on day $i$. Table IV presents the average quantity weighted prices for all auction-load combinations that were considered in Table III. As expected, the average quantity weighted prices in most cases are considerably different from the average prices.

*2) Stage 2—Study of Market Concentration and Market Power:* In this section, we first present the definition of a market concentration indicator, known as Herfindahl–Hirschman index (HHI), and a new variant of HHI based on market performance. Also presented is the definition of Lerner index (LI), which measures market power. Thereafter, we define a new market power index named *quantity modulated price index (QMPI)*. The numerical values for all the indexes for 12 possible auction-load-congestion combinations are presented. Though the MP indexes considered here are primarily price/quantity based, as noted in [48], market power can also be manifested through lower quality of products and services and deferment of new entries to the market. MP can also be exercised by the buyers through demand side bidding, which is not considered here.

HHI, an index based on installed capacities, is defined as

$$\text{HHI} = \sum_{j=1}^{N} \left( \frac{P_{\max}^j}{\sum_{j=1}^{N} P_{\max}^j} \times 100 \right)^2$$

where $N$ is the total number of generators, $P_{\max}^j$ is the installed capacity of generator $j$, and the expression within the parenthesis is the percentage of market capacity owned by generator $j$. Clearly, the value of HHI of a monopoly would be 10 000, while the index value would be smaller for a larger number of market participants. Under Department of Justice/Federal Trade Commission (DOJ/FTC) standards, a market with HHI value less than 1000 is considered to be free of market concentration. Markets with HHI values between 1000 and 1800 are considered moderately concentrated, and values greater than 1800 indicate high market concentration. HHI, as defined above, is an ex-ante index, which is static and thus differs from the actual market concentration that corresponds to the dynamic bid-based supply allocations. Hence, such a market performance-based HHI (referred to hereafter as HHI*) could be obtained by simply replacing the installed capacities in the HHI expression by the generator supply quantities $(P^j)$ resulting from the bid-based settlement. The idea of a supply quantity-based HHI index was briefly mentioned in [49]–[51]. We note that, since $P^j \le P_{\max}^j$, it can be easily shown that

$$\sum_{j=1}^{N} \left( \frac{P_{\max}^j}{\sum_{j=1}^{N} P_{\max}^j} \times 100 \right)^2 \le \sum_{j=1}^{N} \left( \frac{P^j}{\sum_{j=1}^{N} P^j} \times 100 \right)^2.$$

That is, the HHI value serves as the lower bound for the possible HHI* values in markets with strategic bidding.

LI, a price-based MP index, is calculated for the network as an average over all generators as

$$\text{LI} = \frac{1}{N} \sum_{j=1}^{N} \frac{f_1^j - f_m^j}{f_1^j}$$

where $f_1^j$ denotes the price received for active power, and $f_m^j$ is the marginal cost of generator $j$. In a perfectly competitive market, where the demand curve is perfectly elastic, the LI equals zero. In a monopolistic market, the generator will use its market power to set its profit-maximizing output in the inelastic portion of the demand curve and charge a price greater than the marginal cost. In that case, the LI is greater than zero. In essence, inelastic demand implies large market power or vice versa. QMPI is proposed here as a modified version of LI that uses AQWP instead of average price. It is defined as

$$\text{QMPI} = \frac{1}{N} \sum_{j=1}^{N} \frac{\text{AQWP}^j - f_m^j}{\text{AQWP}^j}.$$

QMPI, though still a price-based index, indirectly also considers the quantity allocation. Thus, QMPI is not undesirably impacted by scenarios where generators receive high prices without a significant supply allocation. This is in contrast to purely quantity-based HHI and purely price-based LI.

The MP indexes (HHI, LI, and QMPI) were calculated for the sample network for each of the auction mechanisms under different load-congestion scenarios. The load-congestion scenarios are as shown in Table V. The numerical values of the MP indexes are given in Table VI. The HHI*, shown in column 4 of the table, are the index values calculated using average supply

TABLE V
LOAD AND CONGESTION PARAMETERS USED IN MARKET POWER STUDY

| Congestion Parameters | | | |
|---|---|---|---|
| From Bus # | To Bus # | Low Congestion | High Congestion |
| | | Long Term Rating M.V.A | Long Term Rating M.V.A |
| 1 | 4 | 250 | 250 |
| 4 | 5 | 250 | 100 |
| 4 | 10 | 250 | 250 |
| 1 | 10 | 250 | 250 |
| 5 | 6 | 150 | 150 |
| 3 | 6 | 300 | 300 |
| 3 | 11 | 300 | 300 |
| 6 | 11 | 300 | 300 |
| 6 | 7 | 150 | 150 |
| 7 | 8 | 250 | 100 |
| 8 | 2 | 250 | 250 |
| 12 | 2 | 250 | 250 |
| 8 | 9 | 250 | 150 |
| 9 | 4 | 250 | 100 |
| Loads(in MW) and Variances | | | |
| Load Bus # | High | Low | |
| | Load/ Variance | Load/ Variance | |
| 5 | 120/20 | 100/20 | |
| 7 | 110/20 | 90/20 | |
| 9 | 110/20 | 90/20 | |
| 12 | 110/20 | 90/20 | |

TABLE VI
HHI, LI, AND QMPI VALUES FOR THE NETWORK

| Auction | Load | Congestion | HHI* | LI | QMPI |
|---|---|---|---|---|---|
| Uniform | Low | Low | 1697.6 | 0.11 | 0.11 |
| Discriminatory | Low | Low | 1530.6 | 0.12 | 0.13 |
| Second Price | Low | Low | 1644.4 | 0.06 | 0.05 |
| Uniform | High | Low | 1497.2 | 0.12 | 0.13 |
| Discriminatory | High | Low | 1442.7 | 0.18 | 0.18 |
| Second Price | High | Low | 1521.2 | 0.08 | 0.08 |
| Uniform | Low | High | 1778.2 | 0.11 | 0.10 |
| Discriminatory | Low | High | 1587.8 | 0.15 | 0.15 |
| Second Price | Low | High | 1693.0 | 0.06 | 0.06 |
| Uniform | High | High | 1569.7 | 0.15 | 0.13 |
| Discriminatory | High | High | 1432.3 | 0.18 | 0.17 |
| Second Price | High | High | 1496.6 | 0.09 | 0.08 |

quantities of the generators. For all three auction rules and all load-congestion combinations, the HHI* values were between 1432 and 1778, indicating moderate market concentration. The HHI value computed from installed capacities was 1257, which supports the observation made earlier that HHI serves as the lower bound for HHI*. The LI and QMPI values show that a consistently high bid markup is exhibited under discriminatory auction, which is followed by uniform price auction and second price uniform auction. The bid markup for second price uniform auction was found to be 50%–60% lower than the other two auctions, which confirms the common belief that second price uniform auction induces truthful bidding. We also note from Table VI that the QMPI index values are not significantly different from the LI values. This can be attributed to the fact that, for the sample network, the differential between the average price and the AQWP was fairly small.

*3) Stage 3—Sensitivity of Market Power:* A generalized mixed full factorial designed experiment was set up to study the sensitivity of market power to factors such as auction type, load, and congestion. Readers unfamiliar with the subject of experimental design are referred to an excellent easy-to-follow treatment of this topic in [52]. We defined a new *revenue-based measure* of MP, named *revenue-based market power index* (RMPI), for use as the response variable in the designed experiment. RMPI is a measure of network profit (net revenue minus net marginal cost), which is defined as

$$\text{RMPI} = \sum_{j=1}^{N} \bar{f}_1^j \bar{P}^j - \sum_{j=1}^{N} \left( f_m^j P^j \right)$$

where $\bar{f}_1^j$ denote the average price of active power received by generator $j$ over all the runs of simulation, and $f_m^j$ denotes the marginal cost of generator $j$. The term $\bar{P}^j$ denotes the average active power quantity allocation. Clearly, for a market to be competitive, a lower value of RMPI is desirable. The lowest possible value of RMPI is zero, which is attainable only in a perfectly competitive market.

TABLE VII
EXPERIMENTAL COMBINATIONS WITH RESPONSE VALUES

| Auction | Load | Congestion | $RMPI$ |
|---|---|---|---|
| Uniform | Low | Low | 721.0 |
| Discriminatory | Low | Low | 743.8 |
| Second Price | Low | Low | 444.1 |
| Uniform | High | Low | 935.8 |
| Discriminatory | High | Low | 1279.4 |
| Second Price | High | Low | 615.7 |
| Uniform | Low | High | 790.3 |
| Discriminatory | Low | High | 887.3 |
| Second | Low | High | 490.6 |
| Uniform | High | High | 1110.1 |
| Discriminatory | High | High | 1322.8 |
| Second Price | High | High | 699.1 |

In the experimental design, congestion and load factors were studied at two levels (see Table V), while the auction type was studied at three levels (uniform, discriminatory, and second price uniform). This resulted in a total of 12 ($2 \times 2 \times 3$) experimental combinations for a full factorial study. Due to the long duration of simulation run that was needed to learn the bidding strategies for each of the 12 combinations, only a single replicate of the response was used. An estimate of the error sum of squares needed for analysis of variance (ANOVA) was obtained from a normal probability plot of the effect estimates of all factors and interactions. Table VII shows the experimental data. The sum of squares (SS) of the error was assessed by combining the sum of the squares of a main factor (congestion) and three two-level interactions (auction and load, auction and congestion, and load and congestion), which were found to be insignificant from a normal probability plot of the factor and interaction effects. The result of ANOVA is shown in Table VIII, where $F_0$ denotes the calculated F-test statistic ,and $F_{\text{critical}}$ denotes the test limit. The results indicate that only auction type and load are significant factors. It appears from the result that, for the sample network, the generators need to consider only auction type and network load condition while making bidding decisions in DA energy markets. It is also interesting to note, from the average prices presented earlier in

TABLE VIII
ANOVA RESULTS WITH RMPI AS RESPONSE VARIABLE

| Source of Variation (Label) | $F_0$ | $F_{critical}$ | Significant Factor |
|---|---|---|---|
| Auction (X) | 33.3 | 5.14 | Significant |
| Load (A) | 34.3 | 5.99 | Significant |
| Load(A)-Congestion(B) | 0.5 | 5.14 | Not |
| Auction(X) (ABX) | | | Significant |

Table III, that the generators in a discriminatory auction raise their bids under higher loads, which gives the impression that the auction and load interaction might be a significant factor. This, however, is not supported by the ANOVA results.

## V. CONCLUDING REMARKS

The primary contributions of this paper include the formulation of a stochastic game model for the energy market and its reinforcement learning-based solution methodology. Though the literature is rich with game theoretic treatments for deregulated energy markets, almost all of these papers present the problem within a mathematical programming framework. These models are very well suited to analyze static (one shot) versions of the energy market games with large number of market participants. However, under stochastic demand variations within a modeling horizon, e.g., an hour or a day, the equilibrium obtained from the above models for a given value of the demand may not be optimal for the entire horizon. Hence, the stochastic demand realizations during a modeling horizon should be taken into account, particularly when significant variations are expected. This is accomplished in this paper via the CMDP model, which formulates the POOLCO DA energy market as a nonzero sum average reward stochastic game.

The reinforcement learning-based solution method for the CMDP model presented in this paper is definitely a first cut approach to a very difficult problem. We are currently working on developing theoretical conditions under which the learning-based solution algorithm may provide Nash equilibrium solutions. Scalability of the RL-based solution approach still remains a challenge in solving stochastic games involving actual power networks. This is due to the complexity of the ac version of OPF that our model considers and the large dimensionality of the state-action space. Nonetheless, the RL-based method, which is developed here, allows us to tackle the stochastic game in DA markets in its entirety. Further research is needed before the issues of convergence, optimality, and scalability are fully addressed.

## ACKNOWLEDGMENT

The authors would like to thank the three anonymous referees who review provided comments on the original manuscript of this paper. Addressing of the comments has made the presentation in the revised manuscript much more focused and concise.

## REFERENCES

[1] P. Cramton, "Electricity market design: The good, the bad, and the ugly," in *Proc. Hawaii Int. Conf. System Sciences*, 2003.
[2] N. Fabra, N. H. Von der Fehr, and D. Harbord, Designing electricity auctions: Uniform, discriminatory and vickrey, Tech. Rep., EWPA, 2002.
[3] N. H. von der Fehr and D. Harbord, Competition in electricity spot markets: Economic theory and international experience Oslo Univ., Dept. Econ., Tech. Rep., 1998.
[4] P. D. Klemperer and M. A. Meyer, "Supply function equilibria in oligopoly under uncertainty," *Econometrica*, vol. 57, no. 6, pp. 1243–1277, 1989.
[5] W. Elmaghraby and S. Oren, "Efficiency of multi-unit electricity auctions," *Energy J.*, vol. 20, pp. 89–116, 1999.
[6] J. Contreras, O. Candiles, J. I. de la Feunte, and T. Gomez, "Auction design in day-ahead electricity markets," *IEEE Trans. Power Syst.*, vol. 16, no. 1, pp. 88–96, Feb. 2001.
[7] J. Nicolaisen, V. Petrov, and L. Tesfatsion, "Market power and efficiency in a computational electricity market with discriminatory double-auction pricing," *IEEE Trans. Evol. Comp.*, vol. 5, no. 5, pp. 504–523, Oct. 2001.
[8] Y. S. Son, R. Baldick, K. Lee, and S. Siddiqi, "Short-term electricity market auction game analysis: Uniform and pay-as-bid pricing," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1990–1998, Nov. 2004.
[9] Y. Ren and F. Galiana, "Pay-as-bid versus marginal pricing-part ii: Market behavior under strategic generator offers," *IEEE Trans. Power Syst.*, vol. 19, no. 4, pp. 1777–1783, 2004.
[10] C. A. Berry, B. F. Hobbs, W. A. Meroney, R. P. ONeill, and W. R. Stewart Jr., "Understanding how market power can arise in network competition: A game theoretic approach," *Util. Pol.*, vol. 8, pp. 139–158, 1999.
[11] S. Borenstein and S. Bushnell, "An empirical analysis of the potential for market power in California's electricity industry," *J. Ind. Econ.*, pp. 285–322, 1999.
[12] S. Steven, *Power System Economics*. Piscataway, NJ: IEEE Press, 2002.
[13] T. Mount, "Market power and price volatility in restructured markets for electricity," in *Proc. 32nd Hawaii Int. Conf. System Sciences*, 1999.
[14] S. Spear, "The electricity market game," *J. Econ. Theory*, vol. 109, pp. 300–323, 2003.
[15] S. Borenstein, S. Bushnell, E. Kahn, and S. Stoft, "Market power in California electricity markets," *Util. Pol.*, vol. 5, no. 3, pp. 219–236, 1995.
[16] D. Bunn and F. Oliveira, "Evaluating individual market power in electricity markets via agent-based simulation," *Ann. Oper. Res.*, vol. 121, pp. 57–77, 2003.
[17] S. M. Harvey and W. W. Hogan, Market Power and Market Simulations Harvard Univ., Tech. Rep., 2002, Center for Business and Government John F. Kennedy School of Government.
[18] J. Filar and K. Vrieze, Competitive Markov Decision Processes New York, 1997, Springer Verlag.
[19] B. F. Hobbs, "Linear complementarity models of Nash-Cournot competition in bilateral and poolco power markets," *IEEE Trans. Power Syst.*, vol. 16, no. 2, pp. 194–202, May 2001.
[20] W. Jing-Yuan and Y. Smeers, "Spatial oligopolistic electricity models with Cournot generators and regulated transmission prices," *Oper. Res.*, vol. 47, no. 1, pp. 102–112, 1999.
[21] J. Yao, S. Oren, and I. Adler, "Computing Cournot equilibria in two settlement electricity markets with transmission constraints," in *Proc. 37th Hawaii Int. Conf. System Sciences*, 2003.
[22] J. Hu and M. P. Wellman, "Nash q-learning for general-sum stochastic games," *J. Mach. Learn. Res.*, vol. 4, pp. 1039–1069, 2003.
[23] J. Li, "Learning average reward irreducible stochastic games: Analysis and applications," Ph.D. dissertation, Dept. Ind. Manage. Syst. Eng., Univ. South Florida, Tampa, 2003.
[24] B. F. Hobbs, B. M. Carolyn, and J. S. Pang, "Strategic gaming analysis for electric power systems: An MPEC approach," *IEEE Trans. Power Syst.*, vol. 15, no. 2, pp. 638–645, May 2000.
[25] B. F. Hobbs, "Equilibrium market power modeling for large scale power systems," in *Proc. Power Eng. Soc. Summer Meeting*, 2001, vol. 1, pp. 558–563.
[26] A. Rudkevich, M. Duckworth, and R. Rosen, Modeling electricity pricing in a deregulated generation industry: The potential for oligopoly pricing in a poolco Tellus Institute, Boston, MA, Tech. Rep., 1997, pp. 3411–.
[27] J. Boucher and Y. Smeers, "Alternative models of restructured electricity systems, part 1: No market power," *Oper. Res.*, vol. 49, no. 6, pp. 821–838, 2001.
[28] B. Hobbs and U. Helman, "Complementarity-based equilibrium modeling for electric power markets," in *Modeling Prices in Competitive Electricity Markets*, D. Bunn, Ed. New York: Wiley, 2004, ch. chapter 3.

[29] O. Daxhelet and Y. Smeers, "Variational inequality models of restructured electric systems," in *Applications and Algorithms of Complementarity*, M. C. Ferris, O. L. Mangasarian, and J. S. Pang, Eds. Dordrecht, The Netherlands: Kluwer, 2001.
[30] K. H. Lee and R. Baldick, "Tuning of discretization in bimatrix approach to power system market analysis," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 830–836, May 2003.
[31] K. H. Lee and R. Baldick, "Solving three-player games by the matrix approach with application to an electric power market," *IEEE Trans. Power Syst.*, vol. 18, no. 4, pp. 1573–1580, Nov. 2003.
[32] R. Ragupathi and T. K. Das, "A stochastic game approach for modeling wholesale energy bidding in deregulated power markets," *IEEE Trans. Power Syst.*, vol. 19, no. 2, pp. 849–856, May 2004.
[33] G. R. Gajjar, S. A. Khaparde, P. Nagaraju, and S. A. Soman, "Application of actor-critic learning algorithm for optimal bidding problem of a Genco," *IEEE Trans. Power Syst.*, vol. 18, no. 1, pp. 11–18, Feb. 2003.
[34] T. K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck, "Solving semi-Markov decision problems using average reward reinforcement learning," *Manage. Sci.*, vol. 45, no. 4, pp. 560–574, 1999.
[35] V. Manfredi and S. Mahadevan, "Hierarchical reinforcement learning using graphical models," in *Proc. ICML Workshop Rich Representation Reinforcement Learning*, Bonn, Germany, 2005.
[36] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "reinforcement learning: A survey," *J. Artif. Intell.*, vol. 4, pp. 237–285, 1996.
[37] R. D. Zimmerman, C. E. Murillo-Sanchez, and D. D. Gan, MAT-POWER a MATLAB power system simulation package, user manual, version 3.0.0 Power System Engineering Research Center, Cornell Univ., , Tech. Rep., 2005.
[38] M. L. Puterman*, Markov Decision Processes*. New York: Wiley, 1994.
[39] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.
[40] A. Gosavi*, Simulation-Based Optimization: Parametric Optimization Techniques and Reinforcement Learning*, ser. Operations Research/Computer Science Interfaces. New York: Springer, 2003.
[41] J. Abounadi, D. Bertsekas, and V. S. Borkar, "Learning algorithms for Markov decision processes with average cost," *SIAM J. Control Optim.*, vol. 40, no. 3, pp. 681–698, 2002.
[42] A. Gosavi, N. Bandla, and T. K. Das, "A reinforcement learning approach to airline seat allocation for multiple fare classes with overbooking," *IIE Trans. (Special Issue on Advances on Large-Scale Optimization for Logistics, Production, and Manufacturing Systems)*, vol. 34, no. 9, pp. 729–742, 2002.
[43] L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, 1996.
[44] K. K. Ravulapati, J. Rao, and T. K. Das, "A reinforcement learning approach to stochastic business games," *IIE Trans. Sched. Logist.*, vol. 36, pp. 373–385, 2002.
[45] C. Darken, J. Chang, and J. Moody, "Learning rate schedules for faster stochastic gradient search," in *Neural Networks for Signal Processing 2—Proceedings of the 1992 IEEE Workshop*, D. A. White and D. A. Sofge, Eds., Piscataway, NJ, 1992, IEEE Press.
[46] R. Zimmerman and D. D. Gan, MATPOWER a MATLAB power system simulation package, user manual Power System Engineering Research Center, Cornell Univ., Tech. Rep., 1997.
[47] PowerWorld Simulator Version 8.0 Glover/Sarma Build 11/02/01, Pow-erWorld Corporation. Urbana, IL, 2004 [Online]. Available: http://www.powerworld. com.
[48] A. K. David and F. Wen, "Market power in generation markets," in *Proc. 5th Int. Conf. Power System Control, Operation, Management*, Hong Kong, 2000, pp. 242–248.
[49] A. K. Kian, R. J. Thomas, B. C. Zimmerman, R. Lesieutre, and T. D. Mount, "Identifying the potential for market power in electric power systems in real-time," in *Proc. 37th Hawaii Int. Conf. System Sciences*, 2004.
[50] E. Williams and R. A. Rosen, A better approach to market power analysis Tellus Institute, Tech. Rep., 1999.
[51] N. McCarthy, Market size, market structure, and market power in the Irish electricity industry ESRI Working Paper No. 168.
[52] D. C. Montgomery*, Design and Analysis of Experiments*. New York: Wiley, 2001.

**Vishnuteja Nanduri** (S'06) received the M.S. degree in industrial engineering in 2005 from University of South Florida, Tampa, where he is currently pursuing the Ph.D. degree in the Industrial Engineering Department.

His research is in the field of stochastic game theoretic modeling of deregulated electricity markets and related RL-based solution approaches. He is also currently involved in the research and development of mathematical models for early detection and treatment of cancer.

Mr. Nanduri is a student member of Institute of Industrial Engineers (IIE) and the Institute for Operations Research and Management Sciences (INFORMS).

**Tapas K. Das** (M'06) is a Professor of industrial and management systems engineering at the University of South Florida, Tampa. His current research includes design of reliable deregulated electric power markets; process control and monitoring using wavelet-based multiresolution analysis approaches; and working with K–12 schools. His research on all the above three areas is funded by the National Science Foundation. His research interest also includes developing methodologies for medical diagnosis and cancer drug response analysis.

Dr. Das is a Fellow of Institute of Industrial Engineers (IIE) and member of the Institute for Operations Research and Management Sciences (INFORMS).