# Reinforcement Learning-Based Control of Residential Energy Storage Systems for Electric Bill Minimization

Chenxiao Guan, Yanzhi Wang, Xue Lin, Shahin Nazarian, and Massoud Pedram
Department of Electrical Engineering, University of Southern California, Los Angeles, CA
{chenxiag, yanzhiwa, xuelin, snazaria, pedram}@usc.edu

*Abstract*—**Incorporating residential-level photovoltaic energy generation and energy storage systems have proved useful in utilizing renewable power and reducing electric bills for the residential energy consumer. This is particular true under dynamic energy prices, where consumers can use PV-based generation and controllable storage modules for peak shaving on their power demand profile from the grid. In general, accurate PV power generation and load power consumption predictions and accurate system modeling are required for the storage control algorithm in most previous works. In this work, the reinforcement learning technique is adopted for deriving the optimal control policy for the residential energy storage module, which does not depend on accurate predictions of future PV power generation and/or load power consumption results and only requires partial knowledge of system modeling. In order to achieve higher convergence rate and higher performance in non-Markovian environment, we employ the $TD(\lambda)$-learning algorithm to derive the optimal energy storage system control policy, and carefully define the state and action spaces, and reward function in the $TD(\lambda)$-learning algorithm such that the objective of the reinforcement learning algorithm coincides with our goal of electric bill minimization for the residential consumer. Simulation results over real-world PV power generation and load power consumption profiles demonstrate that the proposed reinforcement learning-based storage control algorithm can achieve up to 59.8% improvement in energy cost reduction.**

## I. Introduction

The traditional electrical power grid is an interconnected transmission network, which moves the electric power from power generators until it reaches users/consumers through long distances. Since the end user profiles often significantly change according to the day of week and time of day, the power grid must be able to support the worst-case demand of power to all end users [1]. On the other hand, the emerging *smart power grid* will exploit digital technology allowing for bidirectional communication between the utilities and the corresponding consumers to meet the expected growth of end user power consumption at the worst case [2], [3]. Integrating a substantial amount of renewable power sources, such as photovoltaic (PV) or wind power sources, into both residential and power grid levels of the smart grid will increase the efficiency of the current electricity infrastructure and reduce environmental impacts [4]. Smart meters and smart distribution system enable two-way information flow, real-time reporting of power grid status and outages and effective interconnection of renewable power sources. These technologies allow monitoring power consumption, automated control of power consumption of smart devices and appliances, dynamic pricing policies of electricity, and fault detection/tolerance in the smart grid, etc.

Although integrating residential-level renewable (PV) power generations into the smart grid proves useful in reducing fossil fuel consumption, several issues need to be addressed for realizing the full benefits. First, there is a mismatch between the peak solar power generation time (typically at noon) and the peak residential-level power consumption time (typically in the evening.) This time skew means that the PV power cannot be optimally utilized for peak shaving. Besides, the daily distribution of the PV output power is almost fixed, depending on the solar irradiance [5], which also restricts the ability of peak shaving for residential consumers.

In order to mitigate the issues mentioned above, one way is to introduce an energy storage module for houses equipped with PV modules [6]. The proposed energy storage module stores power from the smart grid during off-peak hours of each day and (or) from the PV system, and offers power to the residential consumer during the peak period of the day for peak shaving and electricity cost reduction since electricity price is the most expensive during the peak period. Therefore, the design of dynamic pricing-aware energy control algorithm for the residential energy storage system is a critical task for the smart grid to deliver on its promises.

The energy storage capacity of the storage system is limited due to the relatively high cost of the (battery) storage elements. Hence, it is very important for the controller to predict the PV power generation and load power consumption so that it can optimally control the storage system to minimize the electricity cost. For example, reference [6] presented PV power generation and load power consumption predictions specifically designed to help a residential storage controller. However, in many cases the load power consumption predictions result in relatively significant prediction errors, which will affect the performance of the residential storage controller. Besides being robust to prediction inaccuracies, the residential storage controller should also be resilient and robust with respect to inaccuracies and variabilities arisen from modeling, aging, and cell-level variability of PV and storage systems, as well as efficiency variations of power conversion circuitry.

In this paper, we use the reinforcement learning technique for deriving the optimal control policy for the residential energy storage module, which does not depend on accurate predictions of future PV power generation and/or load power consumption results and only requires partial knowledge of system modeling. More specifically, the reinforcement learning-based storage control does not need information of the power conversion efficiencies of various DC/DC converters and DC/AC inverters, but needs information to precisely estimate the remaining energy in the storage module. In order to achieve higher convergence rate and higher performance in non-Markovian environment, we employ the $TD(\lambda)$-learning algorithm to derive the optimal energy storage system control policy, and carefully define the state and action spaces, and reward function in the $TD(\lambda)$-learning algorithm such that the objective of the reinforcement learning algorithm coincides with our goal of electric bill minimization for the residential consumer. Simulations over real PV power generation and load power consumption profiles demonstrate that the proposed reinforcement learning-based storage control algorithm can achieve up to 59.8% improvement in energy cost reduction.

## II. Reinforcement Learning Background

Reinforcement learning provides a mathematical framework for discovering or learning strategies that map situations onto actions with the goal of maximizing a cumulative reward function [8]. The learner and decision-maker is called the agent. The thing it interacts with, comprising everything outside the agent, is called the environment. The agent and environment interact continually, the agent selecting actions and the environment responding to those actions and presenting new states to the agent. The environment also gains rewards, which are special numerical values that the agent tries to maximize over time.

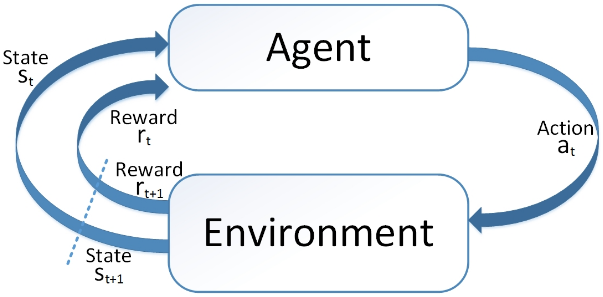

Fig. 1. The agent-environment interaction in reinforcement learning.

Fig. 1 illustrates the agent-environment interaction in reinforcement learning. Specifically, the agent and environment interact at each of a sequence of discrete time steps, i.e., $t = 0, 1, 2, 3, \cdots$. At each time step $t$, the agent receives some representation of the environment state, i.e., $s_t \in S$, where $S$ is the set of possible states, and on that basis selects an action, i.e., $a_t \in A(s_t) \subseteq A$, where $A(s_t)$ is the set of actions available in state $S_t$ and $A$ is the set containing all possible actions. One time step later, in part as a consequence of its action, the agent receives a numerical reward, i.e., $r_{t+1} \in R$, and finds itself in a new state, i.e., $s_{t+1}$.

A policy, denoted by $\pi$, of the agent is a mapping from each state $s \in S$ to an action $a \in A$ that specifies the action $a = \pi(s)$ that the agent will choose when the environment is in state $s$. The ultimate goal of an agent is to find the optimal policy, such that the *value function*

$$V^{\pi}(s) = E\{\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1} | s_t = s\} \quad (1)$$

is maximized for each state $s \in S$. The value function $V^{\pi}(s)$ is the *expected return* when the environment starts in state $s$ at time step $t$ and follows policy $\pi$ thereafter. $0 < \gamma < 1$ is a parameter called the *discount rate* that ensures the infinite sum $\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1}$ converges to a finite value. More importantly, $\gamma$ reflects the uncertainty in the future. $r_{t+k+1}$ is the reward received at time step $t + k + 1$.

## III. System Description

### A. System Architectures

In this paper, we consider a residential consumer equipped with PV power generation and energy storage modules as shown in Fig. 2. The primary purpose of energy storage modules in this system is to serve critical loads during a utility outage and offer power to residential consumer during the peak period of the day for peak shaving and electricity cost reduction. The PV and storage modules are connected to a residential DC bus via DC-DC converters. The smart grid and the residential AC load are connected to the AC bus which is further connected to the residential DC bus via AC/DC inverter and rectifier.

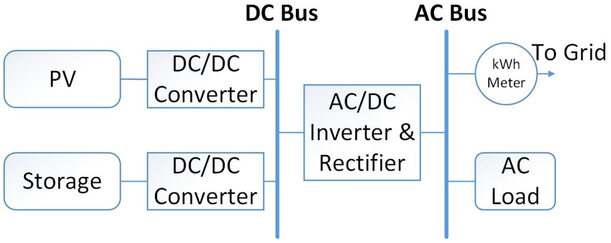

Fig. 2. Block diagram showing the interface between PV module, storage system, residential load, and the smart grid.

We adopt enhanced communication and control support as shown in Fig. 3. The utility smart grid can provide interactive anti-islanding control and dynamic energy pricing information. The system control function shown in the figure represents the core part of the system. The functions including: inverter and converter control, energy storage management, could all be incorporated into a single device, or be separate devices communicating with each other. The energy storage controller can supply energy to the energy storage modules when utility-supplied electricity is at lower cost, and can convert the output of energy storage module to AC load during the peak period of the day. Both of them could be operated in coordination with the PV energy generation.

We adopt a slotted time model, i.e., all system constraints and decisions are provided for discrete time intervals of equal and fixed length. We divide each day into $T$ time slots, each with duration $D$. Hence, we use $T = 96$ and $D = 15$ minutes.

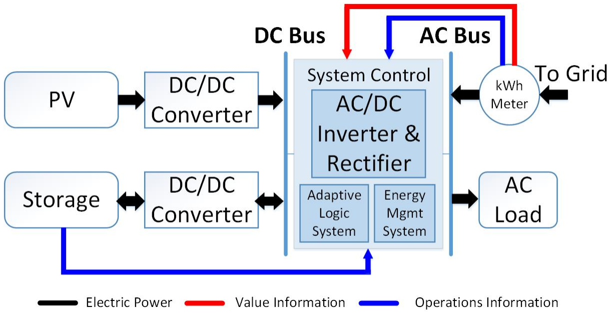

Fig. 3. Residential system with enhanced communication and control structures.

Consider the day-ahead energy pricing scenario with billing period of one day. At the beginning of the day, the smart grid will announce the price signal $Price[t]$ at each time slot $t$. The residential AC load at time slot $t$ of that day is denoted by $P_{load}[t]$. The output power levels of PV and energy storage modules at time slot $t$ are denoted by $P_{pv}[t]$ and $P_{st}[t]$, respectively, where $P_{st}[t]$ can be positive, negative, or zero which means discharging the storage, charging the storage, or idle respectively. The power required from the smart grid, i.e., the grid power consumption, at time slot $t$ is denoted by $P_{grid}[t]$. $P_{grid}[t]$ is positive when electric power is drawn from the smart grid to the residential system. Meanwhile, we should consider the efficiency of each converter. The efficiencies of the DC/DC converter between PV and DC bus, the DC/DC converter energy storage module and DC bus, and the DC/AC inverter and rectifier between DC bus and AC bus are denoted by $\eta_{pv}[t]$, $\eta_{st}[t]$, and $\eta_{AC/DC}[t]$, respectively, at time slot $t$.

The power flowing from the DC bus to the DC/AC inverter (i.e., on the left side of the DC/AC inverter), denoted by $P_{DC/AC,left}[t]$, is given by

$$\begin{cases} P_{pv}[t]\cdot\eta_{pv}[t]+P_{st}[t]\cdot\eta_{st}[t], & if P_{st}[t]\geq 0\\ P_{pv}[t]\cdot\eta_{pv}[t]+P_{st}[t]\cdot\frac{1}{\eta_{st}[t]}, & if P_{st}[t]<0\end{cases} \tag{2}$$

On the other hand, the power flowing from the DC/AC inverter (i.e., on the right side of the DC/AC inverter), denoted by $P_{DC/AC,right}[t]$, is given by

$$P_{DC/AC,right}[t]=P_{load}[t]-P_{grid}[t] \tag{3}$$

Please note that both $P_{DC/AC,left}[t]$ and $P_{DC/AC,right}[t]$ can be positive or negative, and they satisfy the following relationship:

$$\begin{aligned} &\eta_{AC/DC}[t]\cdot P_{DC/AC,left}[t]\\ &\qquad =P_{DC/AC,right}[t],\ \text{if } P_{DC/AC,left}[t]>0\\ &\frac{1}{\eta_{AC/DC}[t]}\cdot P_{DC/AC,left}[t]\\ &\qquad =P_{DC/AC,right}[t],\ \text{if } P_{DC/AC,left}[t]<0 \end{aligned} \tag{4}$$

The total energy cost we pay in a billing period (one day) is given by:

$$Cost_E=\sum_t Price[t]\cdot\max\{0,P_{grid}[t]\}\cdot D, \tag{5}$$

which implies that selling power back to the grid, i.e., $P_{grid}[t]<0$, will not get reimbursed. This is due to realistic considerations.

The residential energy storage controller has knowledge of the current power consumption values $P_{load}[t]$, $P_{pv}[t]$, and $P_{grid}[t]$ from real-time measurement, and it controls the discharging/charging power $P_{st}[t]$ of the storage module. On the other hand, the power conversion efficiencies $\eta_{pv}[t]$, $\eta_{st}[t]$, and $\eta_{AC/DC}[t]$ are functions of the input/output power of the corresponding converter/inverter, and the energy storage controller has no prior knowledge of these efficiency values for realistic concerns.

### *B. Energy Storage Module Modeling*

In the residential energy system, the energy storage controller requires a precise estimation of the current energy stored in the storage module, and thus, an accurate modeling of the energy storage module is required.

The most significant cause of power losses in the storage system, which is typically made of lead-acid batteries or Li-ion batteries, is the rate capacity effect of batteries [10], [11]. To be more specific, high discharging current of the battery will reduce the amount of available energy that can be extracted from the battery, thereby reducing the battery's service life between fully charged and fully discharged states [10]. In other words, high-peak pulsed discharging current will deplete much more of the battery's stored energy than a smooth workload with the same total energy demand. We use discharging efficiency of a battery to denote the ratio of the battery's output current to the degradation rate of its stored charge. Then the rate capacity effect specifies the fact that the discharging efficiency of a battery decreases with the increase of the battery's discharging current. The rate capacity effect also affects the energy loss in the battery during the charging process in a similar way.

The rate capacity effect can be captured using the Peukert's formula, an empirical formula specifying the battery charging and discharging efficiencies as functions of the charging current $I_c$ and discharging current $I_d$, respectively:

$$\eta_{rate,c}(I_c)=\frac{1}{(I_c/I_{ref})^{\alpha_c}},\eta_{rate,d}(I_d)=\frac{1}{(I_d/I_{ref})^{\alpha_d}}, \tag{6}$$

where $\alpha_c$ and $\alpha_d$ are peukert's coefficients, and their values are typically in the range of $0.1-0.3$; $I_{ref}$ denotes the reference current of the battery, which is proportional to the battery's nominal capacity $C_{nom}$. Typically, $I_{ref}$ is set to $C_{nom}/20$, indicating that it takes 20 hours to fully discharge the battery using discharging current $I_{ref}$.

We name $I_c/I_{ref}$ and $I_d/I_{ref}$ the battery's normalized charging current and normalized discharging current, respectively. Notice that the efficiency values $\eta_{rate,c}(I_c)$ and $\eta_{rate,d}(I_d)$ in Eqn. (6) are greater than 100% if the magnitude of the normalized charging or discharging current is less than one, which implies that the above-mentioned Peukert's formula is not accurate in this case. We modify the Peukert's formula such that the efficiency values $\eta_{rate,c}(I_c)$ and $\eta_{rate,d}(I_d)$

become equal to 100% if the magnitude of the normalized charging/discharging current is less than one. In other words, the battery suffers from no rate capacity effect in this case.

We denote the increase/degradation rate of storage energy in the $t$-th time slot by $P_{st,in}[t]$, which may be positive (i.e., discharging from the storage, and the amount of stored energy decreases), negative (i.e., charging the storage, and the amount of stored energy increases), or zero. Based on the modified Peukert's formula, the relationship between $P_{st,in}[t]$ and the storage output power $P_{st}[t]$ is given by:

$$\begin{cases} V_{st} \cdot I_{st,ref} \cdot (\frac{P_{st,in}[t]}{V_{st} \cdot I_{st,ref}})^{\beta_1} & , if \frac{P_{st,in}[t]}{V_{st} \cdot I_{st,ref}} > 1 \\ P_{st,in}[t] & , if -1 \le \frac{P_{st,in}[t]}{V_{st} \cdot I_{st,ref}} \le 1 \\ -V_{st} \cdot I_{st,ref} \cdot (\frac{|P_{st,in}[t]|}{V_{st} \cdot I_{st,ref}})^{\beta_1} & , if \frac{P_{st,in}[t]}{V_{st} \cdot I_{st,ref}} < -1 \end{cases} \tag{7}$$

where $V_{st}$ is the storage terminal voltage and is supposed to be (near-) constant; $I_{st,ref}$ is the reference current of the storage system, which is proportional to its nominal capacity $C_{st,nom}$ given in Ampere-Hour (Ahr); coefficient $\beta_1$ is in the range of $0.8 - 0.9$, whereas coefficient $\beta_2$ is in the range of $1.1 - 1.3$.

The residential energy storage controller estimates the remaining energy in the storage module using the Coulomb counting method [12], i.e., the remaining energy $E_{st}[t]$ at the begin of time slot $t$ (i.e., at the end of time slot $t-1$) is estimated via:

$$E_{st}[t] = E_{st,ini} - \sum_{t'=1}^{t-1} P_{st,in}[t'] \cdot D \tag{8}$$

where $E_{st,ini}$ is the initial energy stored in the storage module at the beginning of day.

## IV. Reinforcement learning based energy storage system control

### A. Motivations

Reinforcement learning provides a efficient solution to the problems in which (i) with the change of system states, different actions should be taken, and both the current states and the selected action determine the future state; (ii) an expected return will be optimized cumulatively instead of immediately; (iii) the agent only needs knowledge of the current state and the reward it receives, which means it is a Markov process; (iv) the system might be non-stationary to some degree. These properties make reinforcement learning different from other machine learning techniques, model-based optimal control and dynamic programming, and Markov decision process-based approach respectively.

On the other hand, the energy storage system control possesses all of the four properties mentioned above. (i) During a whole day, the PV power generation, energy storage level, load power consumption and electricity price require different operation modes and actions, and also the future energy storage level depends on the charging/discharging current. (ii) The energy storage system aims at minimizing the total electricity cost during a whole day rather than electricity cost rate (price) at a certain time step. (iii) The energy storage system control agent does not have *a priori* knowledge of a whole day, while it has only the knowledge of the current PV power generation, load power consumption and energy storage level as a result of the action taken. (iv) The actual consumer load consumption profiles are non-stationary. Hence, the reinforcement learning technique better suits the energy storage system than other optimization methods.

### *B. State, Action and Reward of the Reinforcement Learning Algorithm*

*1) State Space:* We define the state space of the energy storage system control problem as a finite number of states, each represented by the residential load consumption, PV power generation, energy storage level, and energy price:

$$\begin{aligned} S =& \{s = [P_{load}, P_{pv}, E_{st}, Price]^T | P_{load} \in \mathbf{P_{load}}, P_{pv} \in \mathbf{P_{pv}}, \\ & E_{st} \in \mathbf{E_{st}}, Price \in \mathbf{Price}\}, \end{aligned} \tag{9}$$

where $\mathbf{P_{load}}$, $\mathbf{P_{pv}}$, $\mathbf{E_{st}}$, and $\mathbf{Price}$ are respectively the finite sets of residential load power consumption levels, residential PV power generation levels, energy storage levels in the storage module, and electricity prices. Discretization is required when defining these finite sets.

At each time slot $t$, the storage controller has knowledge of $Price[t]$ which is pre-announced by the smart grid controller at the beginning of day; it measures the $P_{pv}[t]$ and $P_{load}[t]$ levels; and estimates the energy storage level $E_{st}[t]$ using Eqn. (8). In this way the storage controller knows the current state at time slot $t$. In other words, the reinforcement learning framework is *partially model-free* in that it does not need information of the power conversion efficiencies of various DC/DC converters and DC/AC inverters, but needs information to precisely estimate the remaining energy in the storage module.

The current state space is comprised of four dimensions, i.e., $P_{load}$, $P_{pv}$, $E_{st}$, and $Price$, which makes the total number of states high. The complexity and convergence speed of reinforcement learning algorithms are proportional to the number of state-action pairs [8]. In order to reduce computation complexity and accelerate convergence, we use $P_{load} - P_{pv}$ to replace both $P_{load}$ and $P_{pv}$ in the state vector, and thus the state space becomes:

$$\begin{aligned} S =& \{s = [P_{load} - P_{pv}, E_{st}, Price]^T | P_{load} - P_{pv} \in \mathbf{P_{netload}}, \\ & E_{st} \in \mathbf{E_{st}}, Price \in \mathbf{Price}\} \end{aligned} \tag{10}$$

Please note that $P_{load} - P_{pv}$ could be either positive or negative. The proposed replacing $P_{load}$ and $P_{pv}$ by $P_{load} - P_{pv}$ in the state vector is intuitive because (i) when $P_{load} - P_{pv}$ is less than zero, the excessive power consumption can be used to charge the storage, and (ii) when $P_{load} - P_{pv}$ is large, it is more desirable to discharge storage in order to provide power for the residential loads.

*2) Action Space:* We define the action space of the energy storage system control problem as a finite number of actions, in which each action represents a specific discharging/charging power of the energy storage module:

$$A = \{a = P_{st} | P_{st} \in \mathbf{P_{st}}\}, \tag{11}$$

The set $\mathbf{P_{st}}$ contains within it a finite number of current values in the rage of $[-P_{st,max}, P_{st,max}]$. $P_{st} > 0$ denotes discharging the energy storage module; $P_{st} < 0$ denotes charging the energy storage module. Discretization is required when defining this finite set $\mathbf{P_{st}}$.

*3) Reward Function:* We define the reward that the reinforcement learning agent receives after taking action $a$ at state $s$ as the negative value of the electricity cost in that time step, i.e., $-Price[t] \cdot \max\{0, P_{grid}[t]\} \cdot D$. In this reward function, $P_{grid}[t]$ is pre-announced by the smart grid controller at the beginning of day and $Price[t]$ can be measured by the residential controller. Remember from Section II that the reinforcement learning-based storage controller aims at maximizing the expected return, i.e., the discounted sum of rewards. Therefore, by using the negative value of the electricity cost in a time step as the reward, the total electricity cost will be minimized while maximizing the expected return.

### *C. $TD(\lambda)$-Learning Algorithm for Energy Storage System Control*

To derive the optimal energy storage control policy, we adopt a specific type of reinforcement learning technique, namely the $TD(\lambda)$-learning algorithm [9], due to its higher convergence rate and higher performance in the non-Markovian environment (compared with the simplest Q-learning method.) In $TD(\lambda)$-learning, a value function $Q(s,a)$ is associated with each state-action pair $(s,a)$, which approximates the expected (discounted) cumulative reward when taking action $a$ at state $s$. There are two basic steps in the $TD(\lambda)$-learning algorithm: action selection and $Q$-value update.

*1) Action Selection:* The most straightforward approach for action selection is to always select the current best action with the highest $Q$ value. However, this approach is at the risk of getting stuck at a sub-optimal solution. A judicious reinforcement learning agent should thus exploit the best action known so far to gain high rewards and meanwhile explore the other candidate actions to find a potentially better choice. We address this *exploration versus exploitation* issue by separating the overall learning procedure into two phases: The first phase is the exploration phase, in which the $\epsilon$-greedy policy is adopted, i.e., the current best action is chosen with probability of $1 - \epsilon$, whereas all the other actions are chosen with the same probability. The second phase is the exploitation phase, in which the action with the highest $Q$ value is always selected for reward maximization.

*2) Updating Q-Values Using Eligibility Traces:* Suppose that action $a_t$, i.e., $P_{st}[t]$, is taken in state $s_t$, i.e., $[P_{load}[t], P_{pv}[t], E_{st}[t], Price[t]]^T$, at time step (time slot) $t$, and reward $r_{t+1}$ and new state $s_{t+1}$ are observed at this time step. Then at the next time step $t+1$, the TD($\lambda$)-learning algorithm updates $Q$ value for each state-action pair $(s,a)$ as:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \cdot e(s,a) \cdot \delta, \tag{12}$$

where $\alpha$ is the *learning rate* coefficient, $e(s,a)$ is the *eligibility* of the state-action pair $(s,a)$ specifying the frequency that state-action pair $(s,a)$ is encountered in the past, and $\delta$ is calculated as

$$\delta \leftarrow r_{t+1} + \gamma max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t), \tag{13}$$

where $\gamma$ denotes the *discount rate* in reinforcement learning algorithm.

At time step $t+1$, the eligibility $e(s,a)$ of each state-action pair $(s,a)$ is updated by

$$e(s,a) \leftarrow \begin{cases} \gamma \cdot \lambda \cdot e(s,a) + 1, & s = s_t \cap a = a_t \\ \gamma \cdot \lambda \cdot e(s,a), & otherwise \end{cases} \tag{14}$$

to reflect the degree to which state-action pair $(s,a)$ has been selected in recent past, in which $\lambda$ is a constant value between 0 and 1.

*3) Algorithm Description:* The pseudo code of the TD($\lambda$)-learning algorithm for energy storage module control is described in Algorithm 1.

**Algorithm 1** TD($\lambda$)-Learning Algorithm for Residential Energy Storage Controller:

```
Initialize $Q(s,a)$ arbitrarily for all the state-action pairs.
for For each time step $t$ do
    Choose action $a_t$ for state $s_t$ using the exploration-exploitation policy discussed in Section IV part C.
    Take action $a_t$, observe reward $r_{t+1}$ and the next state $S_{t+1}$.
    $\delta \leftarrow r_{t+1} + \gamma max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)$.
    $e(s_t, a_t) \leftarrow e(s_t, a_t) + 1$.
    for For all state-action pair $(s,a)$ do
        $Q(s,a) \leftarrow Q(s,a) + \alpha \cdot e(s,a) \cdot \delta$.
        $e(s,a) \leftarrow \gamma \cdot \lambda \cdot e(s,a)$.
    end for
end for
```

## V. Experimental Results

In this section we provide experimental results on the effectiveness of the proposed reinforcement-learning based residential storage module control algorithm. The PV power profiles used in our experiments are measured at Duffield, VA, in the year 2007, whereas the residential load consumption data comes from the Baltimore Gas and Electric Company, also measured in the year 2007 [13]. We add some random peaks to the load consumption profiles. Fig. 4 illustrates the synthesized day-ahead electricity price function during a day, which has peak-hours in the evening and off-peak hours mainly in the morning.

We compare the performance in electric bill reduction of the proposed reinforcement learning-based storage control algorithm with baseline algorithm. The baseline algorithm charges the storage module during the off-peak hours (00:00 to 03:59) with constant charging power and discharges the storage module during the peak hours (20:00 to 21:59) with

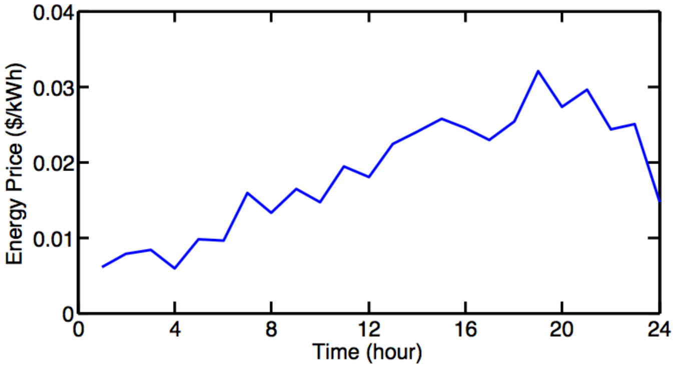

Fig. 4. Synthesized day-ahead electricity price function during a day.

constant discharging power in order to perform peak shaving. Fig. 5 illustrates the total energy cost of each billing period (each day) over a whole year. One can clearly observe that the proposed algorithm consistently outperforms the baseline algorithm in terms of reducing electric bill.

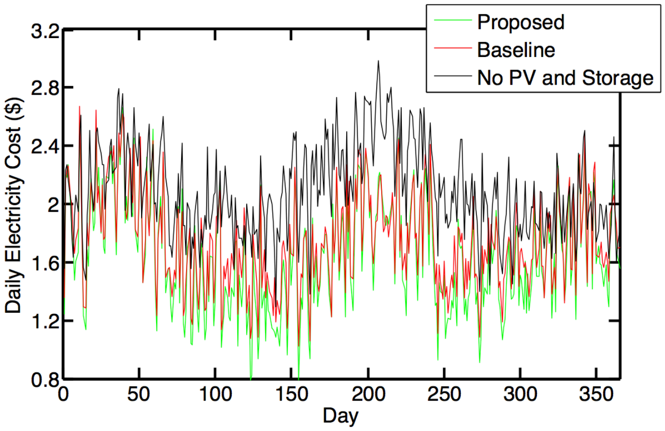

Fig. 5. Comparison between the original energy cost of the residential consumer, energy cost equipped with the proposed algorithm, and energy cost equipped with the baseline algorithm on each day over a year.

We define the *cost saving capability* of a storage module control algorithm (the proposed algorithm or baseline algorithm) to be the electric cost saving over a billing period (one day) due to the additional storage module, compared with the same residential smart grid consumer without PV and energy storage systems. We compare the cost saving capabilities of the proposed algorithm versus the baseline algorithm on every month (summed over all the days over the month) throughout a year, as illustrated in Table I. Experimental results demonstrate that the proposed storage control algorithm achieves a maximum improvement of 59.8% on the monthly cost saving capability, which occurs in October.

## VI. Conclusion

In this paper, we use the reinforcement learning technique for deriving the optimal control policy for the residential energy storage module, which does not depend on accurate predictions of future PV power generation and/or load power consumption results and only requires partial knowledge of system modeling. More specifically, the reinforcement learning-based storage control does not need information of the power conversion efficiencies of various DC/DC converters and DC/AC inverters, but needs information to precisely estimate the remaining energy in the storage module. We employ the $TD(\lambda)$-learning algorithm to derive the optimal energy storage system control policy in order to achieve higher convergence rate and higher performance in non-Markovian environment. We carefully define the state and action spaces, and reward function in the $TD(\lambda)$-learning algorithm such that the objective of the reinforcement learning algorithm coincides with our goal of electric bill minimization for the residential consumer.

TABLE I
Monthly saving of electricity cost of the proposed algorithm and baseline algorithm ($).

| | Jan | Feb | Mar | Apr | May | Jun |
|---|---|---|---|---|---|---|
| Proposed | 6.1668 | 7.3146 | 13.9817 | 14.3264 | 19.6353 | 19.1056 |
| Baseline | 4.3619 | 6.2966 | 11.7838 | 9.8755 | 14.1859 | 17.0373 |
| | Jul | Aug | Sep | Oct | Nov | Dec |
| Proposed | 17.5026 | 17.4447 | 18.9397 | 12.6236 | 8.2157 | 5.3630 |
| Baseline | 15.7595 | 15.8013 | 14.6985 | 7.9001 | 6.1563 | 3.1549 |

## Acknowledgement

This work is supported in part by a grant from the Directorate for Computer & Information Science & Engineering of the National Science Foundation.

## References

[1] L. D. Kannberg, D. P. Chassin, J. G. DeSteese, S. G. Hauser, M. C. Kintner-Meyer, R. G. Pratt, L. A. Schienbein and W. M. Warwick, "GridWise$^{TM}$: The benefits of a transformed energy system," PNNL14396, Pacific Northwest National Laboratory. Sep. 2004.
[2] S. Massoud Amin and B. F. Wollenberg. "Toward a smart grid: Power delivery for the 21st century," *IEEE Power and Energy Magazine*, 3(5), pp. 34-41. 2005.
[3] S. Keshav and C. Rosenberg. "How internet concepts and technologies can help green and smarten the electrical grid," in *ACM SIGCOMM Computer Communications Review*, 2011.
[4] S. Caron and G. Kesidis, "Incentive-based energy consumption scheduling algorithms for the smart grid," in *Proc. Smart Grid Commun. Conf.*, 2010.
[5] Y. Kim, N. Chang, Y. Wang and M. Pedram. "Maximum power transfer tracking for a photovoltaic-supercapacitor energy system," in *Proc. of International Symposium on Low Power Electronics and Design (ISLPED)*, 2010.
[6] Y. Wang, S. Yue, L. Kerofsky, S. Deshpande and M. Pedram, "A hierarchical control algorithm for managing electrical energy storage systems in homes equipped with PV power generation," *Proc. of IEEE Green Technology Conference (GTC)*, Apr. 2012.
[7] Y. Wang, X. Lin and M. Pedram, "Adaptive control for energy storage systems in households with photovoltaic modules," *IEEE Transactions on Smart Grid*, 5(2), pp. 992-1001. 2014.
[8] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*, 1998.
[9] R. S. Sutton, Learning to predict by the methods of temporal differences,?*Machine Learning*, vol. 3, pp. 9-44, 1988.
[10] D. Linden and T. B. Reddy, *Handbook of Batteries*. McGraw-Hill Professional, 2001.
[11] D. Doerfeel and S. A. Sharkh, "A critical review of using the Peukert's equation for determining the remaining capacity of lead-acid and lithium-ion batteries," *Journal of Power Sources*, 2006.
[12] K. S. Ng, C. S. Moo, Y. P. Chen, and Y. C. Hsieh, "Enhanced coulomb counting method for estimating state-of-charge and state-of-health of lithium-ion batteries," *Applied Energy*, 2009.
[13] Baltimore Gas and Electric Company, *Historical Load Data*, https://supplier.bge.com/LoadProfiles_EnergySettlement/historicalloaddata.htm.