Model-Free Learning-Based Online Management of Hybrid Electrical Energy Storage Systems in Electric Vehicles

Siyu Yue, Yanzhi Wang, Qing Xie, Di Zhu, and Massoud Pedram
Dept. of Computer Engineering
University of Southern California
Los Angeles, CA, United States, 90089
{siyuyue, yanzhiwa, xqing, dizhu, pedram}@usc.edu

Naehyuck Chang
Dept. of EECS/CSE
Seoul National University
Seoul, Korea, 151-747
naehyuck@elpl.snu.ac.kr

***Abstract*—To improve the cycle efficiency and peak output power density of energy storage systems in electric vehicles (EVs), supercapacitors have been proposed as auxiliary energy storage elements to complement the mainstream Lithium-ion (Li-ion) batteries. The performance of such a hybrid electrical energy storage (HEES) system is highly dependent on the implemented management policy. This paper presents a model-free reinforcement learning-based approach to dynamically manage the current flows from and into the battery and supercapacitor banks under various scenarios (combinations of EV specs and driving patterns). Experimental results demonstrate that the proposed approach achieves up to 25% higher efficiency compared to a Li-ion battery only storage system and outperforms other online HEES system control policies in all test cases.**

Keywords—Electric Vehicle; Hybrid Energy Storage Systems; Reinforcement Learning

I. INTRODUCTION

There has been a steady increase in the market share of full electric vehicles (FEVs) and hybrid electric vehicles (HEVs) over the past few years [1]. Compared to conventional internal combustion engine (ICE) vehicles, FEV and HEV have better energy efficiency [2]. These improvements mainly result from the fact that the efficiency of an electric motor (typically 80-90%) is significantly higher than that of an ICE (typically around 20%) [3]. In addition, FEVs and HEVs can easily recover a portion of kinetic energy back to electric energy when the vehicle is decelerating, i.e., by using the *regenerative braking* technique. They also have the ability to completely shut down the electric motor when the vehicle is in the stand-by mode [2].

In most recent mass production FEVs and HEVs, Li-ion battery is preferred over other types of batteries. The main advantage of Li-ion battery is its high cell voltage, high energy density, and long lifetime. Li-ion battery also has less self-discharge and no memory effect [5]. However, the power density (defined as the maximum amount of power provided per unit volume or mass) of Li-ion battery is low. In addition, a high charging or discharging current significantly degrades Li-ion battery's cycle efficiency due to its *rate capacity effect* [18], and severely deteriorates its cycle life [6]. Unfortunately, high charging and discharging current is rather common in the application of FEV and HEV. The peak power demand during acceleration is usually several times larger than that of constant speed driving [7]. Regenerative braking also produces high charging currents [9].

To combat the disadvantages and improve the cycle efficiency of Li-ion batteries, researchers have proposed to combine supercapacitors (also known as ultracapacitors) with Li-ion batteries to form a hybrid electrical energy storage (HEES) system [4]. Contrary to Li-ion battery, supercapacitor has superior power density and negligible rate capacity effect, despite its low energy density and high capital cost. Therefore, supercapacitors can be used as an auxiliary energy storage buffer to take care of peak discharging and charging power, while Li-ion batteries provide the average power demand.

Researchers have proposed a few control policies for the hybrid energy storage system in FEV and HEV, which can be categorized into offline optimization-based policies and online decision-based policies. Offline management policies assume given driving profiles, and their optimality relies on the accuracy of their models of motors, vehicle dynamics, conversion circuitry, and electrical storage systems. For example, Wang *et al.* investigate both the design and control problem of a hybrid storage system, with much simplification on the storage element model and the driving profile [14]. Moreno *et al.* propose an optimal control policy based on the complete knowledge of the storage system's model, DC-DC converter's model and the driving profile [11]. Sangyoung *et al.* employ charge migration from the battery to the supercapacitor to improve the overall efficiency of the storage system, considering the rate capacity effect of Li-ion battery during both charging and discharging [9].

Most online management policies are heuristic-based approaches. Shah *et al.* propose a simple control algorithm to manage the HEES system, where supercapacitor is used only during sudden vehicle acceleration or deceleration [13]. Cao *et al.* elaborate the possible configurations of hybrid storage systems and propose a hysteretic feedback control method to maintain supercapacitor terminal voltage [12]. Ortúzar *et al.* use a threshold-based approach that limits the charging and discharging current of the battery bank [15]. Miller *et al.* applied feedback control on the supercapacitor voltage so that the fast transient power demand is handled by the supercapacitor [16]. Thounthong *et al.* consider a similar HEES system where the fuel cell is equipped as the main storage unit instead of Li-ion batteries [10]. Other than heuristic-based management policies, Moreno and Ortúzar *et al.* also propose to use neural network based approach to control the hybrid storage system [11][15].

This work is supported in part by the Software and Hardware Foundations program of the NSF's Directorate for Computer & Information Science & Engineering.

In this work, we focus on the management of HEES system for FEVs. To the best of our knowledge, this is the first paper that presents a model-free online reinforcement learning (RL) based approach to manage HEES systems in FEVs. Unlike offline optimization-based policies, our approach is model-free, i.e., it does not need to know the model of vehicle dynamics, motors, conversion circuitry, etc. It treats these models as a black box and solely relies on the input and output to optimize its decision. Furthermore, it adjusts the management policy to adapt to different driving profiles. Our approach is also different from the neural network based one in that neural network is supervised learning technique based on golden results used for training, whereas our approach learns on the fly. We perform simulations for various FEV specifications and different driving profiles.

II. System Overview

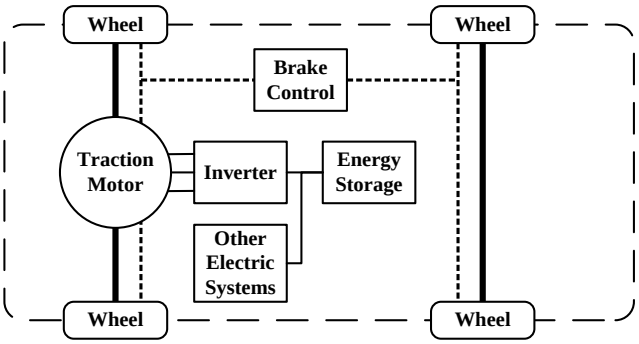

Fig. 1. Powertrain of a 2WD FEV.

Fig. 1 shows the powertrain structure of a two-wheel drive (2WD) FEV. The system consists of an *electric traction motor*, an *energy storage system*, a *brake control system*, power conversion circuitry including an *IGBT-based inverter* and *switching-mode power converters* (shown in Fig. 3), and other electric systems such as air-conditioning, lighting and in-vehicle entertainment systems.

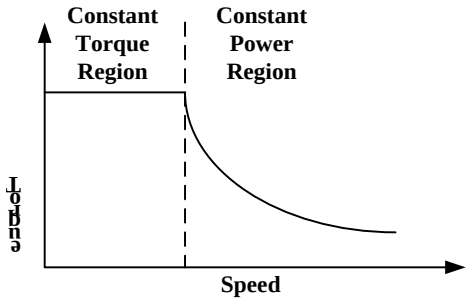

Fig. 2. Torque-speed relationship of an electric motor.

During normal driving the energy storage system provides power to the traction motor, and the motor applies torque to wheels through the axle shafts. Unlike ICEs, the output torque of electric motors are high throughout the full speed range, therefore the use of multiple transmission gears is not commonly required. Fig. 2 shows a typical max-torque versus speed curve of an electric motor. As shown in the figure, the operation of an electric motor is usually divided into two regions: a constant-torque region and a constant-power region. In the constant-torque region, the maximum torque applied is limited by the tire grip or the consideration of passenger comfort, and the output power of the electric motor is proportional to the speed of the vehicle. In contrast, the maximum torque is limited by the maximum power of the motor in the constant-power region.

During braking, the electric motor works as a generator to recycle kinetic energy and store it back to the energy storage system (regenerative braking) or use it to heat up vehicle interior through a thermal resistor (dynamic braking). The decelerating torque generated by the motor alone is usually too small to slow down the vehicle as quickly enough. Thus, a conventional hydraulic braking system is still required.

A. Hybrid Energy Storage System

Fig. 3 shows the HEES system structure. It employs parallel structure as it is flexible (the terminal voltages of both energy storage banks varies and may not match the voltage on the DC bus) [17]. The system consists of a battery bank as the main storage unit and a supercapacitor as an auxiliary storage unit. Three main factors contribute to the power loss in such a HEES system, including the internal resistance of storage banks, the rate capacity effect of the Li-ion battery bank, and the power dissipation in the two switching-mode power converters.

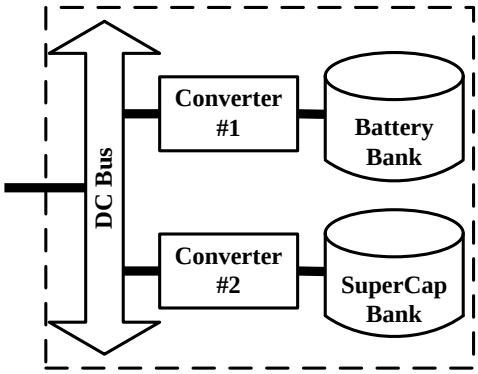

Fig. 3. Hybrid electrical energy storage system.

The power loss due to internal resistance is given by $I^2 r_{int}$. The internal resistance r_{int} of a Li-ion battery is a function of the *state-of-charge* (SoC) of the battery, and is generally different from charging and discharging. The SoC of a battery is defined as the ratio between the stored charge and the full charge capacity.

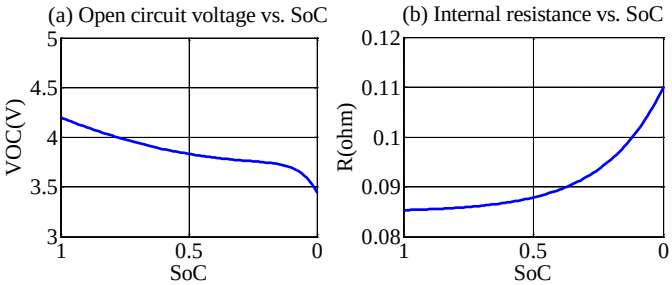

Fig. 4. Li-ion battery cell characteristics.

Fig. 4 illustrates the open circuit voltage and internal resistance as a function of the SoC for a typical Li-ion battery cell during discharging. There exists several methods to estimate the SoC of a Li-ion battery accurately, such as Coulomb counting [19], extended Kalman filtering [20], etc.

Rate capacity effect describes the fact that the actual charge change rate I_a inside a battery is a sublinear (superlinear) function of its charging (discharging) current I_c (I_d). The Peukert's law is a widely adopted empirical equation to evaluate the rate capacity effect [18]:

$$I_a \triangleq Q_f \frac{\mathrm{d}SoC_b(t)}{\mathrm{d}t} = I_{ref}\left(\frac{I_c}{I_{ref}}\right)^{k_c} \text{ or } I_{ref}\left(\frac{I_d}{I_{ref}}\right)^{k_d} \quad (1)$$

where k_c and k_d are the Peukert's constants ($0 < k_c < 1 < k_d < 2$) for charging and discharging, respectively. I_{ref} is the reference current and is usually calculated by $Q_f/20$ where Q_f (in Ah) is the full charge capacity of a battery. Supercapacitors do not suffer from rate capacity effect in contrast.

Two switching-mode power converters convert the terminal voltage of both storage banks to match the DC bus voltage setting and regulate input/output currents of both banks by feedback control. The conversion efficiency of a converter η_{conv} is defined as:

$$\eta_{conv} = \frac{P_{out}}{P_{in}} = \frac{V_{out}I_{out}}{V_{in}I_{in}} \quad (2)$$

where V_{out} and I_{out} are the output voltage and current, V_{in} and I_{in} are the input voltage and current of the converter, respectively. The efficiency of a converter is usually not constant, but dependent on its input and output voltages and currents [9].

B. Vehicle Dynamics

A common model used to calculate the total traction force F_T of a vehicle is based on the following equation [8]:

$$F_T = F_a + F_g + F_R + F_{AD}$$
$$= m \cdot a + m \cdot g \cdot \sin\theta + C_R \cdot m \cdot g \cdot \cos\theta + \frac{1}{2} C_A \cdot \rho \cdot A \cdot v^2 \quad (3)$$

where F_a is the acceleration force, F_g is the component of vehicle gravity force along the road with slope θ, F_R is the rolling friction force, and F_{AD} is the aerodynamic resistance force. Note that C_R is the rolling resistance coefficient which is dependent on the condition of both the road and the tire, C_A is the aerodynamic resistance coefficient, ρ is the air density, A is the frontal area of the vehicle, and v is the velocity of the vehicle.

Given the total traction force, the driving power demand is:

$$P_d = F_T \cdot v/\eta_m \quad (4)$$

where η_m is the efficiency of the electric motor (including the inverter) and is a function of the angular speed of the motor $\omega_m = v \cdot \mu/r$ and the torque of the motor $T_m = F_T \cdot r/\mu$ where r is the radius of the tire and μ is the axle ratio. Note that in regenerative braking mode ($T_m < 0$), the hydraulic braking system may provide braking torque in addition to that provided by the electric motor.

We define P_{total} as the total power drawn from the energy storage system:

$$P_{total} = P_d + P_e \quad (5)$$

where P_e accounts for the power demand of all other electrical systems in the vehicle.

III. Problem Formulation

The control variables in the HEES system management problem in FEVs are the charging/discharging currents of the battery bank and the supercapacitor bank. More precisely, at any time, the driving power demand, regenerative braking power and the power demand of other electrical systems are given by the position of the gas or the brake paddle and the vehicle dynamics. Therefore, the hybrid storage system controller only concerns how to distribute the total power demand defined in (5) between these two banks.

Depending on the sign of P_{total} and the decision made by the system controller, the HEES system is in one of the six modes as in TABLE I. A positive value of P_b or P_s means the corresponding energy storage bank is providing power and a negative of P_b or P_s means the corresponding storage bank is being charged.

In Mode 1, both storage banks are supplying power to the motor. In Mode 6, both storage banks are being charged. In all other modes, one storage bank is supplying power to or getting charged by the electric motor and the other storage bank. Fig. 5 shows the power flow of Mode 3, where the battery bank is providing power to the motor as well as charging the supercapacitor bank.

TABLE I. Operation Modes of the HEES System.

Mode	Total Power P_{total}	Battery Output Power P_b	SuperCap Output Power P_s
1	$P_{total} \geq 0$	$P_b \geq 0$	$P_s \geq 0$
2	$P_{total} \geq 0$	$P_b < 0$	$P_s \geq 0$
3	$P_{total} \geq 0$	$P_b \geq 0$	$P_s < 0$
4	$P_{total} < 0$	$P_b < 0$	$P_s \geq 0$
5	$P_{total} < 0$	$P_b \geq 0$	$P_s < 0$
6	$P_{total} < 0$	$P_b < 0$	$P_s < 0$

Let $I_{d,b}$ and $I_{d,s}$ ($I_{d,b}, I_{d,s} \geq 0$) denote the discharging current of the battery bank and the supercapacitor bank, respectively, $I_{c,b}$ and $I_{c,s}$ ($I_{c,b}, I_{c,s} \leq 0$) denote the charging current of these two banks. Note that $I_{d,b}$ and $I_{c,b}$ cannot be non-zero simultaneously. The same rule is applied to $I_{d,s}$ and $I_{c,s}$. We combine both equations in one for writing simplicity:

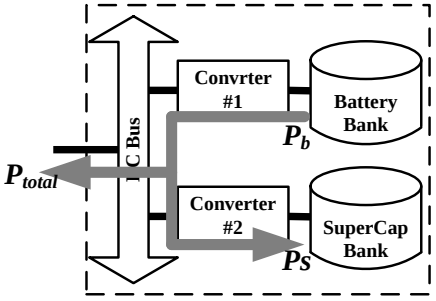

Fig. 5. Mode 3 power flow in the HEES system.

$$P_b = V_b^{OC}\left(I_{d,b} + I_{c,b}\right) - \left(I_{d,b}^2 R_{d,b} + I_{c,b}^2 R_{c,b}\right)$$
$$P_s = V_s^{OC}\left(I_{d,s} + I_{c,s}\right) - \left(I_{d,s}^2 R_{d,s} + I_{c,s}^2 R_{c,s}\right) \quad (6)$$

The equation for P_{total} is thusly:

$$P_{total} = V_b^{OC}\left(\eta_{d,b} I_{d,b} + I_{c,b}/\eta_{c,b}\right) - \left(\eta_{d,b} I_{d,b}^2 R_{d,b} + I_{c,b}^2 R_{c,b}/\eta_{c,b}\right) + V_s^{OC}\left(\eta_{d,s} I_{d,s} + I_{c,s}/\eta_{c,s}\right) - \left(\eta_{d,s} I_{d,s}^2 R_{d,s} + I_{c,s}^2 R_{c,s}/\eta_{c,s}\right) \quad (7)$$

where $\eta_{d,b}$, $\eta_{c,b}$, $\eta_{d,s}$ and $\eta_{c,s}$ are the discharging and charging efficiency of both banks determined by the efficiency of the corresponding converter.

We formally formulate the HEES system management problem in FEVs as an online decision making problem. The goal of HEES system management is to optimize the overall objective:

Objective: Minimizing the total energy drawn from the HEES system for a complete driving cycle:

$$E_{total} = \int_0^T V_b^{OC}(t) I_a(t) \mathrm{d}t + \frac{1}{2} C_s \left(\left(V_s^{OC}(0)\right)^2 - \left(V_s^{OC}(T)\right)^2\right) \quad (8)$$

where T is the total driving time, $I_a(t)$ is the actual charge change rate inside the battery given in (1), and is positive when it is discharging. The online management policy will make decision periodically at each epoch *t* throughout operation.

Find: (at each epoch t)
- The battery bank discharge or charge current: $I_{d,b}(t)$, $I_{c,b}(t)$
- The supercapacitor bank discharge or charge current: $I_{d,s}(t)$, $I_{c,s}(t)$

Given: (at each epoch t)
- Driving power demand $P_d(t)$
- Power demand of other electrical systems $P_e(t)$
- DC bus voltage: $V_{DCBus}(t)$
- Li-ion battery bank state-of-charge: $SoC_b(t)$
- Supercapacitor bank open circuit voltage: $V_s^{OC}(t)$
- Vehicle speed: $v(t)$

Subject to: (at each epoch t)
- Power constraint given by (7).
- Supercapacitor bank terminal voltage constraint:

$$V_{s,min}^{OC} \le V_s^{OC}(t) = V_s^{OC}(0) - \int_0^t \left(I_{d,s}(\tau) + I_{c,s}(\tau)\right) d\tau / C_s \le V_{s,max}^{OC} \quad (9)$$

Note that at any time t, the controller only has the data now and the past data. We do not assume any prior knowledge about the future.

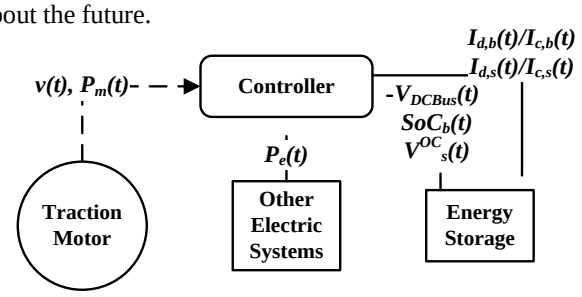

Fig. 6. HEES Control System.

Fig. 6 shows the information and control flow in the HEES control system. The controller gathers information from the motor, other electrical systems, as well as the energy storage system itself. It makes a decision on how to regulate charging/discharging current of battery bank and supercapacitor bank based on this information and the presented algorithm in Section IV. Next it sets outputs of the converters to the desired current levels based on the decision. The controller does not need to have any knowledge about the future driving profile, or the underlying models for the power conversion circuitry, electric motor and the vehicle dynamics. The only required information is the V_b^{OC} versus SoC_b relationship of the Li-ion battery in order to calculate the total energy consumption, which is the optimization objective. Note that if this relationship is unavailable, V_b^{OC} may be substituted with the battery bank's rated terminal voltage as an approximation. Section IV describes how one can learn a judicious control policy without knowing all those models.

IV. Problem Solution

A. Reinforcement Learning Background

Reinforcement learning (RL) is a widely used machine learning technique when the problem of interest can be modeled as an interaction system between an objective-oriented agent and an environment with uncertainty [22]. Fig. 7 illustrates how the agent interacts with the environment: At decision epoch t_k, the system is at state s_k, and the agent chooses action a_k. As a result of the action, the system transitions to state s_{k+1} at time t_{k+1} and yields a reward R_k to the agent. In RL, a decision epoch is defined as the time point when the agent is able to make an action.

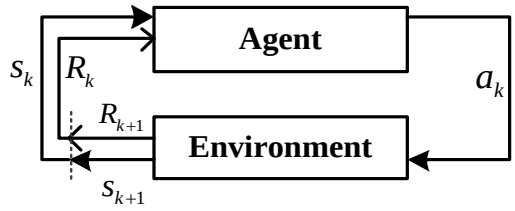

Fig. 7. The agent environment interaction [22].

We denote the state space by S and the set of available actions by A. The reward R is a function of the current state and action: $S \times A \to R$. The ultimate goal of RL is to find a policy $\pi(s) = a$ for the agent, which chooses action $a \in A$ for each state $s \in S$, in order to maximize the *discounted cumulative reward* collected during an infinite time span: $\sum_{i=0}^{\infty} \gamma^i R_i$. Note that the discount factor γ is required not only to make the above summation finite, but more importantly to reflect the uncertainty in the future.

We employ a specific type of RL algorithm, namely TD(λ)-learning [23]. In TD(λ)-learning technique, a Q value, denoted by $Q(s,a)$, is associated with each state-action pair (s,a), which approximates the expected discounted cumulative reward of taking action a starting at state s. At each decision epoch, Q values are updated based on the collected reward during the current time period, and a new action is chosen based on the updated Q values of corresponding state-action pairs.

B. Proposed RL-based Algorithm

To apply the RL method to the HEES system management problem in EVs, we first define the state and action space, and the reward function. Then we present the RL-based algorithm.

1) State, Action and Reward Definition:

State Space: $\{P_{total}\} \times \{v\} \times \{\widehat{v'}\} \times \{V_s^{OC}\} \times \{V_{DCBus}\}$, where $\widehat{v'}$ is the predicted vehicle acceleration between the current and the next decision epoch, and the rest are defined in Section III.

The complexity and convergence speed of TD(λ)-learning algorithm is proportional to the number of state-action pairs; therefore, it is better to include only the most critical inputs (and exclude the rest) in the state space. For example, the battery SoC is excluded in the state space since it does not have significant impact on the power conversion efficiency of converters, compared to the SoC of the supercapacitor bank. V_{DCBus} may be excluded as well if it does not change dramatically over time. Meanwhile, because the predicted vehicle acceleration is strongly correlated with the motor power, we add it to the state space. By predicting future acceleration of the vehicle, we equip the controller with the information of the peak power demand or regenerative power in the near future, which helps the controller to make the decision of taking corresponding actions such as pre-charging or discharging the supercapacitor bank.

Action Space: $\{I_{d,b}/I_{c,b}\}$, where $I_{d,b}$ ($I_{c,b}$) denote the discharge (charge) current of the battery bank and is from a set of predefined values.

We only need to determine the charging or discharging current of the battery bank. The supercapacitor bank current is determined automatically from the total power demand P_{total} using (7). In practice, the supercapacitor bank current is determined by a feedback control mechanism in the sense that we configure Converter #1 in Fig. 3 to supply the desired amount of current and Converter #2 in the same figure to maintain the voltage on the DC bus. For example, if P_{total} is higher than the power provided by both storage banks, the voltage on the DC bus slightly drops, and the supercapacitor bank supplies more power in response to raise the DC bus voltage back to the desired level.

To keep the supercapacitor bank terminal voltage within the lower and upper bounds, we add additional logic to protect the supercapacitor from being over-charged or over-discharged when its terminal voltage is higher than the maximum limit or lower than the minimum limit.

The proposed RL algorithm for the online hybrid energy storage system management problem is invoked periodically with a period of Δt. If the total power demand changes between any two decision epochs, the battery bank current remains the same while the supercapacitor bank current adapts to the change of total power demand because Converter #2 performs voltage regulation of the DC bus.

Reward Function: Reward R_k collected between decision epochs t_{k-1} and t_k is defined as:

$$R_k = -\int_{t_{k-1}}^{t_k} V_b^{OC}(t) I_a(t) \mathrm{d}t - \frac{1}{2} C_s \left(\left(V_s^{OC}(t_{k-1})\right)^2 - \left(V_s^{OC}(t_k)\right)^2 \right) \quad (10)$$

Note that R_k is positive if the HEES system is being charged and negative if the system is being discharged. The reward function is defined as such because the overall objective for the HEES system management problem in FEVs is to minimize the energy drawn from the HEES system.

2) Action Selection

A straightforward approach for action selection is to always choose the action with the highest Q value. If we do so, however, we are at the risk of getting stuck in a sub-optimal solution [24]. A judicious RL agent should exploit the best action known so far to gain rewards while in the meantime exploring all possible actions to find a potentially better choice. We address this *exploration versus exploitation* issue by breaking our learning procedure into two phases: In the exploration phase, *ε-greedy-policy* is adopted, i.e., the current best action is chosen only with probability $1-\varepsilon$. In the exploitation phase, the action with the highest Q value is always chosen.

3) Q-Value Update Equation

TD(λ)-learning algorithm updates every Q value according to the following equation at decision epoch t_k:

$$Q(s,a) \xleftarrow{update} Q(s,a) + \alpha \cdot e(s,a) \cdot \left(R_k + \gamma \max_{a_k} \{Q(s_k,a_k)\} - Q(s,a) \right) \quad (11)$$

where s_k and a_k are the state and action at decision epoch t_k. The coefficient α controls the *learning rate* and γ is the *discount factor*. $e(s,a)$ denotes the *eligibility* of the state-action pair (s,a), reflecting the degree to which the particular state-action pair has been chosen in the recent past. The eligibility e of all state-action pairs is updated at each decision epoch by:

$$e(s,a) \xleftarrow{update} \lambda \cdot e(s,a) + \delta((s,a),(s_k,a_k)) \quad (12)$$

where λ is a constant between 0 and 1, and $\delta(x,y)$ is the Kronecker delta function:

$$\delta(x,y) = \begin{cases} 1, x = y \\ 0, x \neq y \end{cases} \quad (13)$$

In practice, we do not have to update the Q values and eligibility e of all state-action pairs. We only keep a list of M most recent state-action pairs since the eligibility of all other state-action pairs is at most λ^M which is negligible when M is large enough.

4) Vehicle Acceleration Prediction

We first predict the speed of the vehicle at the next decision epoch using a second-order autoregressive (AR) model [25]:

$$\hat{v}(t_{k+1}) = \varphi_1 v(t_k) + \varphi_2 v(t_{k-1}) + \varepsilon_{k+1} \quad (14)$$

where φ_1 and φ_2 are two AR coefficients and ε_k is white noise. Both φ_1 and φ_2 are computed online by estimating the first two elements of the autocorrelation of vehicle speed (denoted by ρ_1 and ρ_2) and solving Yule Walker equations:

$$\begin{cases} \varphi_1 + \varphi_2 \rho_1 = \rho_1 \\ \varphi_1 \rho_1 + \varphi_2 = \rho_2 \end{cases} \quad (15)$$

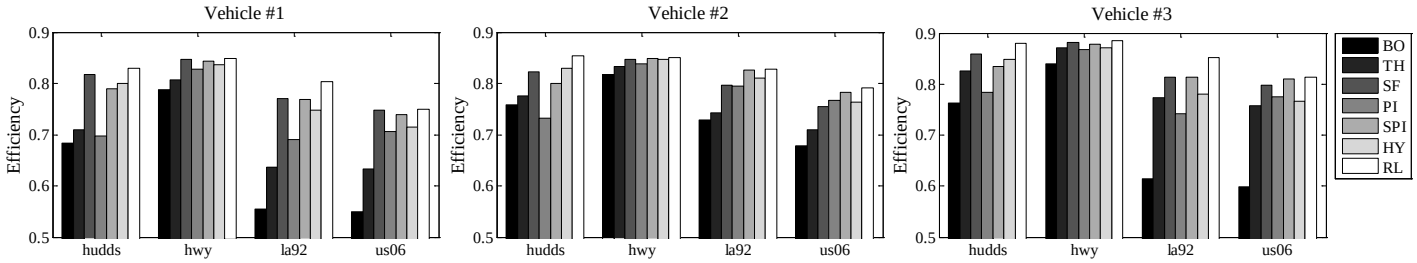

Fig. 8. Electrical energy storage system's efficiency of different vehicles and different driving profiles.

Then the predicted vehicle acceleration between the current decision epoch and the next decision epoch is given by:

$$\hat{v}' = \frac{\hat{v}(t_{k+1}) - v(t_k)}{\Delta t} = \frac{(\varphi_1 - 1)v(t_k) + \varphi_2 v(t_{k-1})}{\Delta t} \quad (16)$$

5) Algorithm Description

In summary, the pseudo code of the above TD(λ)-learning algorithm at each decision epoch is:

Update AR model coefficients for vehicle acceleration prediction
Compute current state s
Find action a that maximizes $Q(s, a)$
Update the Q values of M previous states
Update eligibility e of $M - 1$ previous states and the current state
If in exploration phase
    Randomly select a new action a' with probability ε
Execute action a (or a')

The time complexity of the above algorithm is $O(|A| + M)$ where $|A|$ denotes the number of available actions and M is the number of previous state-action pairs kept in memory.

V. Simulation Results

A. Simulation Setup

We conduct simulations on three different vehicle specifications (as shown in TABLE II.) and four driving profiles (as shown in TABLE III.), using SIMES [30], which has detailed models of various energy storage elements and power converters for validating the model-free RL technique. We adopt the motor efficiency model from ADVISOR [3], theoretical permanent magnet synchronous motor (SM) model in [28], and theoretical induction motor (IM) model in [29], respectively. We use the vehicle dynamics model presented in Section II.2 to convert the driving profile to the driving power demand. Without loss of generality, the power demand of other electrical systems is assumed to be constant as it is very small compared with the driving power demand.

In the proposed RL algorithm, we configure the number of state-action pairs to be 160 and use a control interval of 10s. The Q value of each state-action pair is initialized to 0.

TABLE II. Vehicle specifications.[1]

Vehicle Specs	1	2	3
Mass (kg)	970	1493	2108
Motor Type	SM	SM	IM
Tire Radius (m)	0.287	0.316	0.352
Axle ratio	4.0	7.94	9.73
Li-ion Capacity (kWh)	17.6	24	65
SC Size (F)	5	7	15

[1] The specifications of the three vehicles mainly come from those of a Smart EV, Nissan Leaf and Tesla Model S, respectively. A supercapacitor bank with reasonable capacity is considered for each vehicle.

TABLE III. Driving profile specifications (from [27]).

Profile	Time (s)	Distance (km)	Maximum Speed (km/h)	Maximum Acceleration (m/s2)
hudds	1061	8.9	93	1.96
hwy	766	16.5	96	1.43
la92	1436	15.8	108	3.08
us06	601	12.9	129	3.76

The baseline system and management policies we compare with are as follows, with corresponding parameters optimized for each vehicle model:

1) BO: Li-ion battery only electrical energy storage system;

2) TH: Threshold values are imposed upon the battery bank charging/discharging current. The battery bank current should not exceed the threshold unless the supercapacitor bank terminal voltage goes above $V_{s,max}^{OC}$ or drops below $V_{s,min}^{OC}$;

3) SF: Supercapacitor bank supplies power/get charged first. Battery gets charged/discharged only when the supercapacitor bank terminal voltage goes above $V_{s,max}^{OC}$ or drops below $V_{s,min}^{OC}$;

4) PI: Proportional-Integral (PI) feedback control on V_s^{OC} [16]. Use supercapacitor bank to regulate the voltage on DC bus and use battery bank to perform PI control on V_s^{OC};

5) SPI: Speed dependent PI control on V_s^{OC} [16]. Similar to PI whereas the target voltage is speed-dependent;

6) HY: Hysteretic feedback control on V_s^{OC} where the relationship between the battery bank current and the supercapacitor bank terminal voltage is defined by a hysteretic curve [12].

B. Simulation Results

Fig. 8 shows the electrical energy storage system efficiency (defined as the ratio between energy delivered to the load $\int P_{total} \mathrm{d}t$ and energy drawn from the energy storage system E_{total}) of different vehicles and different driving profiles. As shown in Fig. 8, the efficiency of the battery only system can be as high as around 80% for the hwy profile when the vehicle is driving at relatively constant speed for most of the time. However, its efficiency drops significantly to below 55% in some cases when the vehicle spends more time on quick acceleration. The incorporation of a supercapacitor bank enhances the efficiency by up to 25% using the presented RL algorithm. An advantage of the presented RL algorithm based management policy over other baseline policies is that the RL based

policy consistently outperforms all other baseline policies in all cases of different driving profiles and vehicle specifications, demonstrating that it adapts itself to different driving conditions well.

TABLE IV. ENERGY DISSIPATED IN STORAGE SYSTEMS.

Management Policy	BO	TH	SF	PI	SPI	HY	RL
Energy Dissipated (%)	29.9	25.6	18.2	22.1	18.1	19.2	16.2

The average percentages of energy dissipated in the storage system of different systems and management policies are shown in TABLE IV. The presented RL algorithm reduces energy dissipation by 46% compared to the battery only system and 10% compared to the best baseline management policy. The decrease of energy dissipation in the storage system not only improves the efficiency, but also reduces heat generated in the storage system, thereby enhances the lifetime of Li-ion batteries and lowers security risks.

VI. CONCLUSION

In this paper a model-free reinforcement learning based algorithm to manage the hybrid electrical energy storage systems in electric vehicles is presented. Simulations are conducted using realistic models of energy storage systems and actual parameters of existing commercial electrical vehicles. Simulation results shows that the presented approach improves the efficiency of the energy storage system by up to 25% compared to a Lithium-ion battery only energy storage system and outperforms all other online control policies in all cases with different vehicle specifications and different driving profiles.

REFERENCES

[1] http://www.electricdrive.org/index.php?ht=d/sp/i/20952/pid/20952

[2] C. C. Chan, "An Overview of Electric Vehicle Technology," *Proc. of the IEEE*, vol. 81, no. 9, pp. 1202-1213, 1993.

[3] K. B. Wipke, M. R. Cuddy, and S. D. Burch, "ADVISOR 2.1: a user-friendly advanced powertrain simulation using a combined backward/forward approach," *IEEE Trans. on Vehicular Technology*, vol. 48, pp. 1751-1761, 1999.

[4] M. Pedram, N. Chang, Y. Kim, and Y. Wang, "Hybrid electrical energy storage systems," *Proc. of International Symposium on Low Power Eletronics and Design*, pp. 363-368, Aug. 2010.

[5] A. Affanni, A Bellini, G. Franceschini, P. Guglielmi, and C. Tassoni "Battery choice and management for new-generation electric vehicles," *IEEE Trans. on Industrial Electronics*, vol. 52, no. 5, pp.1343 -1349, 2005.

[6] Q. Xie, X. Lin, Y. Wang, M. Pedram, D. Shin, and N. Chang, "State of health aware charge management in hybrid electrical energy storage systems," *Proc. of Design Automaion and Test in Europe*, Mar. 2012.

[7] http://www.transportation.anl.gov/pdfs/B/805.PDF

[8] J. Y. Wong, *Theory of ground vehicles*. Wiley-Interscience, 2001.

[9] S. Park, Y. Kim, and N. Chang, "Hybrid Energy Storage Systems and Battery Management for Electric Vehicles," *Proc. of Design Automation Conference*, pp. 97:1-97:6, June 2013.

[10] P. Thounthong, S. Raël, and B. Davat, "Control strategy of fuel cell/supercapacitors hybrid power sources for electric vehicle," *Journal of Power Sources*, vol. 158, no. 1, pp. 806-814, 2006.

[11] J. Moreno, M. E. Ortúzar, and J. W. Dixon, "Energy-management system for a hybrid electric vehicle, using ultracapacitors and neural networks," *IEEE Trans. on Industrial Electronics*, vol. 53, no. 2, pp. 614-623, 2006.

[12] J. Cao, and A. Emadi, "A new battery/ultra-capacitor hybrid energy storage system for electric, hybrid and plug-in hybrid electric vehicles," *Proc. of Vehicle Power and Propulsion Conference*, Sep. 2009.

[13] V. Shah, R. Chaudhari, P. Kundu, and R. Maheshwari, "Performance analysis of hybrid energy storage system using hybrid control algorithm with BLDC motor driving a vehicle," *Proc. of 2010 IEEE Joint International Conference on Power Electronics, Drives and Energy Systems*, 2010.

[14] L. Wang, E. G. Collins, and H. Li, "Optimal design and real-time control for energy management in electric vehicles," *IEEE Trans. on Vehicular Technology*, vol. 60, no. 4, pp. 1419-1429, 2011.

[15] M. Ortúzar, J. Moreno, and J. Dixon, "Ultracapacitor-based auxiliary energy system for an electric vehicle: Implementation and evaluation," Proc. *IEEE Trans. on Industrial Electronics*, vol. 54, no. 4, pp. 2147-2156, 2007.

[16] J. M. Miller, U. Deshpande, T. J. Doughertym, and T. Bohn, "Power electronic enabled active hybrid energy storage system and its economic viability," *Proc. of Applied Power Electronic Conference*, pp. 190-198, 2009.

[17] S. M. Lukic, S. G. Wirasingha, F. Rodriguez, J. Cao, and A. Emadi, "Power management of an ultracapacitor/battery hybrid energy storage system in an HEV," *Proc of Vehicle Power and Propulsion Conference*, 2006.

[18] D. Linden and T. B. Reddy, *Handbook of Batteries*. McGrew-Hill Professional, 2001.

[19] 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*, vol. 86, no. 9, pp. 1506-1511, 2009.

[20] O. Barbarisi, F. Vasca, and L. Glielmo, "State of charge Kalman filter estimator for automotive batteries," *Control Engineering Practice*, vol. 14, no. 3, pp. 267-275, 2006.

[21] Y. Kim, N. Chang, Y. Wang, and M. Pedram, "Maximum power transfer tracking for a photovoltaic-supercapacitor energy system," *Proc. of International Symposium on Low Power Eletronics and Design*, pp. 307-312, Aug. 2010.

[22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT Press, Cambridge, MA, 1998.

[23] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, vol. 3, no. 1, pp. 9-44, 1988.

[24] M. Wiering, "Explorations in efficient reinforcement learning," PhD thesis, University of Amserdam, 1999.

[25] S. M. Pandit, and S. M. Wu, *Time series and system analysis with applications*, RE Krieger Publishing Company, 1990.

[26] G. U. Yule, "On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers," *Philosophical Trans. of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 226, pp. 267-298, 1927.

[27] http://www.epa.gov/nvfel/testing/dynamometer.htm#area.

[28] P. Pillay, and R. Krishnan, "Modeling, simulation, and analysis of permanent-magnet motor drives. I. The permanent-magnet synchronous motor drive," *IEEE Trans. on Industry Applications*, vol. 25, no. 2, pp. 265-273, 1989.

[29] S. S. Williamson, A. Emadi, and K. Rajashekara, "Comprehensive efficiency modeling of electric traction motor drives for hybrid electric vehicle propulsion applications," *IEEE Trans. on Vehicular Technology*, vol. 56, no. 4, pp. 1561-1572, 2007.

[30] S. Yue, D. Zhu, Y. Wang, N. Chang, and M. Pedram, "SIMES: A Simulator for Hybrid Electrical Energy Storage Systems," *Proc. of International Symposium on Low Power Eletronics and Design*, Sep. 2013