A Deep Reinforcement Learning Framework for Optimizing Fuel Economy of Hybrid Electric Vehicles

Pu Zhao[1], Yanzhi Wang[2], Naehyuck Chang[3], Qi Zhu[4], Xue Lin[1]
[1]Department of ECE, Northeastern University, Boston, MA 02115, USA
[2]Department of EECS, Syracuse University, Syracuse, NY 13244, USA
[3]School of EE, Korea Advanced Institute of Science and Engineering, Yuseong-Gu, Daejeon 34141, Korea
[4]Department of EECS, Northwestern University, Evanston, IL 60208, USA
zhao.pu@husky.neu.edu, ywang393@syr.edu, naehyuck@cad4x.kaist.ac.kr, qzhu@northwestern.edu, xue.lin@northeastern.edu

Abstract—Hybrid electric vehicles employ a hybrid propulsion system to combine the energy efficiency of electric motor and a long driving range of internal combustion engine, thereby achieving a higher fuel economy as well as convenience compared with conventional ICE vehicles. However, the relatively complicated powertrain structures of HEVs necessitate an effective power management policy to determine the power split between ICE and EM. In this work, we propose a deep reinforcement learning framework of the HEV power management with the aim of improving fuel economy. The DRL technique is comprised of an offline deep neural network construction phase and an online deep Q-learning phase. Unlike traditional reinforcement learning, DRL presents the capability of handling the high dimensional state and action space in the actual decision-making process, making it suitable for the HEV power management problem. Enabled by the DRL technique, the derived HEV power management policy is close to optimal, fully model-free, and independent of a prior knowledge of driving cycles. Simulation results based on actual vehicle setup over real-world and testing driving cycles demonstrate the effectiveness of the proposed framework on optimizing HEV fuel economy.

I. Introduction

The development of electric vehicles, both all-electric ones (aka EVs) and hybrid ones (aka HEVs), has been a mitigation to the roaring demand of fossil fuels by the rapidly growing transportation systems and industry. Nowadays, almost all the automobile manufacturers have released their own EV and/or HEV models. Compared with conventional internal combustion engine (ICE)-propelled vehicles, EVs employ electric motors (EMs), which enable higher energy efficiency and zero tailpipe emission [1, 2]. HEVs employ both ICEs and EMs for propulsion to combine the energy efficiency of EMs and a long driving range of ICEs into the same HEVs [3].

The hybrid propulsion system is comprised of an ICE with its associated fuel tank, and one or more EMs with the associated energy storage system (i.e., the battery pack). The ICE provides the primary propulsion by consuming fuel, whereas the EM converts the electrical energy from the battery pack into the secondary propulsion [4, 5]. The EM not only assists the ICE with extra torque, but also serves as an electricity generator during regenerative braking, which recovers the kinetic energy into electrical energy to charge the battery pack [4]. The EM helps the ICE operate in fuel-efficient regions that further improves the fuel economy.

Comparing to conventional ICE vehicles and EVs, the HEVs have a relatively complicated powertrain. The power management policy, which determines the power split between the ICE and the EM to fulfill the demanded propulsion, is essential to the improved fuel economy of HEVs. Various types of HEV power management policies have been investigated, such as rule-based policies [6, 7], global optimization policies [8, 9], real-time optimization policies [10, 11], and reinforcement learning-based policies [12–14].

Inspired by the recent breakthrough of *deep reinforcement learning* (DRL) [15, 16], we develop a deep reinforcement learning framework of the HEV power management to enhance the fuel economy in this work. Generally speaking, the DRL technique is comprised of an offline *deep neural network* (DNN) construction phase and an online deep Q-learning phase. In the offline phase, a DNN is constructed and trained to infer the Q values for the potentially huge amount of state-action pairs. The deep Q-learning performs the optimal action selection and the Q value update in the online phase.

The innovations and contributions of this proposed framework, compared with the previous HEV power management policies, are as follows: (1) It can converge to the optimal power management policy by using the offline constructed DNN and the online Q-learning, while the rule-based policies are usually far from the optimal. (2) Unlike the global optimization policies, it neither relies on *a priori* knowledge of the driving cycles, nor needs the detailed and accurate HEV modeling. (3) It is possible to arrive at the optimal policy applicable to any types of driving cycles, while the real-time optimization policies are quite sensitive to different types of driving cycles. (4) Compared with reinforcement learning, it is able to handle the high dimensions of the state and action space in the HEV power management with accelerated convergence speed. (5) Also enabled by the DRL, prediction of future driving characteristics is incorporated into the state representation, which further enhances the effectiveness of the proposed DRL framework. Simulation results based on actual vehicle setup over real-world and testing driving cycles demonstrate that the proposed DRL framework can improve the HEV fuel economy by up to 56.3%.

978-1-5090-0602-1/18/$31.00 ©2018 IEEE

II. Related Work

The HEV power management coordinates the operation of ICE and EM, to fulfill the propulsion requirement and at the same time minimize the fuel consumption. Many research works have been conducted on the HEV power management to improve the fuel economy. The rule-based HEV power management strategies have been proposed based on intuition, human expertise or fuzzy logic [6, 7]. These approaches are easy for run-time implementation, but they cannot guarantee any kind of optimality. To overcome the shortcomings of the rule-based approaches, the global optimization methods i.e., dynamic programming algorithms, have been employed for HEV power control [8, 9]. The global optimization methods can achieve optimal fuel consumption for specific trips. However, they require *a priori* knowledge of the driving cycles for specific trips and heavily rely on detailed and accurate HEV modeling. The real-time optimization techniques [10, 11], such as the equivalent consumption minimization strategies (ECMS), have been proposed to transform global optimization into an instantaneous optimization problem. Such techniques are effective for run-time control but quite sensitive to the driving cycles.

Reinforcement learning (RL) [17, 18] provides a powerful tool for the decision-maker to "learn" how to "act" optimally. The decision-maker i.e., *agent* can observe the environment's *state* and take an appropriate *action* according to the observed state. A *reward* will be given to the agent as the result of the chosen action. Stimulated by the reward, the agent targets at deriving a policy, by "learning" from its past experience. RL has been applied to the HEV power management for minimizing fuel cost, total operation cost, or joint control with auxiliary systems [12–14]. RL techniques have guaranteed convergence to the optimal policy, but the convergence speed is related to the dimensions of the state and action space.

With the increasing popularity of neural networks, there are also research works on HEV that apply learning techniques. A learning vector quantization neural network [19] has been proposed to identify the driving cycle style. In [20], a fuzzy neural network has been applied to identify urban driving conditions. The authors of [21] have proposed a solution algorithm to the ECMS and an adaptive neural network for driving cycle recognition is utilized to decrease the sensitivity of the algorithm to driving cycle variations. [22] has developed a neural network based trip modeling.

Recently the breakthroughs of DRL in playing Atari [16] and Alpha Go [15] demonstrate a good example to handle the high dimensional state and action space in complicated control problems. Reference [16] presented the pioneering work of the deep reinforcement learning, which successfully learns control policies directly from high-dimensional sensory inputs. It uses a trained deep convolutional neural network and outperforms all previous approaches on six of the games. In [15], a new approach to computer Go has been proposed, in which 'value networks' are used to evaluate board position and 'policy networks' are used to select moves. A novel combination of supervised learning from human expert games and reinforcement learning from games of a self-play is adopted for training these deep neural networks.

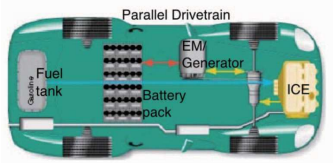

Fig. 1. A parallel hybrid powertrain architecture of an HEV [23].

III. HEV System Architecture

Figure 1 shows a parallel hybrid powertrain, where ICE and EM propel the vehicle in parallel. Thanks to its high energy efficiency and a relatively straightforward structure [5], power management techniques can be applied to it effectively. The HEV models are described below for better understanding of the whole HEV system and faciliating the simulations. However, our approach is totally model-free and does not rely on models to derive a high-fuel economy control policy.

A. HEV Components

A.1 Internal Combustion Engine (ICE)

According to the quasi-static ICE model [24], the ICE fuel efficiency is given by

$$\eta_{ICE}(T_{ICE}, \omega_{ICE}) = T_{ICE} \cdot \omega_{ICE} / (\dot{m}_f \cdot D_f) \quad (1)$$

where T_{ICE} and ω_{ICE} are the torque (in $\mathrm{N} \cdot \mathrm{m}$) and rotational speed (in rad/s) of an ICE, respectively, representing the operating point of the ICE. $\dot{m}_f$ denotes the fuel consumption rate (in g/s) of the ICE, which is a nonlinear function of the operating point, and D_f represents the fuel energy density (in J/g). The ICE operating should remain in a safe range:

$$\omega_{ICE}^{\min} \leq \omega_{ICE} \leq \omega_{ICE}^{\max}, \quad (2a)$$
$$0 \leq T_{ICE} \leq T_{ICE}^{\max}(\omega_{ICE}). \quad (2b)$$

A.2 Electric Motor (EM)

The EM operates in parallel with the ICE. It acts as a motor to propel the vehicle solely or together with ICE. It also operates as a generator to charge the battery pack. The efficiency of the EM [23] is given by

$$\eta_{EM}(T_{EM}, \omega_{EM}) = \begin{cases} (T_{EM} \cdot \omega_{EM})/P_{batt} & T_{EM} \geq 0 \\ P_{batt}/(T_{EM} \cdot \omega_{EM}) & T_{EM} < 0 \end{cases} \quad (3)$$

where T_{EM} and ω_{EM} are the torque and speed of the EM, respectively, and P_{batt} is the output power of the battery pack. When the EM operates as a motor, T_{EM} is positive and the battery pack is being discharged, i.e., $P_{batt} > 0$; when the EM operates as a generator, T_{EM} is negtive and the battery pack is being charged, i.e., $P_{batt} < 0$. To ensure a safe and smooth operation of an EM, the operating point (T_{EM}, ω_{EM}) must be within a certain range as follows:

$$0 \leq \omega_{EM} \leq \omega_{EM}^{\max}, \quad (4a)$$
$$T_{EM}^{\min}(\omega_{EM}) \leq T_{EM} \leq T_{EM}^{\max}(\omega_{EM}). \quad (4b)$$

A.3 Vehicle Dynamics

The vehicle tractive force F_{TR} to support the vehicle speed and acceleration, which are controlled by the brake or accelerator pedal, is derived by

$$F_{TR} = m \cdot a + F_g + F_R + F_{AD}, \tag{5a}$$
$$F_g = m \cdot g \cdot \sin\theta, \tag{5b}$$
$$F_R = m \cdot g \cdot \cos\theta \cdot C_R, \tag{5c}$$
$$F_{AD} = 0.5 \cdot \rho \cdot C_D \cdot A_F \cdot v^2, \tag{5d}$$

where m is the vehicle mass, a is the vehicle acceleration, F_g is the force due to road slope, F_R is the rolling friction force, F_{AD} is the air drag force, θ is the road slope angle, C_R is the rolling friction coefficient, ρ is the air density, C_D is the air drag coefficient, A_F is the vehicle frontal area, and v is the vehicle speed. Given v, a and θ, the tractive force F_{TR} can be derived using (5). Then, the vehicle wheel torque T_{wh} and wheel speed ω_{wh} are given by

$$T_{wh} = F_{TR} \cdot r_{wh}, \tag{6a}$$
$$\omega_{wh} = v / r_{wh}. \tag{6b}$$

The demanded power for propelling the vehicle, i.e., P_{dem} is then calculated as

$$P_{dem} = F_{TR} \cdot v = T_{wh} \cdot \omega_{wh}. \tag{7}$$

A.4 Powertrain Mechanics

The ICE and EM are coupled together through the hybrid powertrain [25], which is commonly comprised of planetary gear sets, to propel the vehicle cooperatively. The speed and torque of the ICE, the EM, and the vehicle wheel satisfy the following speed and torque relation:

$$\omega_{wh} = \frac{\omega_{ICE}}{R(j)} = \frac{\omega_{EM}}{R(j) \cdot \rho_{reg}}, \tag{8a}$$
$$T_{wh} = R(j) \cdot (T_{ICE} + \rho_{reg} \cdot T_{EM} \cdot (\eta_{reg})^{\alpha}) \cdot (\eta_{gb})^{\beta} \tag{8b}$$

where $R(j)$ is the j-th gear ratio, ρ_{reg} is the reduction gear ratio, η_{reg} and η_{gb} are the reduction gear efficiency and the gear box efficiency, respectively. α and β are defined as

$$\alpha = \begin{cases} +1 & T_{EM} \geq 0, \\ -1 & T_{EM} < 0. \end{cases} \tag{9}$$

$$\beta = \begin{cases} +1 & T_{ICE} + \rho_{reg} \cdot T_{EM} \cdot (\eta_{reg})^{\alpha} \geq 0, \\ -1 & T_{ICE} + \rho_{reg} \cdot T_{EM} \cdot (\eta_{reg})^{\alpha} < 0. \end{cases} \tag{10}$$

B. HEV Control

In the actual operation of an HEV, the vehicle speed v and the power demand P_{dem} (or equivalently, the speed v and acceleration a) are determined by the driver through pressing the acceleration or brake pedal. Then the HEV controller needs to control the operation of the ICE, EM and powertrain to make the vehicle meet the target propulsion. Generally, the HEV controller chooses a couple of control variables, such as the battery output power P_{batt} (or equivalently, the battery output current i) and the gear ratio $R(j)$, and then the rest of the variables (i.e., the ICE torque T_{ICE}, the ICE speed ω_{ICE}, the EM torque T_{EM} and the EM speed ω_{EM}) are determined according to the given control variables based on the operating principles of HEV components as discussed previously. This is called the backward-looking optimization approach, which is equivalent to actual HEV management [6, 7].

Fig. 2. The interactions between the agent and environment.

IV. DRL Framework of HEV Power Management

A. Motivations

The complete HEV power management problem exhibits high dimensional state and action space. To deal with such situation, the reinforcement learning-based methods [12–14] need to reduce the state and/or action space in order to make the HEV power management problem tractable. However, the state and action space reduction may decrease the effectiveness of the control and compromise the model-free characteristics. Therefore, we propose the DRL framework of HEV power management, exploiting the capability of DRL to handle the large state and action space in actual control problem. By using the DRL, we can more effectively represent the system state and implement a fully model-free control.

B. Basics of DRL Framework

The learner and decision-maker is called *agent* and the external world of the agent is called *environment*. The agent and environment interact continually with each other. The agent selects actions, and the environment responds to those actions and presents new situations to the agent. The environment also gives rise to rewards, which are specific numerical values that the agent tries to maximize over time. The interaction procedure is illustrated in Figure 2. The DRL technique is comprised of an offline DNN construction phase and an online deep Q-learning phase to solve complicated control problems with a large number of states and a wide action space [16, 26].

The offline phase adopts DNN to derive the correlation between each state-action pair (s, a) of the system under control and its value function $Q(s, a)$. The value function $Q(s, a)$, which represents the expected accumulated (with discount) reward when the system starts at state s and follows action a and certain policy thereafter, is defined as:

$$Q(s, a) = \mathbf{E}\Big[\sum_{k=0}^{\infty} \gamma^k r_k \Big| s_0, a_0\Big] \tag{11}$$

where r_k is the reward received in the k-th time slot and γ is the discount rate.

C. DRL Formulation

In this section, we propose the DRL formulation of the HEV power management problem, formulating the state, action, and reward of DRL to represent the HEV power management problem. We use a slot-time model, i.e., the decision epoch is at the beginning of each time slot with equal length. At each k-th decision epoch i.e., t_k, the HEV system is at state s_k. The HEV controller (i.e., agent) takes an action a_k according to the current state. As a result of the action taken, the agent receives the reward r_k in the k-th time slot i.e., $[t_k, t_{k+1})$.

C.1 State Space

The state space of the DRL is comprised of a finite number of states, each represented by the propulsion power demand, vehicle speed, charge stored in the battery pack, and predicted propulsion power demand for the next time slot, given by

$$S = \Big\{ s = [p_{dem}, v, q, pre]^T \,|p_{dem} \in P_{dem}, \\ v \in V, q \in Q, pre \in \mathrm{P}_{pre} \Big\} \tag{12}$$

where p_{dem} is the power demand for propelling the HEV, v is the vehicle speed, q is the amount of charge stored in the battery pack, and pre is the predicted power demand. P_{dem}, V, Q, and P_{pre} in (12) are, respectively, the finite sets of propulsion power demand levels, vehicle speed levels, levels of charge stored in the battery pack, and predicted power demand levels. Discretization is required when defining these four finite sets. In particular, Q is constructed by discretizing the range of charge stored in the battery pack, i.e., $[q_{min}, q_{max}]$, into a finite number of charge levels:

$$Q = \{q_1, q_2, \cdots, q_N\}, \tag{13}$$

where $q_{min} = q_1 < q_2 < ... < q_N = q_{max}$. Generally, q_{min} and q_{max} are 40% and 80% of the nominal capacity of the battery pack, respectively, for an ordinary HEV.

In the state representation, we incorporate some future driving characteristics (i.e., pre) into consideration for more effective representation and thereby better performance in fuel economy. Incorporating future driving characteristics leads to one additional dimension in the state representation, which increases computation complexity, however, the DRL has the sufficient capability to handle large state space. Also, although both the future velocity and future propulsion power demand could be predicted, predicting the later is more desirable for the DRL. The reason is that the propulsion power demand is more directly related to the action selection than the velocity.

As for the prediction method of future driving characteristics, the randomness of the driving behavior may affect the prediction accuracy. We need a desirable tradeoff between accurate prediction and additional computation complexity. Based on the above mentioned observations, we employ the exponential weighting function to predict the future power demand based on the current measurement data as follows:

$$pre_i \leftarrow (1-\alpha) \cdot pre_{i-1} + \alpha \cdot meas_{i-1} \tag{14}$$

where pre_i is the i-th predicted propulsion power demand, pre_{i-1} is the $(i-1)$-th predicted data, $meas_{i-1}$ is the $(i-1)$-th measured propulsion power demand, and α is the learning rate. Experiments show that the simple function can serve as a desirable prediction method to strike a balance between effective prediction and additional complexity. We are incorporating one-step-ahead prediction into the state space. We can do more steps ahead, but it complicates the state space and prediction accuracy may not be guaranteed. The prediction decision epoch coincides with the deep reinforcement learning decision epoch.

C.2 Action Space

The action space of the DRL is comprised of a finite number of actions, each represented by the discharging current of the battery pack and the gear ratio, i.e.,

$$A = \{a = [i, R(j)]^T \,|i \in I, R(j) \in R\} \tag{15}$$

where an action $a = [i, R(j)]^T$ denotes to discharge the battery pack using current i and choose the j-th gear ratio. The set I contains a finite (discretized) number of discharging current values in the range of $[-I_{max}; I_{max}]$. $i > 0$ denotes discharging the battery pack, and $i < 0$ denotes charging the battery pack. The set R contains all allowable gear ratio values. Usually, there are four or five gear ratio values in total [27].

C.3 Reward Function

The objective of the DRL-based control is to minimize the HEV fuel consumption. Therefore, we define the reward r_k that the agent receives after taking action a_k in state s_k as the negative of the fuel consumption in the k-th time slot, i.e., $-\dot{m}_f \cdot \Delta T$, where ΔT is the length of a time slot, and $\dot{m}_f$ is the fuel consumption rate in that time slot. The DRL agent targets at maximizing the expected return

$$\Sigma_{k=0}^{\infty} \gamma^k \cdot r_k, \tag{16}$$

which is a discounted sum of rewards. Hence, by using the above-mentioned reward function, the overall fuel consumption will be minimized while maximizing the expected return.

D. DRL Procedure

Based on the DRL formulation of the HEV power management problem, we discuss the procedure for deriving the optimal HEV action selection. The proposed DRL procedure of HEV power management comprises an offline DNN construction phase and an online deep Q-learning phase. The key steps are summarized in Algorithm 1.

D.1 Offline DNN Construction

The offline DNN construction phase derives the Q-value estimate for each state-action (s, a) pair. We employ a convolutional neural network as the DNN structure. The first layer is a pooling layer, which reduces the input dimensionality and computation complexity. The following layers are convolutional layers, each followed by rectified linear units (ReLUs) to perform element-wise nonlinearity. The last hidden layer is

Algorithm 1 The DRL Framework of HEV power control

Offline:

1: Simulate the control process using an arbitrary but gradually refined policy for enough long time;
2: Obtain the state transition profile and $Q(s,a)$ value estimates during the process simulation;
3: Store the state transition profile and $Q(s,a)$ value estimates in experience memory D with capacity N_D;
4: Train a DNN with features (s,a) and outcomes $Q(s,a)$;

Online:

5: **for** each execution sequence **do**
6: **for** each decision epoch t_k **do**
7: With probability $1-\varepsilon$ select the action $a_k = \arg\max_a Q(s_k,a)$, otherwise select an action randomly;
8: Perform system control using the chosen action;
9: Observe reward $r_k(s_k,a_k)$ during time period $[t_k, t_{k+1})$ and the new state s_{k+1} at the next epoch;
10: Store transition set (s_k,a_k,r_k,s_{k+1}) in D;
11: Update $Q(s_k,a_k)$ using $\max_{a'} Q(s_{k+1},a')$ and $r_k(s_k,a_k)$ based on the Q-learning updating rule;
12: **end for**
13: Update DNN weight set θ based on the newly updated Q-value estimates in a mini-batch manner;
14: **end for**

a fully-connected layer, followed by ReLUs. The output layer is also a fully-connected layer with outputs for the actions.

To train the DNN, we need enough samples of Q-value estimates of the corresponding state-action (s,a) pairs. The real-world and testing driving cycles are utilized to obtain the Q-value estimates. More specifically, we can drive the vehicle following the driving cycles and use an arbitrary but gradually refined policy for the HEV power management. The state transition profile is recorded in an experience memory D with capacity N_D and also the Q-value estimates are obtained as the accumulative fuel consumption. Based on the stored state transition profiles and Q-value estimates, the DNN is constructed with weight set θ trained using the standard training algorithms [28].

D.2 Online Deep Q-Learning

For the online phase, we adopt the deep Q-learning technique based on the offline-trained DNN to select actions and update Q-values estimates. At each decision epoch t_k of an execution sequence, suppose the HEV system is in state s_k, the DRL agent performs inference using the DNN to obtain the $Q(s_k,a)$ value estimate for each possible action a. The ε-greedy policy selects the action with the maximum $Q(s_k,a)$ value estimate with probability $1-\varepsilon$ and a random action with probability ε. At the next decision epoch, the observed reward $r_k(s_k,a_k)$ after action a_k leads to Q-value updates based on the following updating rule,

$$Q(s,a) \leftarrow Q(s,a) + \alpha \cdot e(s,a) \cdot \delta, \tag{17}$$

where α is a coefficient controlling the *learning rate*, $e(s,a)$ is the *eligibility* of the pair (s,a), and δ is calculated as

$$\delta \leftarrow r_{k+1} + \gamma \cdot \max_{a'} Q(s_{k+1},a') - Q(s_k,a_k). \tag{18}$$

In (18), γ is the discount rate. After the execution of a whole control sequence, the DNN is updated with the newly observed Q-value estimates.

E. Model-Free Property Analysis

Theoretically, the DRL framework could be model-free, i.e., the agent does not require detailed system model to choose actions as long as it can observe the current state and reward as a result of an action previously taken. For the HEV power management, model-free control means that the controller should be able to observe the current state (i.e., propulsion power demand, vehicle speed, battery pack charge level, and power prediction) and the reward (i.e., the negative of fuel consumption in a time slot) as a result of an action (i.e., battery pack discharging current and gear ratio), while the detailed HEV models are not needed by the controller.

In the DRL framework, the HEV controller can use sensors to measure the driver-controller pedal motions to obtain power demand and vehicle speed and observe the current state. The future power demand is predicted according to Eqn. (14). In order to observe the charge level, the Coulomb counting method [29] is required, which is typically realized using a dedicated circuit implementation [30]. The reward can be obtained by measuring the actual fuel consumption. Therefore, the DRL framework is fully model-free i.e., no need for the detailed and accurate HEV modeling.

V. Experimental Results and Discussions

The results of the simulation of the DRL framework for HEV power management are presented in this section. The HEV setup is adopted from the vehicle simulator ADVISOR [23]. The key parameters of the HEV are summarized in Table I. The proposed DRL framework for HEV power management is compared with the rule-based policy [6] and reinforcement learning technique [31] based on both real-word and testing driving cycles. One driving cycle denotes a vehicle speed versus time profile for a specific trip. The experiments make use of both real-world and testing driving cycles developed by different organizations and projects such as U.S. EPA (Environmental Protection Agency) and E.U. MODEM (Modeling of Emissions and Fuel Consumption in Urban Areas project).

First, the fuel consumptions of an HEV with the proposed DRL framework and the rule-based policy are investigated. Table II summarizes the fuel consumptions over some driving cycles. We can observe that the fuel consumptions of the proposed DRL framework are always lower than that of the rule-based policy for all the driving cycles, and the reduction of the fuel consumption can be as high as 56.3%. On average, the fuel consumption can be reduced by 35% with the proposed DRL framework, as shown in the last row of Table II. Then, we compute the MPG values of the proposed DRL framework and the rule-based policy for different driving cycles, as presented in Figure 3. It can be observed that the proposed control framework achieves higher MPG values than the rule-based

TABLE I
HEV KEY PARAMETERS.

Vehicle	Transmission	ICE
$m = 1254kg$	$\rho_{reg} = 1.75$	Peak power 41kW
$C_R = 0.009$	$\eta_{reg} = 0.98$	Peak eff. 34%
$C_D = 0.335$	$\eta_{gb} = 0.98$	**EM**
$A_F = 2m^2$	R(k) = [13.5;7.6;	Peak power 56kW
$r_{wh} = 0.282m$	5.0;3.8;2,8]	Peak eff. 92%
Battery		
Capacity 25 A.h Voltage 240V		

TABLE II
FUEL CONSUMPTION OF THE PROPOSED FRAMEWORK AND THE RULE-BASED POLICY

Driving cycle	Rule-based	Proposed method	reduction
UDDS	412.3g	303.5g	26.4%
NEDC	319.8g	203.5g	36.4%
NYCC	86.1g	37.6g	56.3%
HWFET	364.0g	201.9g	44.5%
Modem1	228.6g	162.6g	28.9%
Modem2	344.9g	225.6g	34.6%
total	1755.7g	1134.7g	35.4%

policy and improves the fuel efficiency. The proposed framework achieves up to 35% MPG improvement.

Furthermore, we demonstrate the effectiveness of introducing the prediction of future propulsion power demand into state representation on the fuel economy. The results are shown in Figure 4. We compare the normalized fuel consumption for several driving cycles under DRL frameworks with and without the prediction. The results show that the framework with prediction decreases fuel consumption and achieves better performance compared with the framework without prediction. The achievements demonstrate the effectiveness of prediction on the fuel economy, and the improvements due to prediction only can be as high as 19%. We also compare our DRL-based framework with the reinforcement learning based power management method [31]. The RL-based method employs $TD(\lambda)$-learning algorithm to derive the power management policy. As shown in Figure 5, we can see DRL-based power control can

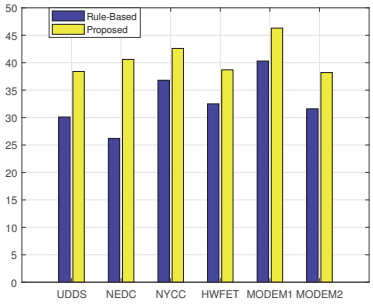

Fig. 3. The MPG values achieved by the proposed DRL framework and the rule-based policy.

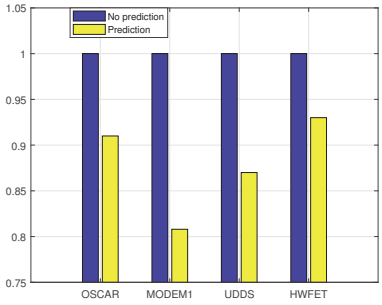

Fig. 4. Normalized fuel consumption of DRL-based HEV control frameworks with and without prediction.

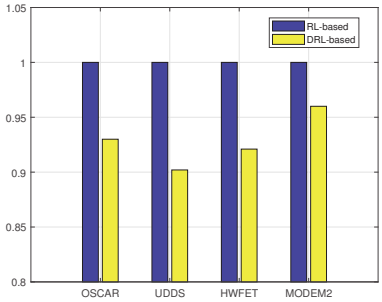

Fig. 5. Normalized fuel consumption of RL-based HEV control framework and DRL-based framework.

achieve better fuel economy than the RL-based framework, and the fuel economy improvement can be as high as 10%. DRL enables larger state space and thereby better control policy can be acheived, while in RL the state space needs to be discretized more coarsely. The results demonstrate the effectiveness of DRL method compared with RL-based framework.

VI. SUMMARY AND CONCLUSIONS

In this work, we propose a DRL based HEV power management framework for optimizing the fuel economy. The DRL technique is comprised of an offline DNN construction phase and an online deep Q-learning phase. The offline phase adopts DNN to derive the correlation between each stateaction pair and its value function. The online Q-learning phase would perform action selection and value updating. The DRL based HEV power management policy is fully modelfree, and independent of a prior knowledge of driving cycles. Simulation results based on actual vehicle setup over realworld and testing driving cycles demonstrate the effectiveness of the proposed framework on optimizing HEV fuel economy.

ACKNOWLEDGMENTS

This work was sponsored in part by NSF awards CCF-1553757, CNS-1704662, and CCF-1733701 and National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2015R1A2A1A09005694).

REFERENCES

[1] C. C. Chan, "The state of the art of electric, hybrid, and fuel cell vehicles," *Proceedings of the IEEE*, vol. 95, no. 4, pp. 704–718, 2007.

[2] S. Pelletier, O. Jabali, and G. Laporte, "Battery electric vehicles for goods distribution: a survey of vehicle technology, market penetration, incentives and practices," *Availabe online: https://www. cirrelt. ca/DocumentsTravail/CIRRELT-2014-43. pdf (accessed on 19 May 2016)*, 2014.

[3] A. Zia, "A comprehensive overview on the architecture of hybrid electric vehicles (hev)," in *2016 19th International Multi-Topic Conference (INMIC)*, Dec 2016, pp. 1–7.

[4] F. R. Salmasi, "Control strategies for hybrid electric vehicles: Evolution, classification, comparison, and future trends," *IEEE Transactions on vehicular technology*, vol. 56, no. 5, pp. 2393–2404, 2007.

[5] C. C. Chan, A. Bouscayrol, and K. Chen, "Electric, hybrid, and fuel-cell vehicles: Architectures and modeling," *IEEE transactions on vehicular technology*, vol. 59, no. 2, pp. 589–598, 2010.

[6] H. Banvait, S. Anwar, and Y. Chen, "A rule-based energy management strategy for plug-in hybrid electric vehicle (phev)," in *American Control Conference, 2009. ACC'09.* IEEE, 2009, pp. 3938–3943.

[7] B. M. Baumann, G. Washington, B. C. Glenn, and G. Rizzoni, "Mechatronic design and control of hybrid electric vehicles," *IEEE/ASME Transactions On Mechatronics*, vol. 5, no. 1, pp. 58–72, 2000.

[8] C.-C. Lin, H. Peng, J. W. Grizzle, and J.-M. Kang, "Power management strategy for a parallel hybrid electric truck," *IEEE transactions on control systems technology*, vol. 11, no. 6, pp. 839–849, 2003.

[9] L. V. Pérez, G. R. Bossio, D. Moitre, and G. O. García, "Optimization of power management in an hybrid electric vehicle using dynamic programming," *Mathematics and Computers in Simulation*, vol. 73, no. 1, pp. 244–254, 2006.

[10] G. Paganelli, M. Tateno, A. Brahma, G. Rizzoni, and Y. Guezennec, "Control development for a hybrid-electric sport-utility vehicle: strategy, implementation and field test results," in *American Control Conference, 2001. Proceedings of the 2001*, vol. 6. IEEE, 2001, pp. 5064–5069.

[11] P. Pisu and G. Rizzoni, "A comparative study of supervisory control strategies for hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 506–518, 2007.

[12] X. Lin, Y. Wang, P. Bogdan, N. Chang, and M. Pedram, "Reinforcement learning based power management for hybrid electric vehicles," in *Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design*. IEEE Press, 2014, pp. 32–38.

[13] Y. Wang, X. Lin, M. Pedram, and N. Chang, "Joint automatic control of the powertrain and auxiliary systems to enhance the electromobility in hybrid electric vehicles," in *Design Automation Conference (DAC), 2015 52nd ACM/EDAC/IEEE*. IEEE, 2015, pp. 1–6.

[14] X. Lin, P. Bogdan, N. Chang, and M. Pedram, "Machine learning-based energy management in a hybrid electric vehicle to minimize total operating cost," in *Computer-Aided Design (ICCAD), 2015 IEEE/ACM International Conference on*. IEEE, 2015, pp. 627–634.

[15] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. Van Den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot *et al.*, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, 2016.

[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," *arXiv preprint arXiv:1312.5602*, 2013.

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

[18] E. Alpaydin, *Introduction to machine learning*. MIT press, 2014.

[19] H. He, C. Sun, and X. Zhang, "A method for identification of driving patterns in hybrid electric vehicles based on a lvq neural network," pp. 3363–3380, 2012.

[20] Y. Tian, X. Zhang, L. Zhang, and X. Zhang, "Fuzzy control strategy for hybrid electric vehicle based on neural network identification of driving conditions," *Control Theory & Applications*, vol. 28, no. 3, pp. 363–369, 2011.

[21] Y. Gurkaynak, A. Khaligh, and A. Emadi, "Neural adaptive control strategy for hybrid electric vehicles with parallel powertrain," in *Vehicle Power and Propulsion Conference (VPPC), 2010 IEEE*. IEEE, 2010, pp. 1–6.

[22] Q. Gong, Y. Li, and Z. Peng, "Power management of plug-in hybrid electric vehicles using neural network based trip modeling," in *2009 American Control Conference*, June 2009, pp. 4601–4606.

[23] N. R. E. Lab, "Advisor 2003 documentation," http://bigladdersoftware.com/advisor/docs/advisor_doc.html.

[24] I. Kolmanovsky and J. Grizzle, "Dynamic optimization of lean burn engine aftertreatment," *Ann Arbor*, vol. 1001, pp. 48 109–2122, 2001.

[25] S. Delprat, J. Lauber, T.-M. Guerra, and J. Rimaux, "Control of a parallel hybrid powertrain: optimal control," *IEEE transactions on Vehicular Technology*, vol. 53, no. 3, pp. 872–881, 2004.

[26] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[27] H. Borhan, A. Vahidi, A. M. Phillips, M. L. Kuang, I. V. Kolmanovsky, and S. Di Cairano, "Mpc-based energy management of a power-split hybrid electric vehicle," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 3, pp. 593–603, 2012.

[28] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[29] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 3. state and parameter estimation," *Journal of Power sources*, vol. 134, no. 2, pp. 277–292, 2004.

[30] T. Instruments, "High-performance battery monitor ic with coulomb counter, voltage, and temperature measurements doc," *ID SLUS521A*, 2002.

[31] X. Lin, Y. Wang, P. Bogdan, N. Chang, and M. Pedram, "Reinforcement learning based power management for hybrid electric vehicles," in *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, Nov 2014, pp. 33–38.