Joint Automatic Control of the Powertrain and Auxiliary Systems to Enhance the Electromobility in Hybrid Electric Vehicles *

Invited

Yanzhi Wang, Xue Lin, and Massoud Pedram
University of Southern California
yanzhiwa@usc.edu, xuelin@usc.edu, pedram@usc.edu

Naehyuck Chang
KAIST
naehyuck@kaist.ac.kr

ABSTRACT

Autonomous driving has become a major goal of automobile manufacturers and an important driver for the vehicular technology. Hybrid electric vehicles (HEVs), which represent a trade-off between conventional internal combustion engine (ICE) vehicles and electric vehicles (EVs), have gained popularity due to their high fuel economy, low pollution, and excellent compatibility with the current fossil fuel dispensing and electric charging infrastructures. To facilitate autonomous driving, an autonomous HEV controller is needed for determining the power split between the powertrain components (including an ICE and an electric motor) while simultaneously managing the power consumption of auxiliary systems (e.g., air-conditioning and lighting systems) such that the overall electromobility is enhanced. Certain (partial) prior knowledge of the future driving profile is useful information for the automatic HEV control. In this paper, methods for predicting driving profile characteristics to enhance HEV power control are first presented. Based on the prediction results and the observed HEV system state (e.g. velocity, battery state-of-charge, propulsion power demand), we propose a reinforcement learning method to determine the power source split between the ICE and electric motor while also controlling the power consumptions of the air-conditioning and lighting systems in the automobile. Experimental results demonstrate significant improvement in the overall HEV system efficiency.

1. INTRODUCTION

Growing concerns about fuel consumption and pollutant emission have forced the automotive industry toward the development of electric and hybrid electric vehicles. Nowadays, most of the major automobile manufacturers have introduced their own electric vehicles (EVs) and/or hybrid electric vehicles (HEVs). Compared with conventional ICE (internal combustion engine)-propelled vehicles, EVs demonstrate much higher energy efficiency and zero tailpipe emission due to the employment of electric motors [1]. However, battery-related concerns have restricted the widespread adoption of EVs [2]. On the other hand, the HEVs that represent a compromise between conventional vehicles and full EVs, can achieve higher fuel economy and lower pollution than conventional ICE-based vehicles and suffer from fewer battery-related concerns compared to EVs.

HEVs feature a hybrid propulsion system comprised of an ICE with an associated fuel tank and one or more electric motors (EMs) with associated energy storage system (batteries), both of which are coupled to the drivetrain. The ICE consumes fuel and provides the primary propulsion, whereas the EM provides the secondary propulsion by consuming electricity stored in the battery pack [3]. Besides assisting the ICE with extra torque, the EM can also serve as an electricity generator to recover kinetic energy during vehicle braking to charge the battery pack. This is called the regenerative braking mechanism, which helps improve the HEV fuel economy [3, 4].

A power management policy for HEVs determines the power split between the ICE and EM to satisfy the speed and torque requirements while ensuring safe and smooth operation of various power components. Since the fuel cost is the major operating cost of an HEV, the majority of previous work on HEV energy management policies aim to reduce fuel consumption and pollution emissions. Rule-based power management strategies have been designed to determine the power split between ICE and EM based on intuition, heuristics, human expertise or fuzzy logic [5, 6]. Although rule-based approaches are effective for real-time supervisory control, their results may be far from optimal. On the other hand, the optimization-based HEV control strategies either minimize the fuel consumption during a trip with a known (deterministic or stochastic) future driving profile [7, 8, 9], or perform real-time supervisory control by converting the amount of battery charge into equivalent fuel consumption [10]. These optimization-based control strategies require *a priori* knowledge of driving cycles and detailed and accurate HEV modeling, and are therefore quite challenging, and likely ineffective, in real-time implementations.

Reinforcement learning (RL) [11] provides a powerful tool for the learning agent (i.e., the decision-maker) to "learn" how to "act" optimally when the exact and accurate system modeling is difficult or even impossible to obtain [12]. The 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, which is a mapping from each possible state to an optimal action, by "learning" from its past experience. The reinforcement learning technique has been applied to the HEV

*This research is supported in part by a grant from National Science Foundation, and the Mid-Career Program and the International Research & Development Program of the NRF of Korea.

DAC '15, June 07 - 11, 2015, San Francisco, CA, USA

http://dx.doi.org/10.1145/2744769.2747933.

energy management problem in [13] for fuel cost minimization. However, there are two important limitations of this work: (i) the proposed RL techniques can be enhanced to make use of certain prediction results about future driving profile characteristics, and (ii) this work and most of the other previous works simply ignore the power consumption of auxiliary systems, which may accounts for 10% - 30% of the overall fuel consumption. Hence, a joint control framework is desirable to simultaneously reduce fuel consumption induced both by propelling the vehicle and by the auxiliary systems[1].

In this paper, we investigate a joint control framework of powertrain and auxiliary systems in an HEV by means of RL, in order to overcome the two aforesaid shortcomings. We minimize fuel cost induced both by propelling the vehicle and by the auxiliary systems since both are critical parts in the overall fuel consumption, and meanwhile maximize a total utility function (representing the degree of desirability) of the auxiliary systems. Unlike some previous approaches, the learning process does not require complete *a priori* information about driving profiles and uses only partial information about the HEV drivetrain modeling, i.e., it can be partially model-free. The learning process properly determines the operating modes of the HEV components, such as battery discharging/charging power/current, gear ratio, operating power of auxiliary systems, etc., based on the proper definition of "states". We properly determine the reward of the RL agent such that the objective of the RL agent coincides with our goal of both minimizing the overall fuel consumption and maximizing the total utility function of the auxiliary systems. The TD(λ)-learning algorithm [11] is employed as the RL algorithm due to its higher convergence rate and higher performance in non-Markovian environment.

In order to further enhance the effectiveness of the RL framework, we incorporate prediction of future driving profile characteristics. The prediction results will serve as a part of the "state" classification in the main RL algorithm, and can enhance the performance of the RL agent because certain partial information of the future characteristics can be provided [11]. An exponential weighting function, although quite simple, can serve as a desirable prediction method of future driving profile characteristics, in order to strike a balance between effectiveness in prediction and additional complexity in the RL algorithm. Simulation results over real-world and testing driving cycles demonstrate the effectiveness of the proposed RL-based joint HEV control mechanism.

2. SYSTEM DESCRIPTION

Although this work aims at designing a smart and (partially) model-free HEV controller that does not need a precise modeling of various HEV components, it is still necessary to understand the fundamental principles of HEV operations. Without loss of generality, we design our power management policy based on the parallel HEV configuration as the one of the most widely employed configurations in state-of-the-art HEVs [4], in which the ICE and EM can deliver power in parallel to drive the wheels. There are five operation modes in a parallel HEV, depending on the energy flows: (i) only the ICE propels the wheels/vehicle, (ii) only the EM propels the wheels/vehicle, (iii) both the ICE and EM propel the wheels/vehicle, (iv) the ICE propels the vehicle and simultaneously drives the EM to charge the battery pack, and (v) the EM charges the battery pack during braking, i.e., the regenerative braking mode. Once again, auxiliary systems are important parts of an HEV and consume a significant part of overall fuel consumption.

2.1 HEV Component Analysis

[1]Although the auxiliary systems are typically powered by battery, the battery is ultimately charged by consuming fuels.

2.1.1 Internal Combustion Engine (ICE)

According to the quasi-static ICE model [14], the ICE fuel efficiency is given by

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

where T_{ICE} and ω_{ICE} are torque (in N·m) and revolution speed (in rad/s) of the ICE, respectively, which represent the operating point of the ICE; $\dot{m}_f$ is the fuel consumption rate (in g/s) of ICE, which is a nonlinear function of the operating point; D_f is the fuel energy density (in J/g). The following constraints should be satisfied to ensure safe and smooth operation of the ICE:

$$\begin{aligned} \omega_{ICE}^{min} \le\ & \omega_{ICE} & \le \omega_{ICE}^{max}, \\ 0 \le\ & T_{ICE} & \le T_{ICE}^{max}(\omega_{ICE}). \end{aligned} \tag{2}$$

2.1.2 Electric Motor (EM)

The EM can operate either as a motor to propel the vehicle or as an electric generator to charge the battery pack. The efficiency of the EM is defined by

$$\eta_{EM}(T_{EM},\omega_{EM}) = \begin{cases} (T_{EM}\cdot\omega_{EM})/(P_{batt}-p_{aux}) & T_{EM}\ge 0 \\ (P_{batt}-p_{aux})/(T_{EM}\cdot\omega_{EM}) & T_{EM}<0 \end{cases} \tag{3}$$

where T_{EM} and ω_{EM} denote the torque and speed of the EM, respectively, P_{batt} is the output power of battery pack, and p_{aux} is the operating power of auxiliary systems. Obviously $P_{batt}-p_{aux}$ is the input power of the EM. When the EM operates as a motor, T_{EM} is positive and $P_{batt}-p_{aux}>0$; when the EM operates as a generator, T_{EM} is negative and $P_{batt}-p_{aux}<0$. The following constraints should be satisfied to ensure safe and smooth operation of the EM:

$$\begin{aligned} 0 \le\ & \omega_{EM} & \le \omega_{EM}^{max}, \\ T_{EM}^{min}(\omega_{EM}) \le\ & T_{EM} & \le T_{EM}^{max}(\omega_{EM}). \end{aligned} \tag{4}$$

2.1.3 Vehicle Dynamics

The electric vehicle is assumed to be a rigid body with four wheels, and the vehicle mass is assumed to be concentrated in a single point. The tractive force F_{TR} to support the vehicle speed and acceleration (which are set by the driver by pressing the braking or acceleration pedals) satisfies

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

where m is the vehicle mass, a is the vehicle acceleration, F_g is the force due to the 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 frontal area of vehicle, and v is the vehicle speed. Given v, a, and θ, the tractive force F_{TR} can be derived using (5). Then, the wheel torque T_{wh} and wheel speed ω_{wh} are related to F_{TR}, v, and the wheel radius r_{wh} as given by

$$\begin{aligned} T_{wh} &= F_{TR}\cdot r_{wh}, \\ \omega_{wh} &= v/r_{wh}. \end{aligned} \tag{6}$$

The demanded power to propel the vehicle, denoted by p_{dem}, is given by

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

2.1.4 Drivetrain Mechanics

The ICE and EM are coupled through the drivetrain to propel the vehicle. The speed and torque of the ICE, the EM, and the wheel

satisfy the following relationship:

$$\omega_{wh} = \frac{\omega_{ICE}}{R(k)} = \frac{\omega_{EM}}{R(k)\cdot\rho_{reg}} \quad (8)$$
$$T_{wh} = R(k)\cdot\left(T_{ICE}+\rho_{reg}\cdot T_{EM}\cdot(\eta_{reg})^{\alpha}\right)\cdot(\eta_{gb})^{\beta}$$

where

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

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

In (8), $R(k)$ is the gear ratio of the k-th gear (there are often a total of four or five gear ratios), ρ_{reg} is the reduction gear ratio, and η_{reg} and η_{gb} are the efficiencies of the reduction gear and the gear box, respectively.

2.1.5 *Auxiliary Systems*

The auxiliary system of HEV is comprised of lighting, air conditioning (or more generally, heating, ventilation, and air conditioning or HVAC), and other battery-powered systems such as GPS. The auxiliary systems may account for 10% - 30% of the overall fuel consumption for an ordinary (fuel-based) vehicle. For HEVs and EVs, it is projected that auxiliary systems will take a larger portion of the overall energy consumption partly because heating of an ordinary vehicle can be partially achieved by the heated internal combustion engine. Hence, the power consumption of auxiliary systems needs to be jointly considered and optimized with the powertrain control for an HEV in order to achieve the global optimal solution. For example, if the battery stored charge is not enough or the battery output power is relatively large, it is desirable to limit the power consumption of auxiliary systems. The effect of auxiliary systems (or more specifically, the HVAC module) can be compensated later after the battery is charged by the ICE or when the battery output power is reduced.

Let p_{aux} denote the total operating power of auxiliary systems, which is a control variable of HEV controller (and partially for the driver.) We adopt a *utility function* $f_{aux}(p_{aux})$ to represent the total *satisfaction level* when applying operating power p_{aux} to the auxiliary systems, which is widely adopted in modeling HVAC systems [15][2]. The utility function is general in the sense that it demonstrates the combination of effects of multiple auxiliary systems such as lighting, HVAC, and other battery-powered systems. In general, the utility function is a uni-modal (quasi-concave) function since neither too high power consumption nor too low power consumption is desirable for the auxiliary system components (for example, too high power consumption for the HVAC means either too hot or too cold, and vice versa.) The utility function $f_{aux}(p_{aux})$ can be either inferred from driver behaviors (e.g., the target temperature set by the driver) or from past learning experiences, and may vary from time to time. The goal of HEV controller is to maximize the total utility function value over the whole driving profile.

2.2 HEV Control Flow

In reality, it is the driver that determines the speed v and the propulsion power demand $p_{dem} = \omega_{wh}\cdot T_{wh}$ profiles (or equivalently, the speed v and acceleration a profiles) for propelling the HEV through pressing the acceleration or brake pedal. Then, the HEV controller determines the operation of ICE, EM, drivetrain, and auxiliary systems[3] so that the HEV meets the target performance (i.e., speed v and acceleration a) and a certain objective function is maximized. This is called the *backward-looking* optimization approach and is equivalent to actual HEV management [5, 6].

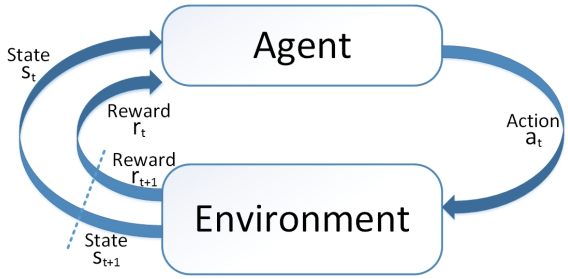

Figure 1: The interactions between agent and environment in RL framework.

In the actual HEV control process, the HEV controller chooses a few control variables, such as battery output power P_{batt} (or equivalently, the battery discharging/charging current i), the gear ratio $R(k)$, and the operating power p_{aux} of the auxiliary systems. The remaining of variables, including the ICE torque T_{ICE} and speed ω_{ICE}, EM torque T_{EM} and speed ω_{EM}, become associate (dependent) variables, the values of which are determined by P_{batt}, $R(k)$, and p_{aux} according to the operating principles of HEV discussed in Section 2.1.

The majority of HEV control strategies in the reference papers, such as dynamic programming-based strategy, model predictive control strategy, and equivalent consumption minimization strategy (ECMS), rely on very detailed HEV system modeling. There are also model-free or partially model-free HEV control strategies that do not rely on detailed HEV system modeling or only need partial HEV modeling, which are preferred due to their flexibility and generality, and ease in implementation. For example, rule-based control strategies only require battery modelings. The proposed RL-based HEV control strategy is also a model-free or partially model-free HEV control framework.

3. REINFORCEMENT LEARNING BASICS

Reinforcement learning (RL) provides a mathematical framework for discovering and learning strategies that map situations onto actions with the goal of maximizing a cumulative reward function [11]. In the RL framework, the learner and decision-maker is called the *agent* and everything outside the agent is called the *environment* (which interacts with the agent). The agent and environment interact continually, the agent selecting actions and environment responding to these actions and presenting new situations to the agent. The environment also gives rise to rewards to the agent, which are special numerical values that the agent tries to maximize over the optimization period.

Figure 1 illustrates the agent-environment interaction in the RL framework at each of a sequence of discrete time steps $t = 0, 1, 2, \ldots$. At each time step t, the agent observes some representation of the environment *state* $s_t \in \mathcal{S}$, and on that basis takes an *action* $a_t \in \mathcal{A}$, where $\mathcal{S}$ and $\mathcal{A}$ are the sets of possible states and actions (in every state), respectively. One time step later, the agent receives a numerical *reward* $r_{t+1} \in R$ and finds the environment in a new state s_{t+1} as a consequence of the action taken.

A policy of the agent, denoted by π, is a mapping from each state $s \in \mathcal{S}$ to action $a \in \mathcal{A}$ that specifies the action $a = \pi(s)$ that the agent will choose in state s. The ultimate goal of the agent is to find the

[2]Please note that this utility function is a simplified version of the actual utility function since the actual utility function of HVAC is not only a function of the instantaneous power consumption but also depends on the previous temperature.

[3]Often the driver also determines operation of auxiliary systems partly, such as lighting systems.

optimal policy, such that the *value function*

$$V^{\pi}(s) = E\left\{\sum_{k=0}^{\infty}\gamma^{k}\cdot r_{t+k+1} \mid s_t = s\right\} \tag{11}$$

is maximized for each state $s \in \mathcal{S}$.

The value function $V^{\pi}(s)$ is the *expected return* when system starts in state s at time t and follows policy π thereafter. $0 < \gamma < 1$ is a parameter named 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, γ reflects the uncertainty and discount in the future [11]. r_{t+k+1} is the reward received at time step $t+k+1$.

4. RL-BASED JOINT CONTROL FRAMEWORK OF POWERTRAIN AND AUXILIARY SYSTEMS

In this section, we present the motivation and details of the proposed RL-based joint control framework of powertrain and auxiliary systems.

4.1 Motivations of Using RL for Joint HEV Control

We use reinforcement learning for the joint control framework of powertrain and auxiliary systems due to the following reasons. (i) HEV energy management policies aim to minimize the total fuel consumption (and maximize the cumulative utility function of auxiliary systems) during a whole driving cycle rather than the instantaneous fuel consumption rate (or instantaneous objective function value) at a certain time step, which is suitable for RL since the latter also aims to optimize an expected cumulative return instead of an immediate reward (11). (ii) During a driving cycle, the change of vehicle speed, power demand, battery charge level (and predicted future driving profile) necessitates different HEV operating modes, which is suitable for RL since an RL agent takes different actions depending on the current state. (iii) The actual driving cycles are non-stationary. Hence, the RL technique is more suitable for the joint HEV power management framework than other optimization methods.

Most of the previous work on HEV power management neglect the power consumption of auxiliary systems (including HVAC, lighting, etc.), which may account for 10% - 30% of the overall fuel consumption of the HEV. The power management results may be sub-optimal by neglecting this important portion of power consumption. In order to mitigate this shortcoming, we aim to develop a more effective joint control framework for HEV propulsion and auxiliary systems, to minimize the overall fuel consumption and maximize the overall objective function of auxiliary systems.

So as to further enhance the effectiveness of the RL-based joint control framework, we incorporate prediction of future driving profile characteristics. The prediction results can serve as a part of the "state" in the RL-based control algorithm, and can enhance the performance of the RL agent by providing partial information of the future characteristics of driving profiles. Details are described in the next subsection.

4.2 Prediction of Future Driving Profile Characteristics

In this subsection, we describe the prediction method of future driving profile characteristics. As one know, the prediction cannot be highly accurate because of the following two reasons: (i) the prediction accuracy is inherently limited by the difficulty and randomness in driving profile prediction, and (ii) a more accurate prediction result (with higher precision levels) will significantly add computation complexity and reduce convergence rate of the RL algorithm, because the prediction results will add at least one dimension to the state space of the RL algorithm. Hence, we need to achieve a desirable tradeoff between the effectiveness in prediction and additional complexity in the RL algorithm.

Another important observation is that although we could predict both the future velocity and future propulsion power demand (or acceleration), predicting the later is more desirable for the RL agent. This is because the propulsion power demand is more directly related to the action chosen (e.g., the battery discharging current, gear ratio, etc.) by the RL agent than the velocity.

Based on the above-mentioned two observations, we adopt the exponential weighting function, which predicts the future data (propulsion power demand) based on the current measurement data as follows:

$$pre_i \leftarrow (1-\alpha)\cdot pre_{i-1} + \alpha \cdot meas_{i-1}, \tag{12}$$

where pre_i is the i-th predicted future data (propulsion power demand), pre_{i-1} is the $(i-1)$-th predicted data, $meas_{i-1}$ is the $(i-1)$-th measured data (propulsion power demand), and α is the learning rate. Experiments show that the exponential weighting function, though quite simple, can serve as a desirable prediction method to strike a balance between effectiveness in prediction and additional complexity in the RL algorithm. Other methods such as artificial neural network (ANN) can also be utilized for future driving profile prediction. Details are omitted due to space limitation.

4.3 Details of the RL Process

4.3.1 State Space

We define the state space of the RL technique as a finite number of states, each represented by the propulsion power demand, vehicle speed, battery pack stored charge level, and predicted driving profile characteristics, given by

$$\mathcal{S} = \left\{ s = [p_{dem}, v, q, pre]^T \mid p_{dem} \in \mathcal{P}_{dem}, v \in \mathcal{V}, q \in \mathcal{Q}, pre \in \mathcal{P}re \right\} \tag{13}$$

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 characteristics of future driving profiles.

Different actions may be taken in different states. For instance, if the propulsion power demand level is negative, i.e., during vehicle braking, the action chosen by the agent (HEV controller) should be charging the battery by using the EM as a generator. On the other hand, if the propulsion power demand level is a very large positive value, the selected action should be discharging the battery to power the EM, which propels the vehicle and provides power for auxiliary systems in assistance with ICE.

A RL agent is able to observe a state from online measurement. In the actual implementation, the current propulsion power demand level p_{dem} and vehicle speed v are obtained by sensors to measure the driver-controlled pedal motion, and future driving profile characteristics are predicted using methods described above. However, the charge level q cannot be obtained from online measurement of battery's terminal voltage, because the terminal voltage of battery pack changes with the charging/discharging current and therefore it is not an accurate indicator of q [16]. In order to observe the charge level q, the Coulomb counting method [17] is required by the RL agent, which is typically realized using a dedicated circuit implementation [18].

$\mathcal{P}_{dem}$, $\mathcal{V}$, $\mathcal{Q}$, and $\mathcal{P}re$ in (13) are, respectively, the finite sets of propulsion power demand levels, vehicle speed levels, levels of charge stored in the battery pack, and predicted driving profile characteristics. Discretization is required when defining these four finite sets. In particular, $\mathcal{Q}$ 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:

$$\mathcal{Q} = \{q_1, q_2, ..., q_N\} \tag{14}$$

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 battery pack, respectively, for an ordinary HEV (charge-sustaining mode).

4.3.2 Action Space and Reduced Action Space

We define the action space of the RL framework as a finite number of actions, each represented by the discharging current of battery pack, the gear ratio, and operating power of auxiliary systems, i.e.,

$$\mathcal{A} = \left\{ a = [i, R(k), p_{aux}]^T | i \in I, R(k) \in R, p_{aux} \in P_{aux} \right\} \tag{15}$$

where an action $a = [i, R(k), p_{aux}]^T$ chosen by the RL agent denotes to discharge the battery using current i, choose the k-th gear ratio, and apply operating power p_{aux} for the auxiliary systems.

The set I in (15) contains within it 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, which depend on the powertrain design. Usually, there are four or five gear ratio values in total [8]. Finally, P_{aux} represents a finite (discretized) set of operating power levels of auxiliary systems.

Alternatively, we define a reduced action space $\mathcal{A}_{re}$, in which an action $a_{re} = [i]^T$ only accounts for the discharging/charging current of the battery. Using this reduced action space, the gear ratio $R(k)$ and auxiliary systems operating power p_{aux} can be selected by solving an optimization problem such that the instantaneous reward function (as shall be discussed later) can be maximized. Since the computation complexity and convergence speed of RL algorithms are proportional to the number of state-action pairs [19], the reduced action space $\mathcal{A}_{re}$ significantly reduces the computation complexity and increase convergence speed of the RL algorithm. Another advantage of the reduced action space is that discretization of p_{aux} in the original action space is no longer required, which in turn enhances the control precision and performance. Of course, there is a side effect that the reduced action space relies on (partial) HEV component modeling when solving the optimization problem. However, due to the significant advantages, we would suggest to use the reduced action space $\mathcal{A}_{re}$ for reduced computation complexity and increased convergence rate, and make the RL agent partially model-free.

4.3.3 Reward Function

The objective of the RL-based joint control mechanism is to minimize the total fuel cost, induced by both propelling the vehicle and auxiliary systems, and to maximize the overall utility function value of the auxiliary systems over the whole driving profile. Therefore, we define the reward r that the agent receives after taking action a in state s as the negative of the fuel consumption plus the utility function value of the auxiliary systems at that time step, i.e., $(-\dot{m}_f + w \cdot f_{aux}(p_{aux})) \cdot \Delta T$, where ΔT is the length of a time step, $\dot{m}_f$ is the fuel consumption in that time step, and w is a weighting factor determining the relative importance of fuel consumption and the auxiliary system utility function. The RL agent targets at maximizing the expected return (11), which is a discounted sum of rewards. Hence, by using the above-mentioned reward function, the overall fuel consumption will be minimized and the overall utility function value will be maximized (of course they are connected through the weighting factor w) while maximizing the expected return.

In an actual reinforcement learning implementation, the RL agent (HEV controller) should be aware of the reward it receives after taking an action, since the observation of reward is critical in deriving the optimal policy. In the above-mentioned reward definition, $\dot{m}_f$ can be obtained by measuring the fuel consumption directly, and utility function $f_{aux}(p_{aux})$ can be either inferred from driver behaviors (e.g., the target temperature set by the driver) or from past learning experience.

4.3.4 TD(λ)-Learning Algorithm for Joint HEV Control

We employ the TD(λ)-learning algorithm [11] to derive the optimal policy for the joint control of powertrain and auxiliary systems, because of (i) its relatively higher convergence rate and (ii) higher performance in non-Markovian environment. In TD(λ)-learning, a Q value, denoted by $Q(s,a)$, is associated with each state-action pair (s,a), where a state s is represented by the propulsion power demand p_{dem}, the vehicle speed v, the battery charge q, and predicted future driving profile characteristics pre, and an action a can be either a complete action or a reduced action as described before. The $Q(s,a)$ value approximates the expected (discounted) cumulative reward of taking action a in state s. Details of the TD(λ) algorithm is summarized as follows.

In the TD(λ)-learning procedure, the Q values are initialized arbitrarily in the beginning of execution. At each time step t, the agent selects an action a_t for current state s_t based on the $Q(s,a)$ values. To avoid the risk of getting stuck at a sub-optimal solution, the *exploration versus exploitation policy* [11] is employed for the action selection, i.e., the agent does not always select the action a with the maximum $Q(s_t,a)$ value for current state s_t. Instead, the current best action is chosen only with probability of $1-\varepsilon$, and the other actions are chosen with equal probability.

Suppose that the chosen action is a_t at time step t, the learning agent observes a new state s_{t+1} and receives reward r_{t+1} at time step $t+1$. Then based on the observed s_{t+1} and r_{t+1}, the agent updates Q values for each state-action pair (s,a), in which the *eligibility* $e(s,a)$ of each state-action pair (s,a) is updated and effectively utilized during Q value updating. The eligibility $e(s,a)$ of a state-action pair (s,a) reflects the degree to which the particular state-action pair has been chosen in the recent past, where λ is a constant between 0 and 1. Due to the usage of the eligibility of state-action pairs, in practice we do not need to update 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 for a large enough M.

Algorithm 1 TD(λ)-Learning Algorithm

```
Initialize Q(s,a) arbitrarily for all the state-action pairs.
for each time step t do
    Choose action a_t for state s_t using the exploration-exploitation policy.
    Take action a_t, observe reward r_{t+1} and next state s_{t+1}.
    δ ← r_{t+1} + γ · max_{a'} Q(s_{t+1}, a') − Q(s_t, a_t).
    e(s_t, a_t) ← e(s_t, a_t) + 1.
    for all state-action pair (s,a) do
        Q(s,a) ← Q(s,a) + α · e(s,a) · δ.
        e(s,a) ← γ · λ · e(s,a).
    end for
end for
```

5. EXPERIMENTAL RESULTS

We simulate the operation of an HEV, the model of which is developed in the vehicle simulator ADVISOR [20]. The key parameters of the HEV are summarized in Table 1. We test our joint HEV control framework and compare with the reinforcement learning (RL) policy [13] and the rule-based policy [5]. We use both real-world and testing driving trip profiles, which are developed and provided by different organizations and projects such as

Table 1: HEV key parameters.

Vehicle	Transmission	ICE
$m = 1254$ kg	$\rho_{reg} = 1.75$	peak power 41kW
$C_R = 0.009$	$\eta_{reg} = 0.98$	peak eff. 34%
$C_D = 0.335$	$\eta_{gb} = 0.98$	**EM**
$A_F = 2$ m^2	$R(k) = [13.5; 7.6;$	peak power 56kW
$r_{wh} = 0.282$ m	$5.0; 3.8; 2.8]$	peak eff. 92%

battery
Capacity 25A·h Voltage 240V

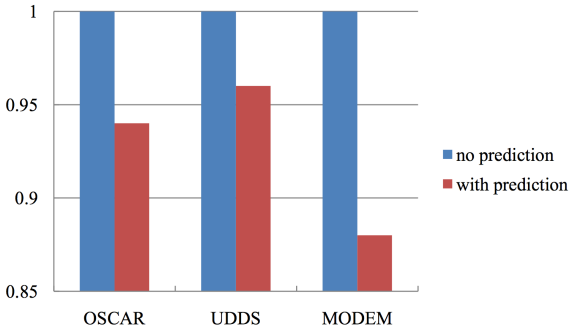

Figure 2: Normalized fuel consumption of RL-based HEV control frameworks with and without prediction.

U.S. EPA (Environmental Protection Agency) and E.U. MODEM (Modeling of Emissions and Fuel Consumption in Urban Areas project).

One improvement of this work over [13] is that we introduce prediction of future driving profile characteristics. First, we measure the fuel economy improvement due to the prediction only. Figure 2 shows the normalized fuel consumption for three driving profiles (i.e., OSCAR, UDDS, and MODEM) under HEV control frameworks with and without the prediction. The fuel economy improvement due to prediction only can be as high as 12%.

Furthermore, we compare the proposed joint control framework with the rule-based policy [5]. We assume the most desirable power consumption of the auxiliary systems is 600W and more or less power consumption from the auxiliary systems will reduce the value of the utility function $f_{aux}(p_{aux})$. Table 2 shows the accumulation of the reward function $(-\dot{m}_f + w \cdot f_{aux}(p_{aux})) \cdot \Delta T$ over whole driving profiles. Please note that $-\dot{m}_f$ is a negative value and also the reward function value is negative. We can observe the proposed control framework always achieves higher reward function values than the rule-based policy. To compare the fuel economy of the proposed and rule-based policy, Figure 3 shows the corresponding MPG values from the two policies for different driving profiles. The proposed framework achieves up to 29% MPG improvement.

6. CONCLUSIONS

In this paper, we first present methods for predicting driving profile characteristics to enhance HEV power control. Based on the prediction results and the observed HEV system state (e.g. velocity, battery state-of-charge, propulsion power demand), we propose a reinforcement learning method to determine the power source split between the ICE and EM while also controlling the power consumptions of the air-conditioning and lighting systems in the automobile. Experimental results demonstrate up to 29% MPG value improvement.

Table 2: Reward function values from the proposed joint control framework and the rule-based policy.

	Proposed	Rule-based
OSCAR	-275.76	-337.50
UDDS	-754.85	-849.25
SC03	-284.14	-319.66
HWFET	-741.12	-861.68

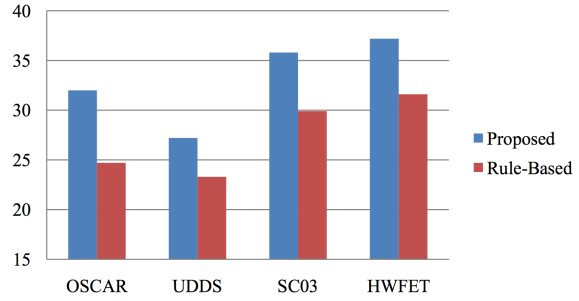

Figure 3: The MPG values achieved by the proposed joint control framework and the rule-based policy.

7. REFERENCES

[1] C. Chan, "The state of the art of electric, hybrid, and fuel cell vehicles," *Proceedings of the IEEE*, 2007.

[2] S. Pelletier and et al., "Battery electric vehicles for goods distribution: A survey of vehicle technology, market penetration, incentives and practices," 2014.

[3] F. R. Salmasi, "Control strategies for hybrid electric vehicles: Evolution, classification, comparison, and future trends," *Vehicular Technology, IEEE Transactions on*, 2007.

[4] C.-C. Chan and et al., "Electric, hybrid, and fuel-cell vehicles: Architectures and modeling," *Vehicular Technology, IEEE Trans*, 2010.

[5] H. Banvait and et al., "A rule-based energy management strategy for plug-in hybrid electric vehicle (phev)," in *ACC'09*.

[6] B. M. Baumann and et al., "Mechatronic design and control of hybrid electric vehicles," *Mechatronics, IEEE/ASME Trans*, 2000.

[7] L. V. Pérez and et al., "Optimization of power management in an hybrid electric vehicle using dynamic programming," *Mathematics and Computers in Simulation*, 2006.

[8] H. Borhan and et al., "Mpc-based energy management of a power-split hybrid electric vehicle," *Control Systems Technology, IEEE Trans*, 2012.

[9] S. J. Moura and et al., "A stochastic optimal control approach for power management in plug-in hybrid electric vehicles," *Control Systems Technology, IEEE Trans*, 2011.

[10] S. Delprat and et al., "Optimal control of a parallel powertrain: from global optimization to real time control strategy," in *Vehicular Technology Conference, 2002.*

[11] R. S. Sutton, "Learning to predict by the methods of temporal differences," *Machine learning*, 1988.

[12] E. Alpaydin, *Introduction to machine learning*. MIT press, 2004.

[13] X. Lin and et al., "Reinforcement learning based power management for hybrid electric vehicles," in *ICCAD*, 2014.

[14] J.-M. Kang, I. Kolmanovsky, and J. Grizzle, "Dynamic optimization of lean burn engine aftertreatment," *Journal of Dynamic Systems, Measurement, and Control*, 2001.

[15] A.-H. Mohsenian-Rad and A. Leon-Garcia, "Optimal residential load control with price prediction in real-time electricity pricing environments," *Smart Grid, IEEE Transactions on*, vol. 1, no. 2, pp. 120–133, 2010.

[16] D. Linden and T. Reddy, "Handbook of batteries, 2002."

[17] G. L. Plett, "Extended kalman filtering for battery management systems of lipb-based hev battery packs: Part 1. background," *JPS*, 2004.

[18] *High-performance battery monitor IC with coulomb counter, voltage and, temperature measurement*. Texas Instruments.

[19] A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 1998.

[20] *ADVISOR 2003 documentation*. National Renewable Energy Lab.