Optimizing Fuel Economy of Hybrid Electric Vehicles Using a Markov Decision Process Model*

Xue Lin, Yanzhi Wang, Paul Bogdan, Naehyuck Chang, and Massoud Pedram

Abstract— **In contrast to conventional internal combustion engine (ICE) propelled vehicles, hybrid electric vehicles (HEVs) can achieve both higher fuel economy and lower pollutant emissions. The HEV features a hybrid propulsion system consisting of one ICE and one or more electric motors (EMs). The use of both ICE and EM increases the complexity of HEV power management, and so advanced power management policy is required for achieving higher performance and lower fuel consumption. This work aims at minimizing the HEV fuel consumption over any driving cycles, about which no complete information is available to the HEV controller in advance. Therefore, this work proposes to model the HEV power management problem as a Markov decision process (MDP) and derives the optimal power management policy using the policy iteration technique. Simulation results over real-world and testing driving cycles demonstrate that the proposed optimal power management policy improves HEV fuel economy by 23.9% on average compared to the rule-based policy.**

I. Introduction

Automobiles have contributed to the development of modern society. However, large amounts of fuel consumption and pollutant emissions resulting from the increasing number of automobiles have drawn attention of researchers and developers towards more energy efficient and environmentally friendly automobiles. The *hybrid electric vehicle* (HEV) has provided a promising solution towards sustainable mobility. In contrast to conventional *internal combustion engine* (ICE) propelled vehicles, HEVs can achieve both higher fuel economy and lower pollutant emissions [1][2][3].

The HEV features a hybrid propulsion system consisting of an ICE and an *electric motor* (EM), both of which may be coupled directly to the drivetrain. The ICE consumes fuel to provide the primary propulsion, whereas the EM converts the stored electrical energy to the secondary propulsion when extra torque is needed. Besides assisting the ICE with extra torque, the EM also serves as a generator for recovering kinetic energy during braking (known as regenerative braking) and storing excess energy from the ICE during coasting. The introduction of the secondary propulsion by the EM allows for a smaller ICE design and makes HEVs more efficient than conventional ICE vehicles in terms of acceleration, hill climbing, and braking energy utilization [4][5].

On the other hand, the use of both ICE and EM increases the complexity of HEV power management and advanced power management policy is required for achieving higher performance and lower fuel consumption. A power management policy for HEVs determines the power split between the ICE and EM to satisfy the speed and torque requirements and, meanwhile, to ensure safe and smooth operation of the involved power components (e.g., ICE, EM, and batteries). Furthermore, a "good" power management policy should result in reduced fuel consumption and lower pollutant emissions. Rule-based power management approaches have been designed based on heuristics, intuition, and human expertise [6][7]. Although rule-based approaches are effective for real-time supervisory control, they may be far from being optimal. Dynamic programming (DP) techniques have been applied to the power management of various types of HEVs [8][9][10]. DP techniques can derive a globally optimal solution that minimizes the total fuel consumption during a whole driving cycle, which is given as a vehicle speed versus time profile for a specific trip. Unfortunately, the DP techniques require *a priori* knowledge of the driving cycles, and therefore they are not applicable for real-time implementation.

The equivalent consumption minimization strategy (ECMS) approach has been proposed to reduce the global optimization problem (as in DP techniques) to an instantaneous optimization problem [11]. However, the ECMS approach strongly depends on the equivalence factors, which convert the electrical energy consumption of EM into the equivalent fuel consumption of ICE. The equivalence factors are quite sensitive to the driving cycles. In other words, equivalence factors that are suitable for a driving cycle may lead to poor performance for other driving cycles. To overcome this challenge, the adaptive-ECMS (A-ECMS) approach has been applied for HEV power management based on driving cycle prediction within a finite horizon [12]. Although the A-ECMS approach has good performance, the detailed driving cycle prediction method has been omitted. Gong *et al.* has provided a trip modeling method using a combination of geographical information systems (GISs), global positioning systems (GPSs), and intelligent transportation systems (ITSs) [13]. However, the driving cycle constructed by this trip modeling method is synthetic and not accurate enough to capture the real driving scenarios.

Our work aims at minimizing the HEV fuel consumption over any driving cycles. Unlike some previous approaches, we do not assume the complete information about the driving cycles to be available to the HEV controller in advance. Therefore, this work proposes to model the HEV power management problem as a Markov decision process

*This work was supported in part by a grant from the Software and Hardware Foundations program of the National Science Foundation, and by the Center for Integrated Smart Sensors funded by Science, ICT & Future Planning as Global Frontier Project (CISS-N01140838).

X. Lin, Y. Wang, P. Bogdan, and M. Pedram are with the Department of Electrical Engineering, University of Southern California, Los Angeles, CA 90089 USA (e-mail: xuelin@usc.edu, yanzhiwa@usc.edu, pbogdan@usc.edu, pedram@usc.edu).

N. Chang is with the Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701 Korea (e-mail: naehyuck@cad4x.kaist.ac.kr).

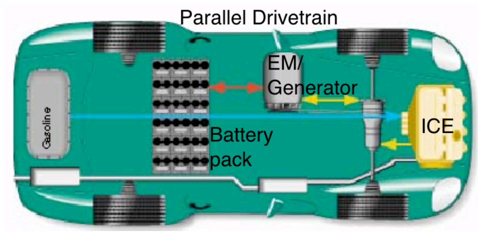

Figure 1. The parallel hybrid drivetrain configuration [14].

(MDP) based on the Markov chain model of driving cycles and the MDP model of the battery pack, which capture the stochastic information in driving cycles, and the *rate capacity* and *recovery effects* of the battery pack, respectively. The optimal power management policy is derived using a policy iteration algorithm, which is a standard dynamic programming-based algorithm for deriving the optimal policy of an MDP. Simulation results over real-world and testing driving cycles demonstrate that the proposed optimal power management policy improves HEV fuel economy by 23.9% on average compared to the rule-based policy.

II. System Modeling

By way of an example and without loss of generality, the proposed power management policy is designed exemplarily for (but not limited to) the parallel hybrid drivetrain configuration displayed in Fig. 1. In an HEV with the parallel hybrid drivetrain (i.e., a parallel HEV), the ICE and EM can deliver power in parallel to drive the wheels. There are five different operation modes in a parallel HEV, depending on the flow of energy: 1) wheels are driven by only the ICE; 2) wheels are driven by only the EM; 3) wheels are driven by both the ICE and EM; 4) battery charging mode: part of the ICE power drives the EM as a generator to charge the battery pack, while the other part of ICE power drives the wheels; 5) regenerative braking mode: the wheels drive the EM as a generator to charge the battery pack when vehicle is braking.

A. Internal Combustion Engine (ICE)

For the study of HEV power management, ICE dynamics is ignored based on the quasi-static assumption [15]. The fuel consumption rate $\dot{m}_f$ (in g/s) of an ICE is a nonlinear function of the ICE speed ω_{ICE} (in rad/s) and torque T_{ICE} (in $\mathrm{N \cdot m}$). The fuel efficiency of an ICE is calculated by

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

where D_f is the fuel energy density (in J/g).

Fig. 2 presents a contour map of the fuel efficiency of an ICE in the *speed-torque* plane. It is a 1.0L VTEC-E SI ICE

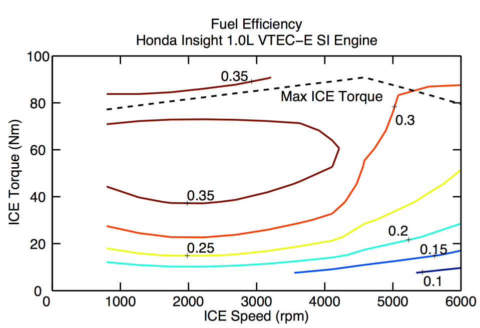

Figure 2. Fuel efficiency map of an ICE.

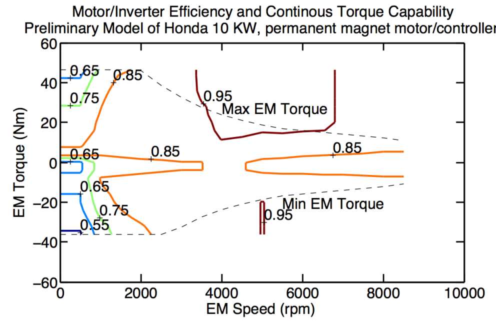

Figure 3. Efficiency map of an EM.

modeled by the advanced vehicle simulator ADVISOR [14]. The ICE has a peak power of 50 kW and a peak efficiency of 0.4. A "good" power management policy should avoid ICE operation point (ω_{ICE}, T_{ICE}) in the low efficiency region. Superimposed on the contour map is the maximum ICE torque $T_{ICE}^{\max}(\omega_{ICE})$ (the dashed line). To ensure safe and smooth operation of an ICE, the following constraints should be satisfied:

$$\omega_{ICE}^{\min} \le \omega_{ICE} \le \omega_{ICE}^{\max}, \quad (2)$$

$$0 \le T_{ICE} \le T_{ICE}^{\max}(\omega_{ICE}). \quad (3)$$

B. Electric Motor (EM)

Fig. 3 presents a contour map of the efficiency of an EM also in the speed-torque plane. It is a permanent magnet EM modeled by ADVISOR. The EM has a peak power of 10 kW and a peak efficiency of 0.96. Let ω_{EM} and T_{EM} respectively denote the speed and torque of the EM. When $T_{EM} \ge 0$, the EM operates as a motor; when $T_{EM} < 0$, the EM operates as a generator. The efficiency of the EM is defined by

$$\eta_{EM}(\omega_{EM}, T_{EM}) = \begin{cases} T_{EM} \cdot \omega_{EM} / P_{batt} & \text{for } T_{EM} \ge 0 \\ P_{batt} / (T_{EM} \cdot \omega_{EM}) & \text{for } T_{EM} < 0 \end{cases} \quad (4)$$

where P_{batt} is the output power of the battery pack. When $T_{EM} \ge 0$, the battery pack is discharging and P_{batt} is a positive value; when $T_{EM} < 0$, the battery pack is charging and P_{batt} is a negative value. Superimposed on the contour map are the maximum and minimum EM torques (the dashed lines) i.e., $T_{EM}^{\max}(\omega_{EM})$ and $T_{EM}^{\min}(\omega_{EM})$, respectively. To ensure safe and smooth operation of an EM, the following constraints should be satisfied:

$$0 \le \omega_{EM} \le \omega_{EM}^{\max}, \quad (5)$$

$$T_{EM}^{\min}(\omega_{EM}) \le T_{EM} \le T_{EM}^{\max}(\omega_{EM}). \quad (6)$$

C. Drivetrain Mechanics

The detailed ADVISOR drivetrain model is not suitable for dynamic optimization due to its high number of states. Thus, we adopt a simplified but sufficiently accurate drivetrain model as in [16][17]. The following equations describe the drivetrain mechanics, showing the mechanical coupling between different components and the vehicle.

- Speed relation

$$\omega_{wh} = \frac{\omega_{ICE}}{R(k)} = \frac{\omega_{EM}}{R(k) \cdot \rho_{reg}}. \quad (7)$$

- Torque relation

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

ω_{wh} and T_{wh} are the wheel speed and torque, respectively. $R(k)$ is the gear ratio of the k-th gear. ρ_{reg} is the reduction gear ratio. η_{reg} and η_{gb} are the reduction gear efficiency and gear box efficiency, respectively. α equals $+1$ if $T_{EM} \ge 0$, and -1 otherwise. β equals $+1$ if $T_{ICE} + \rho_{reg} \cdot T_{EM} \cdot (\eta_{reg})^{\alpha} \geq 0$, and -1 otherwise.

D. Vehicle Dynamics

The vehicle is considered as a rigid body with four wheels and the vehicle mass is assumed to be concentrated in a single point. The following force balance equation describes the vehicle dynamics:

$$m \cdot a = F_{TR} - F_g - F_R - F_{AD}. \quad (9)$$

m is the vehicle mass, a is the vehicle acceleration, and F_{TR} is the total tractive force. The force due to road slope is given by

$$F_g = m \cdot g \cdot \sin\theta, \quad (10)$$

where θ is road slope angle. The rolling friction force is given by

$$F_R = m \cdot g \cdot \cos\theta \cdot C_R, \quad (11)$$

where C_R is rolling friction coefficient. The air drag force is given by

$$F_{AD} = 0.5 \cdot \rho \cdot C_D \cdot A_F \cdot v^2, \quad (12)$$

where ρ is air density, C_D is air drag coefficient, A_F is the vehicle frontal area, and v is the vehicle speed. Given v, a and θ, the total tractive force F_{TR} can be derived using (9)~(12). Then, the wheel speed and torque are related to F_{TR}, v, and wheel radius r_{wh} by

$$\omega_{wh} = v / r_{wh}, \quad (13)$$

$$T_{wh} = F_{TR} \cdot r_{wh}. \quad (14)$$

E. Backward-Looking Optimization

In this work, the backward-looking optimization approach [13] is adopted, which implies that the HEV controller determines the operation of ICE and EM conforming to the mechanical relationships imposed by the drivetrain, so that the vehicle meets the target performance (i.e., target speed and acceleration.) Specifically, from the vehicle speed v, acceleration a (inferred from vehicle speed change), and road slope angle θ, the required wheel speed ω_{wh} and torque T_{wh} can be derived by (9)~(14). Then, the five variables, i.e., the ICE speed ω_{ICE} and torque T_{ICE}, the EM speed ω_{EM} and torque T_{EM}, and the gear ratio $R(k)$, should satisfy (7) and (8), which are essentially three equations. The HEV controller chooses two of the five variables, say, T_{EM} and $R(k)$, as the control variables. The rest of variables (i.e., ω_{ICE}, T_{ICE}, and ω_{EM}) become dependent (associate) variables, the values of which are determined by T_{EM} and $R(k)$ accordingly. The results of the HEV power management policy are the fuel consumption rate of the ICE, and the battery pack output power (associated with the EM.)

III. Markov Decision Process (MDP) Modeling

Markov decision processes (MDPs) provide a powerful mathematical tool for sequential decision making in situations where outcomes are partly random and partly under the control of a decision maker [18]. MDPs have been widely applied to many areas including robotics, automated control, and dynamic power management for embedded systems [19].

A. MDP Concepts and Definitions

We focus on a discrete-time finite-state Markov decision process of a continual process-control task, which is best suited for modeling the HEV power management problem. The whole time horizon is discretized into a sequence of time steps, indexed by $t = 0, 1, 2, 3, \ldots$. There are a finite set $\mathcal{S}$ of states and a finite set $\mathcal{A}$ of actions. At each time step, the process is in some state $s \in \mathcal{S}$, and the decision maker may choose any action $a \in \mathcal{A}_s$, where $\mathcal{A}_s (\subseteq \mathcal{A})$ is the set of actions available for state s. Matrix $\boldsymbol{P}^a$ is the state transition probability matrix, where $\boldsymbol{P}^a_{ss'}$ denotes the probability that action a in state s at time step t will lead to state s' at time step $t+1$. We have $\sum_{s' \in \mathcal{S}} \boldsymbol{P}^a_{ss'} = 1$. Matrix $\boldsymbol{R}^a$ is the immediate reward matrix, where $\boldsymbol{R}^a_{ss'}$ denotes the (expected) immediate reward received after transition to state s' (at time step $t+1$) from state s (at time step t) under action a. Matrices $\boldsymbol{P}^a$ and $\boldsymbol{R}^a$ (for all $a \in \mathcal{A}$) completely specify the most important aspects of the dynamics of an MDP.

A policy, denoted by π, for the decision maker is a mapping from each state $s \in \mathcal{S}$ to an action $a \in \mathcal{A}$ that specifies the action $a = \pi(s)$ that the decision maker will choose when the process is in state s. The MDP optimization problem targets at finding the optimal policy, such that

$$V^{\pi}(s) = E\{\textstyle\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1} \,|\, s_t = s\} \quad (15)$$

is maximized for each state $s \in \mathcal{S}$. The *value function* $V^{\pi}(s)$ is the expected return when the process starts in state s at time step t and follows policy π thereafter. γ is a parameter, $0 < \gamma < 1$, called the *discount rate* that ensures the infinite sum (i.e., $\sum_{k=0}^{\infty} \gamma^k \cdot r_{t+k+1}$) converges to a finite value. r_{t+k+1} is the immediate reward received at time step $t+k+1$, the value of which can be obtained from indexing matrices $\boldsymbol{R}^a_{ss'}$ with $s = s_{t+k}$, $s' = s_{t+k+1}$, and $a = \pi(s_{t+k})$.

For any policy π and any state $s \in \mathcal{S}$, the following consistency condition holds between $V^{\pi}(s)$ and $V^{\pi}(s')$, where s' is a possible successor state of s:

$$V^{\pi}(s) = \textstyle\sum_{s'} \boldsymbol{P}^a_{ss'} \cdot \left(\boldsymbol{R}^a_{ss'} + \gamma \cdot V^{\pi}(s')\right), \quad (16)$$

where $a = \pi(s)$.

B. Stochastic Driving Cycle Modeling

A driving cycle is given as a vehicle speed versus time profile for a specific trip. The HEV controller aims at deriving a power management policy to minimize the fuel consumption during a whole driving cycle. If the HEV controller has *a priori* knowledge of a whole driving cycle at the beginning of a trip, the global optimal policy can be derived using dynamic programming (DP) techniques [8][9][10]. However, such dependency on *a priori* knowledge has become a major deterrent to utilizing the DP approach, i.e., the difficulty of implementation for the real-time control.

We capture the stochastic information in driving cycles using a discrete-time Markov chain model, which predicts the probability distribution of states in the next time step given the state in the current time step. Similar to [20][21], we define the state space for the Markov chain model as a finite number of states, each represented by the power demand and vehicle speed levels:

$$\begin{aligned} \mathcal{S}^D &= \{s^D = [P_{\text{dem}}, v]^{\mathrm{T}} \,|\, P_{\text{dem}} \in \mathcal{P}_{\text{dem}}, v \in \mathcal{V}\} \\ &= \{s^D_1, s^D_2, \ldots, s^D_M\}, \end{aligned} \quad (17)$$

where $P_{\text{dem}} = T_{wh} \cdot \omega_{wh}$ is the power demand for propelling an HEV, $\mathcal{P}_{\text{dem}}$ and $\mathcal{V}$ are respectively the finite sets of power demand levels and vehicle speed levels (discretization is required because our MDP model is suitable for discrete-state spaces), and M is the total number of states in $\mathcal{S}^D$. Then in the state transition probability matrix $\boldsymbol{P}^D$ of the Markov chain, the element $\boldsymbol{P}^D_{i,j}$ denotes the probability that the state in the next time step is s^D_j given that the current state is s^D_i:

$$\boldsymbol{P}^D_{i,j} = \Pr\left(s^D_{t+1} = s^D_j | s^D_t = s^D_i\right). \tag{18}$$

To estimate these state transition probabilities, one needs observation data for both power demand and vehicle speed. We obtain these observations from real-world and testing driving cycle profiles. These profiles provide histories of vehicle speed versus time, and we use the vehicle dynamics to extract corresponding power demand histories as follows:

$$P_{\text{dem}} = F_{TR} \cdot v \tag{19}$$
$$= m\frac{dv}{dt}v + mg\sin\theta\, v + mg\cos\theta\, C_R v + 0.5\rho C_D A_F v^3,$$

which is derived based on Section II-D.

In this work, we use real-world and testing drive cycles representing both highway and city driving, developed by different organizations and projects such as EPA (US Environmental Protection Agency), INRETS (French National Institute for Transport and Safety Research), and MODEM, to compute the observation data and then derive the state transition probability matrix $\boldsymbol{P}^D$ using the maximum likelihood estimation method [22]. In reality, the state transition probability matrix can be updated, when new driving scenarios are observed by the controller.

C. *Stochastic Battery Modeling*

Although the battery pack in an HEV does not couple with the drivetrain directly, it is an important and active power component for an HEV since it provides electrical energy to power the EM and also stores electrical energy generated from the EM (as a generator) during regenerative braking. A comprehensive understanding of the battery model is necessary for deriving power management policy.

The *state* of a battery pack is represented by the amount of charge stored in the battery pack. The majority of literature on HEV power management adopts a simple battery model as follows [3]:

$$Q_t = Q^{ini} - \sum_{k=0}^{t} I_k \cdot \Delta T, \tag{20}$$

where Q_t is the amount of charge stored in the battery pack at the end of time step t, Q^{ini} is the amount of charge stored in the battery pack at the beginning of time step 0, I_t is the discharging current of the battery pack at time step t ($I_t < 0$ means battery charging), and ΔT is the length of a time step. However, this battery model ignores the *rate capacity effect*, which causes the most significant power loss when the battery pack charging/discharging current is high [23]. We know that the battery pack charging/discharging current is high during deceleration and acceleration, and therefore the rate capacity effect should be considered. The rate capacity effect specifies that if the battery pack is discharging ($I > 0$), the actual charge decreasing rate inside the battery pack is higher than I; and if the battery pack is charging ($I < 0$), the actual charge increasing rate inside the battery pack is lower than $|I|$. In addition, the battery model mentioned above also ignores the *recovery effect*, which specifies that the battery pack can partially recover charge loss in previous discharges if relaxation time is allowed in between discharges [23].

A stochastic battery model was proposed in [24]. Inspired by that model, we propose a Markov decision process model for the battery pack in an HEV that captures both the rate capacity effect and the recovery effect. We define the state space for the MDP battery model by discretizing the range of the stored charge of the battery pack i.e., $[Q_{min}, Q_{max}]$ into a finite number of charge levels:

$$\mathcal{S}^{BA} = \{s^{BA}_1, s^{BA}_2, \dots, s^{BA}_N\}, \tag{21}$$

where $Q_{min} = s^{BA}_1 < s^{BA}_2 < \cdots < s^{BA}_N = Q_{max}$. Usually, Q_{min} and Q_{max} are 40% and 80% of the battery pack capacity, respectively, to ensure "healthy" operation of the battery pack [12]. An action a ($\in \mathcal{A}$) taken by the decision maker is to discharge the battery pack with a current value of I, where $I > 0$ denotes discharging, $I < 0$ denotes charging, and $I = 0$ denotes idle. The battery charging/discharging current range $[-I^{\max}, I^{\max}]$ is discretized into a finite number L of values. Then the cardinality of $\mathcal{A}$ equals L.

Next, we need to derive the state transition probability matrix $\boldsymbol{Pb}^a$, where the element $\boldsymbol{Pb}^a_{i,j}$ denotes the probability that action a in state s^{BA}_i at time step t will lead to state s^{BA}_j at time step $t+1$. Specifically, if action a is to set the battery pack in idle ($I = 0$), the battery pack has a probability to switch to a state with higher charge level due to the recovery effect; if action a is to discharge the battery ($I > 0$), the battery pack will switch to some state(s) with lower charge level and the charge decrease is larger than $I \cdot \Delta T$ due to the rate capacity effect; and if action a is to charge the battery pack ($I < 0$), the battery pack will switch to some state(s) with higher charge level and the charge increase is smaller than $|I| \cdot \Delta T$ due to the rate capacity effect. More details are omitted due to space limitation.

D. *Markov Decision Process Modeling of HEV Power Management*

In this section, we propose to model the HEV power management problem as an MDP based on the Markov chain model of driving cycles and the MDP model of the battery pack derived in the previous sections. We define the state space for the MDP model of HEV power management as

$$\mathcal{S} = \left\{s_{i,j} = \left[s^D_i;\ s^{BA}_j\right] \middle|\ s^D_i \in \mathcal{S}^D, s^{BA}_j \in \mathcal{S}^{BA}\right\}. \tag{22}$$

An action a ($\in \mathcal{A}$) taken by the decision maker is to discharge the battery pack with a current value of I, where $I > 0$ denotes discharging, $I < 0$ denotes charging, and $I = 0$ denotes idle. The battery charging/discharging current range $[-I^{\max}, I^{\max}]$ is discretized into a finite number L of values. Then the cardinality of $\mathcal{A}$ equals L.

a) *State Transition Probability Matrix*

Now, we need to derive the state transition probability matrix $\boldsymbol{P}^a$ for the MDP, where the element $\boldsymbol{P}^a_{(i,j),(i',j')}$ denotes the probability that action a in state $s_{i,j}$ at time step t will lead to state $s_{i',j'}$ at time step $t+1$. The value of $\boldsymbol{P}^a_{(i,j),(i',j')}$ can be derived by

$$\boldsymbol{P}^a_{(i,j),(i',j')} = \boldsymbol{P}^D_{i,i'} \cdot \boldsymbol{Pb}^a_{j,j'}, \tag{23}$$

where $\boldsymbol{P}^D$ and $\boldsymbol{Pb}^a$ are the state transition probability matrices of the Markov chain model of driving cycles and the MDP model of the battery pack, respectively. Because these two models are independent from each other, $\boldsymbol{P}^a_{(i,j),(i',j')}$ can be derived by multiplying the corresponding matrix elements in $\boldsymbol{P}^D$ and $\boldsymbol{Pb}^a$ directly.

b) *Immediate Reward Matrix*

Next, we need to derive the immediate reward matrix $\boldsymbol{R}^a$ for the MDP model of HEV power management, where the element $\boldsymbol{R}^a_{(i,j),(i',j')}$ denotes the immediate reward received after taking action a in state $s_{i,j}$. MDP optimization problems aim to maximize the discounted sum of the immediate rewards in the long run as shown in (15). We define the immediate reward as the negative of the fuel consumption in a time step, and therefore by solving the MDP optimization problem we can minimize the overall fuel consumption.

The fuel consumption in a time step depends on state $s_{i,j}$ and action a at the time step. Suppose in state $s_{i,j}$, the power demand is P_{dem} and the vehicle speed is v; and action a specifies that the battery pack discharging current is I. Then, the wheel speed and torque are calculated by

$$\omega_{wh} = v/r_{wh}, \quad (24)$$

$$T_{wh} = P_{\text{dem}} \cdot r_{wh}/v. \quad (25)$$

The battery output power is calculated by

$$P_{batt} = V^{OC} \cdot I - R_{batt} \cdot I^2, \quad (26)$$

where V^{OC} is the open-circuit voltage of the battery pack and R_{batt} is the internal resistance of the battery pack. Please note that if $P_{batt} < 0$, the battery pack is being charged and its input power is $|P_{batt}|$. In order to derive the fuel consumption in the time step, we need to solve the following ***fuel optimization*** (FO) problem:

Given the values of ω_{wh}, T_{wh}, and P_{batt}, ***find*** the EM torque T_{EM} and gear ratio $R(k)$ to ***minimize*** the fuel consumption rate $\dot{m}_f$ ***subject to*** (2)~(8).

Usually, there are about five values that $R(k)$ can assume. For each of the possible $R(k)$ values, we first calculate ω_{ICE} and ω_{EM} using (7), next calculate T_{EM} using (4) while satisfying (5)~(6), and then calculate T_{ICE} using (8) while satisfying (2)~(3). With ω_{ICE} and T_{ICE}, the fuel consumption rate $\dot{m}_f$ is obtained based on the ICE model. We choose the $R(k)$ value that results in the minimum $\dot{m}_f$ i.e., $\dot{m}_f^{opt}$. We calculate $\boldsymbol{R}^a_{(i,j),(i',j')}$ as $-\dot{m}_f^{opt} \cdot \Delta T$. Please note that the immediate reward $\boldsymbol{R}^a_{(i,j),(i',j')}$ is independent of the successor state $s_{i',j'}$. Therefore, the four-dimensional matrix $\boldsymbol{R}^a_{(i,j),(i',j')}$ can be reduced to a two-dimensional matrix $\boldsymbol{R}^a_{(i,j)}$.

E. MDP Optimal Policy Derivation

In Section III-D, we have modeled the HEV power management problem as an MDP with the four essential tuples: the state set $\mathcal{S}$, the action set $\mathcal{A}$, the state transition probability matrix $\boldsymbol{P}^a$, and the immediate reward matrix $\boldsymbol{R}^a$. Now, we will derive the optimal policy π to maximize the value function $V^\pi(s_{i,j})$ for all states $s_{i,j} \in \mathcal{S}$. We adopt the policy iteration algorithm, which is a dynamic programming (DP)-based algorithm to derive the optimal policy for an MDP [18]. The policy iteration algorithm is based on the consistency condition given by (16), which we rewrite as

$$V^\pi(s_{i,j}) = \sum_{(i',j')} \boldsymbol{P}^a_{(i,j),(i',j')} \cdot \left(\boldsymbol{R}^a_{(i,j)} + \gamma \cdot V^\pi(s_{i',j'})\right), \quad (27)$$

where $s_{i',j'}$ is a possible successor state of $s_{i,j}$.

The policy iteration algorithm consists of two basic steps: policy evaluation and policy improvement. The policy evaluation step derives the value function for each state $s_{i,j} \in \mathcal{S}$ with a given policy $\pi(s_{i,j})$ through iteration. The policy improvement step, for each state $s_{i,j} \in \mathcal{S}$, changes the action from the existing policy to a potentially new action that results in a larger value function. If no further improvement can be done, the policy improvement will terminate with the optimal policy; otherwise, the new policy will go through policy evaluation and improvement steps once more.

In summary, the policy iteration is a DP-based algorithm to derive the optimal policy for an MDP. It will result in an optimal HEV power management policy that specifies the action a to take for the HEV, when the HEV is in some state $s_{i,j}$. The action a itself is represented as a discharging/charging current level of the battery pack, and the actual control variables i.e., the EM torque T_{EM} and gear ratio $R(k)$, are obtained through solving the FO problem, for which the solution has been derived in Section III-D-b. The policy iteration algorithm is executed offline, and therefore the execution complexity is not a significant concern.

IV. Experimental Results

The HEV model used for this study is based on Honda Insight Hybrid model developed in ADVISOR [14]. Key parameters are summarized in Table I. We compare our proposed optimal power management policy derived from the MDP model with the rule-based power management policy described in [7].

First, we compare the value functions as defined in (15) of the proposed policy π^{opt} and the rule-based policy π^{rb}. The value function $V^\pi(s)$ of a policy π demonstrates the negative of the expected discounted sum of fuel consumption in the long run, which is to be maximized by π^{opt} and π^{rb}. It is equivalent to minimizing the fuel consumption. We compare $\text{avg}_s V^\pi(s)$ value, which is the average value function over all states, of π^{opt} and π^{rb}. If a discount rate $\gamma = 0.9$ in (15) is used, we obtain $\text{avg}_s V^{\pi^{opt}}(s) = -3.41$ and $\text{avg}_s V^{\pi^{rb}}(s) = -5.53$, which shows the proposed

Policy Iteration Algorithm

1. Initialization
   $V(s_{i,j}) \in \mathcal{R}$ and $\pi(s_{i,j}) \in \mathcal{A}_{s_{i,j}}$ arbitrarily for all $s_{i,j} \in \mathcal{S}$
2. Policy Evaluation
   Repeat
   - $\Delta \leftarrow 0$
   - For each $s_{i,j} \in \mathcal{S}$:
     - $v \leftarrow V(s_{i,j})$
     - $V(s_{i,j}) \leftarrow \sum_{(i',j')} \boldsymbol{P}^{\pi(s_{i,j})}_{(i,j),(i',j')} \cdot \left(\boldsymbol{R}^{\pi(s_{i,j})}_{(i,j)} + \gamma \cdot V(s_{i',j'})\right)$
     - $\Delta \leftarrow \max(\Delta, |v - V(s_{i,j})|)$

   Until $\Delta < \theta$ (a small positive number)
3. Policy Improvement
   $policy\_stable \leftarrow true$
   For each $s_{i,j} \in \mathcal{S}$:
   - $b \leftarrow \pi(s_{i,j})$
   - $\pi(s_{i,j}) \leftarrow \arg\max_a \sum_{(i',j')} \boldsymbol{P}^a_{(i,j),(i',j')} \cdot \left(\boldsymbol{R}^a_{(i,j)} + \gamma \cdot V(s_{i',j'})\right)$
   - If $b \neq \pi(s_{i,j})$, then $policy\_stable \leftarrow false$

   If $policy\_stable$, then stop; else go to 2

TABLE I. HONDA INSIGHT HYBRID COMPONENT PARAMETERS.

Vehicle C_D	0.32	ICE Max. power (kW)	50
Vehicle A_F (m^2)	1.48	ICE Max. Torque (Nm)	89.5
Vehicle r_{wh} (m)	0.3	EM Max. power (kW)	10
Vehicle m (kg)	1000	Battery capacity (Ah)	6.5
Reduction gear ratio ρ_{reg}	1.4	Battery voltage (V)	144

policy achieves 38.3% reduction in fuel consumption. If a discount rate $\gamma = 0.95$ is used, we obtain $\text{avg}_s V^{\pi^{opt}}(s) = -6.63$ and $\text{avg}_s V^{\pi^{rb}}(s) = -10.46$, which shows the proposed policy achieves 36.6% reduction in fuel consumption. Overall, the proposed policy outperforms the rule-based policy in terms of value function.

Next, we test the fuel consumption of the proposed policy and rule-based policy on real-world and testing driving cycles. The fuel consumptions over some driving cycles are summarized in Table II. We can observe that the proposed policy always results in lower fuel consumption and the maximum reduction in fuel consumption is as high as 46.93%. We plot the ICE operation points over a driving cycle on the ICE fuel efficiency map in Fig. 4. The "x" points are from the rule-based policy and the "o" points are from our proposed policy. We can observe that the operation points from the proposed policy are more concentrated on the high efficiency region of the ICE. Furthermore, we compare the overall fuel economy of the proposed policy and the rule-based policy over 17 real-world and testing driving cycles with a total driving time of five hours and both local and highway driving conditions. The rule-based policy achieves a MPG value of 46 and the proposed policy achieves a MPG value of 57, demonstrating the proposed policy improves the fuel economy by 23.9%.

V. CONCLUSION

In contrast to conventional internal combustion engine (ICE) propelled vehicles, hybrid electric vehicles (HEVs) can achieve both higher fuel economy and lower pollutant emissions. The HEV features a hybrid propulsion system consisting of one ICE and one or more electric motors (EMs). The use of both ICE and EM increases the complexity of HEV power management and advanced power management policy is required for achieving higher performance and lower fuel consumption. This work aims at minimizing the HEV fuel consumption over any driving cycles, the complete information of which is not available to the HEV controller in advance. This work proposes to model the HEV power management problem as a Markov decision process (MDP) and derives the optimal power management policy using the policy iteration technique.

TABLE II. FUEL CONSUMPTION: PROPOSED AND RULE-BASED POLICIES.

Driving Cycle	Rule-based Policy	Proposed Policy	Reduction
HWFET	339.94 g	313.57 g	7.75 %
IM240	92.20 g	48.93 g	46.93 %
LA92	585.26 g	353.17 g	39.66 %
NEDC	319.71 g	202.49 g	36.66 %
NYCC	86.06 g	50.94 g	40.81 %
UDDS	363.95 g	204.86 g	43.71 %

REFERENCES

[1] C. C. Chan, "The state of the art of electric, hybrid, and fuel cell vehicles," *Proceedings of the IEEE*, vol. 95, pp. 704-718, Apr. 2007.
[2] F. R. Salmasi, "Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends," *IEEE Trans. Vehicular Technology*, vol. 56, pp. 2393-2404, Sep. 2007.
[3] M. Ehsani, et al., *Modern electric, hybrid electric, and fuel cell vehicles: fundamentals, theory, and design*, CRC press, 2009.
[4] M. Ahman, "Assessing the future competitiveness of alternative powertrains," *Int. J. of Vehicle Design*, pp. 309-331, 2003.
[5] A. Emadi, et al., "Topological overview of hybrid electric and fuel cell vehicular power system architectures and configurations," *IEEE Trans. Vehicular Technology*, vol. 54, pp. 763-770, May 2005.
[6] H. D. Lee, E. S. Koo, S. K. Sul, and J. S. Kim, "Torque control strategy for a parallel-hybrid vehicle using fuzzy logic," *IEEE Industry Applications Magazine*, vol. 6, pp. 33-38, Nov. 2000.
[7] N. J. Schouten, et al., "Fuzzy logic control for parallel hybrid vehicles," *IEEE Trans. Control Systems Technology*, May 2002.
[8] A. Brahma, et al., "Optimal energy management in series hybrid electric vehicles," in *Proc. American Control Conf.*, 2000, pp. 60-64.
[9] C. C. Lin, et al., "Power management strategy for a parallel hybrid electric truck," *IEEE Trans. Control Systems Technology*, Nov. 2003.
[10] L. V. Perez, et al., "Optimization of power management in an hybrid electric vehicle using dynamic programming," *Mathematics and Computers in Simulation*, vol. 73, pp. 244-254, Nov. 2006.
[11] G. Paganelli, et al., "Control development for a hybrid-electric sport-utility vehicle: strategy, implementation and field test results," in *Proc. American Control Conf.*, 2001, pp. 5064-5069.
[12] P. Pisu, and G. Rizzoni, "A comparative study of supervisory control strategies for hybrid electric vehicles," *IEEE Trans. Control Systems Technology*, vol. 15, pp. 506-518, May 2007.
[13] Q. Gong, Y. Li, and Z. R. Peng, "Trip-based optimal power management of plug-in hybrid electric vehicles," *IEEE Trans. Vehicular Technology*, vol. 57, pp. 3393-3401, Nov. 2008.
[14] National Renewable Energy Lab. ADVISOR 2003 documentation. http://bigladdersoftware.com/advisor/docs/advisor_doc.html.
[15] J. M. Kang, et al., "Dynamic optimization of lean burn engine aftertreatment," *J. Dyn. Sys., Meas., Control*, Jun. 2001.
[16] S. Delprat, et al., "Control of a parallel hybrid powertrain: optimal control," *IEEE Trans. Vehicular Technology*, May 2004.
[17] S. Kermanil, et al., "Predictive control for HEV energy management: experimental results," in *Proc. Vehicle Power and Propulsion Conf.*, 2009, pp. 364-369.
[18] R. S. Sutton, and A. G. Barto. *Reinforcement Learning: An Introduction.* The MIT Press, Cambridge, MA, 1998.
[19] L. Benini, et al., "Policy optimization for dynamic power management," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*, vol. 18, pp. 813-833, Jun. 1999.
[20] C. C. Lin, et al., "A stochastic control strategy for hybrid electric vehicles," in *Proc. American Control Conf.*, 2004, pp. 4710-4715.
[21] S. J. Moura, et al., "A stochastic optimal control approach for power management in plug-in hybrid electric vehicles," *IEEE Trans. Control Systems Technology*, vol. 19, pp. 545-555, May 2011.
[22] T. W. Anderson, and L. A. Goodman, "Statistical inference about Markov chains," *Annals Math. Stat.*, vol. 28, pp. 89-110, Mar. 1957.
[23] D. Linden, and T. B. Reddy. *Handbook of Batteries*. McGrew-Hill Professional, 2001.
[24] D. Panigrahi, C. Chiasserini, et al., "Battery life estimation of mobile embedded systems," in *Proc. International Conf. VLSI Design*, 2001.

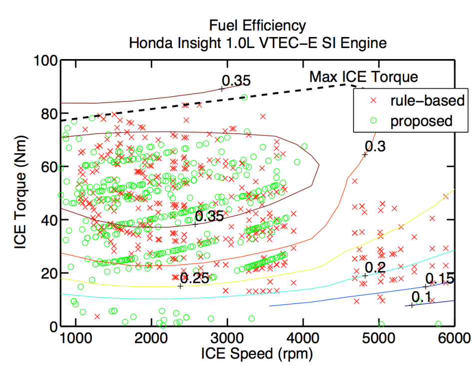

Figure 4. ICE operation points from proposed and rule-based policies.