2015 IEEE 18th International Conference on Intelligent Transportation Systems

Pareto Efficient Policy for Supervisory Power Management Control

Andreas A. Malikopoulos, *Member, IEEE*
Energy & Transportation Science Division
Oak Ridge National Laboratory, Oak Ridge, USA
e-mail: andreas@ornl.gov

***Abstract*—In this paper we address the problem of online optimization of the supervisory power management control in parallel hybrid electric vehicles (HEVs). We model HEV operation as a controlled Markov chain using the long-run expected average cost per unit time criterion, and we show that the control policy yielding the Pareto optimal solution minimizes the average cost criterion online. The effectiveness of the proposed solution is validated through simulation and compared to the solution derived with dynamic programming using the average cost criterion.**

I. Introduction

Hybrid electric vehicles (HEVs) have attracted considerable attention due to their potential to reduce petroleum consumption and greenhouse gas emissions. Implementing online a power management control algorithm to distribute the power demanded by the driver optimally to the available subsystems, e.g., the internal combustion engine, motor, generator, and battery, constitutes a challenging control problem and has been the object of intense study for the last decade. The deterministic formulation of dynamic programming (DP) has been widely used to benchmark the fuel economy of HEVs by providing the maximum theoretical efficiency over a given driving cycle [1]. DP has been extended to a stochastic formulation to derive an optimal control policy for a family of driving cycles. Lin, Peng, and Grizzle [2] proposed a stochastic DP approach using the discounted cost criterion where the one-stage cost was the weighted sum of fuel consumption, NO_x, and particulate matter, with a penalty for state-of-charge (SOC) deviation. The control policy was derived offline by using the policy iteration method. Tate, Grizzle, and Peng [3] used a shortest path stochastic DP formulation to address the minimization of a weighted sum of fuel consumption and tailpipe emissions for an HEV equipped with a dual mode electrically variable transmission, and derived the optimal solution offline by solving a linear program.

Although DP can provide the optimal solution in both the deterministic and stochastic formulation of the power management control problem, the computational burden associated with deriving the optimal control policy prohibits online implementation in vehicles, and it can grow intractable as the size of the problem increases. To address these issues, research efforts have been concentrated on developing online power management algorithms. The main aspects of these algorithms are concerned with the self-sustainability of the electrical path, which must be guaranteed for the entire driving cycle, and the fact that limited a priori knowledge of the future driving conditions is available. Such algorithms consist of an instantaneous optimization problem that accounts for storage system SOC variation through the equivalent fuel consumption (EFC). The latter is evaluated by considering average energy paths leading from the fuel to the energy storage of the electrical path.

Paganelli *et al.* [4] introduced the equivalent consumption minimization strategy (ECMS) that optimizes the power split and the gear ratio while assigning a nonlinear penalty function for SOC deviation in a parallel HEV. Sciarretta, Back, and Guzzella [5] proposed an ECMS algorithm in which the EFC is evaluated under the assumption that every variation in the SOC will be compensated in the future by the engine running at the current operating point. Musardo, Rizzoni, and Staccia [6] presented an adaptive ECMS algorithm that periodically computes the equivalence factor and refreshes the control parameters based on the current driving conditions to maximize fuel economy for a parallel HEV. There has been also a significant amount of work using model predictive control (MPC) to address this problem but mainly in power split HEVs [7] and series HEVs (see [8] and the references therein).

Pisu and Rizzoni [9] compared three algorithms that can be implemented in real time, i.e., a rule-based algorithm, an adaptive ECMS (A-ECMS), and an $\mathcal{H}_\infty$ control, with DP showing that A-ECMS outperforms the other two. Ambuhl and Guzzella [10] presented an ECMS-based algorithm using information received from a global positioning system. Fuel consumption for different driving styles can vary up to 30% [11], [12]. To address different driving styles, Huang, Tan, and He [13] developed a statistical approach to distinguish automatically the driving styles in HEVs. A detailed survey of the supervisory power management control algorithms that have been reported in the literature to date can be found in [14].

In this paper, we address the problem of optimizing the power management control in a parallel HEV configuration.

This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy. The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes.

This research was supported by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for DOE. This support is gratefully acknowledged.

978-1-4673-6596-3 2015

U.S. Government Work Not Protected by U.S. Copyright
DOI 10.1109/ITSC.2015.457
IEEE computer society

We model HEV operation as a controlled Markov chain and solve the stochastic optimal control problem using the long-run expected average cost criterion. We propose a new approach, treating the control problem as a multiobjective optimization problem, and we show that the control policy yielding the Pareto optimal solution for the one-stage cost is an optimal control policy that minimizes the average cost criterion.

The contribution of this paper is the development of an online supervisory controller for a parallel HEV that can yield an optimal solution to the stochastic control problem. The proposed solution uses the efficiency maps of the engine and the motor corresponding to their steady-state operation. Although the supervisory controller in HEVs designates the nominal set points for each subsystem for the lower-level controllers, the implications of the solution in transient operation need further investigation. One potential approach to address this is to learn the transient operation of the system corresponding to the driver' driving style [15] and account for it as discussed in [16]–[18].

The remainder of the paper proceeds as follows. In Section II, we formulate the problem. In Section III, we present the control policy that yields the Pareto optimal solution and show that it minimizes the average cost criterion. In Section IV, we demonstrate the effectiveness of the efficiency of the proposed control algorithm in a parallel HEV and compare it with the DP solution. Finally, in Section V we present concluding remarks.

II. Problem Formulation

In our analysis, we denote random variables with uppercase letters and their space of realizations by script letters. Subscripts denote time, and subscripts in parenthesis denote a subsystem; for example, $X_{t(q)}$ denotes the random variable of the subsystem q at time t. For N subsystems, the shorthand notation $X_{t(1:N)}$ denotes the vector $(X_{t(1)}, X_{t(2)}, ..., X_{t(N)})^T$.

A. Modeling HEV Operation as a Controlled Markov Chain

For this study, we used a parallel HEV with a diesel engine and automatic transmission. In this configuration both the engine and electric motor can provide the power demanded by the driver, either separately or in combination. Because the engine and motor speed depend on the vehicle speed, the available controllable variables are the engine and motor torque. The objective of the power management controller is to guarantee the self-sustainability of the electrical path and distribute the power demanded by the driver optimally between the engine and the motor to maximize HEV efficiency. The controller observes the SOC of the battery, the engine and motor speed, and then computes the optimal engine and motor torque, T^*_{eng} and T^*_{mot} based on the power demanded by the driver, P_{driver}.

We consider the HEV as a system with a finite state space, $\mathcal{S} \subset \mathbb{R}^n$, and a finite control space, $\mathcal{U} \subset \mathbb{R}^m, n, m \in \mathbb{N}$, from which the power management controller selects control actions. In our formulation the state space is the entire range of the engine and motor speed, $\mathcal{S} \subset \mathbb{R}^2$, and the control space $\mathcal{U}$ is the vector of engine and motor torque, $\mathcal{U} \subset \mathbb{R}^2$. The control space can be expanded and include also gear selection depending on the HEV configuration.

In most of the work reported in the literature and discussed in the previous section, the SOC of the battery has been used as a component of the state. However, this may lead to a significantly large state space with implications for increasing the computational burden associated with solving the problem. In our approach, SOC is correlated to an additional power demand by means of one-on-one mapping (Fig. 1). This mapping corresponds to the maximum charging rate of the battery at the current SOC. Depending on the SOC value, there is a corresponding amount of power P_{SOC} that needs to be provided to the battery to stay at the target SOC. This additional amount is added to the driver's power demand. It aims to provide an increasing power request, P_{SOC} as SOC drops up to a certain value designated by the maximum charging rate of the battery as a function of SOC. If SOC is above the target value, then P_{SOC} is zero as no power is required to be added to the battery.

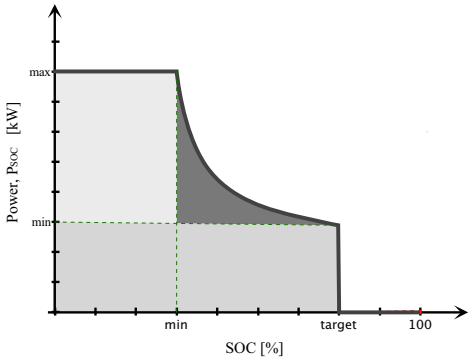

Fig. 1. Power required from the battery with respect to the state of charge (SOC).

The evolution of the state occurs at each of a sequence of stages $t = 0, 1, ...$, and it is portrayed by the sequence of the random variables $X_{t(1:2)} = (X_{t(1)}, X_{t(2)}))^T = (N_{eng}, N_{mot})^T \in \mathcal{S}$ and $U_{t(1:2)} = (U_{t(1)}, U_{t(2)}))^T = (T_{eng}, T_{mot})^T \in \mathcal{U}$, corresponding to the HEV state (engine and motor speed) and control action (engine torque and motor torque) respectively. A state-dependent constraint is incorporated in our problem formulation, i.e., for each state $X_{t(1:2)} = i \in \mathcal{S}$ a nonempty set $\mathcal{C}(i) \subset \mathcal{U}$ of admissible control actions (engine and motor torque) is given. The latter implies that at each state $i \in \mathcal{S}$, the control action set $\mathcal{C}(i) \subset \mathcal{U}$ should include only the control actions that satisfy the physical constraints of the engine and the motor.

Definition 2.1: The set of admissible state/action pairs is defined as

$$\Gamma := \{(X_{t(1:2)}, U_{t(1:2)}) | X_{t(1:2)} = i \in \mathcal{S}, U_{t(1:2)} \in \mathcal{C}(i)\},$$

where Γ is the intersection of a closed subset of $\mathbb{R}^2 \times \mathbb{R}^2$ with the set $\mathcal{S} \times \mathcal{U}$. That is, Γ is closed with respect to the induced topology on $\mathcal{S} \times \mathcal{U}$, and thus it is compact. It follows that for each state $i \in \mathcal{S}$, $\mathcal{C}(i)$ is compact.

Definition 2.2: The function μ is defined that map the state space to the control action space, $\mu : \mathcal{S} \to \mathcal{U}$ such that $\mu(i) \in \mathcal{C}(i), \forall i \in \mathcal{S}$.

Let Π be the set of all sequences $\pi = \{\mu(1), \mu(2), ..., \mu(|\mathcal{S}|)\}$. Each sequence in Π is called a

stationary control policy and operates as follows. Associated with each state $i \in \mathcal{S}$ is the function $\mu(i) \in \mathcal{C}(i)$. If at any time the centralized controller finds the system in state i, then the controller always chooses the action based on the function $\mu(i)$. A stationary policy depends on the history of the process only through the current state, and thus to implement a stationary policy, the controller needs only to know the current state of the system; past states and control actions are irrelevant. The advantages for implementation of a stationary policy are apparent as it uses the storage of less information than required to implement a general policy.

At each stage t, the controller observes the engine and motor speed, $X_{t(1:2)} = i \in \mathcal{S}$, which is a function of the vehicle speed, and executes an action, $U_{t(1:2)} = \mu(X_{t(1:2)})$ (engine and motor torque), from the feasible set of actions, $U_{t(1:2)} \in \mathcal{C}(i)$, at that state. At the same stage t, an uncertainty, $W_{t(1:2)}$, is incorporated in the system consisting of the power demanded by the driver, P_{driver}, and the power required by the battery to reach its target value, P_{SOC}. At the next stage, $t+1$, the system transits to the state $X_{t+1(1:2)} = j \in \mathcal{S}$ and a one-stage cost, $k(X_{t(1:2)}, U_{t(1:2)})$, corresponding to the engine's fuel consumption and motor's efficiency is incurred.

Assumption 2.3: The one-stage expected cost, $k(X_{t(1:2)}, U_{t(1:2)})$, is continuous and bounded.

After the transition to the next state, a new action is selected and the process is repeated. The state transition from one state to another is imposed by a discrete-time equation that describes the dynamics of the system (HEV) of the form

$$X_{t+1(1:2)} = f_t(X_{t(1:2)}, U_{t(1:2)}, W_{t(1:2)}), \tag{1}$$

where $W_{t(1:2)}$ is the disturbance (driver's power demand, P_{driver} and the power required by the battery to reach its target value, P_{SOC}) of the HEV at time t. We have complete observation of the system's state $X_{t(1:2)}$.

Assumption 2.4: The driver's pedal position is a sequence of independent random variables, independent of the initial state $X_{0(1:2)}$.

Assumption 2.4 imposes a condition yielding that the state $X_{t+1(1:2)}$ depends only on $X_{t(1:2)}$ and $U_{t(1:2)}$. Namely, the evolution of the state is a controlled Markov chain and can be represented by a conditional probability, $P(X_{t+1(1:2)} = j | X_{t(1:2)} = i, U_{t(1:2)})$. The completed period of time over which the system is observed is called the *decision-making horizon* and is denoted by T. The horizon can be either finite or infinite; the infinite decision-making horizon is considered for this problem. This is because we are concerned with deriving an optimal control policy, π, that will optimize the efficiency of the HEV in the long-term and not necessarily for a specific period of time. The assumption of an infinite number of stages is never satisfied in practice. However, it is a reasonable approximation for problems involving a finite but very large number of stages, as for example, in the HEV power management control problem where we are interested in optimizing HEV efficiency over the driver's commute.

III. A Multiobjective Optimization Framework for the Solution of the Power Management Control Problem

A. The Average Cost Criterion

Infinite horizon problems are interesting as their analysis is insightful, and the implementation of optimal policies is straightforward. The optimal policies are typically stationary as described in the previous subsection. For the power management control problem formulated here, we select the average cost criterion as we wish to optimize HEV efficiency with respect to any different driver and commute on average per time. Thus we are concerned with deriving a stationary optimal control policy to minimize the long-run average cost per unit time:

$$J(\pi) = \lim_{T\to\infty} \frac{1}{T+1} \mathbb{E}^{\pi} \left[\sum_{0}^{T} k(X_{t(1:2)}, U_{t(1:2)}) \right]. \tag{2}$$

To guarantee that the limit in (2) exists we impose the following assumption.

Assumption 3.1: For each stationary control policy $\pi = \{\mu(1), \mu(2), ..., \mu(|\mathcal{S}|)\}$, where $|\mathcal{S}|$ is the cardinality of the system's state space, the Markov chain $\{X_{t(1:2)} | t = 1, 2, ...\}$ has a single ergodic class.

Namely, for each stationary policy $\pi \in \Pi$ there is a unique probability distribution (row vector) $\beta^{\pi} = (\beta_1, \beta_2, ..., \beta_{|\mathcal{S}|})$, such that $\beta^{\pi} = \beta^{\pi} \cdot \mathbb{P}^{\pi}$, where $\mathbb{P}$ is the transition probability matrix, with $\sum_{i\in\mathcal{S}} \beta_i = 1$. A proof of this assertion may be found in [[19], p. 227]. Under our assumption, it is known [[20], p.175] that

$$\lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} [\mathbb{P}^{\pi}]^t = \mathbf{1} \cdot \beta^{\pi}, \tag{3}$$

where $\mathbf{1} = (1, 1, ..., 1)^T$ is the column vector whose elements are all unity. Substituting (3) into (2) shows that the long-run average cost, $J(\pi)$, does not depend on the initial state and is given more simply as

$$J^{\pi} = \beta^{\pi} \cdot k^{\pi}, \tag{4}$$

where $k^{\pi} = \Big(k(1, \mu(1)), k(2, \mu(2)), ..., k(|\mathcal{S}|, \mu(|\mathcal{S}|)) \Big)^T$ is the column vector of the cost function. Consequently, a stationary control policy is optimal if

$$J^* = \inf \{ J^{\pi} | \pi \in \Pi \}. \tag{5}$$

Since we assume $\mathbb{P}^{\pi}$ to be continuous, it follows that β^{π} is continuous, and since k^{π} is also assumed continuous, so is J^{π}. Hence, by compactness of $\mathcal{U}$, an optimal stationary control policy exists. Our objective is to derive a stationary control policy that minimizes the long-run expected average cost of the HEV. Various methods can be used to solve (5) offline and derive the optimal control policy that minimizes the long-run expected average cost J. In this paper, we seek the theoretical framework that will yield the optimal control policy online while the subsystems interact with each other.

B. HEV Equilibrium Operating Point

In our approach, the power management controller is faced with the task of selecting control actions (engine and motor torque) in several time steps to minimize the average cost. In the HEV configuration adopted here, the engine and the motor are coupled together and their speed is a function of the vehicle speed depending on the gear ratio of the transmission. At each stage t, the controller needs to optimally split the torque demanded by the driver, T_{driver}, between the engine and motor, T^*_{eng} and T^*_{mot} respectively, to optimize fuel economy. Using a myopic approach, namely, selecting the engine to provide a portion of the driver's requested torque so as to operate the engine at a minimum brake specific fuel consumption (BSFC), may result in operating the motor at a lower efficiency, thus wasting energy. Wasting the battery's energy affects fuel economy since this energy will be provided back to the battery from the engine to maintain SOC close to target value.

Consequently, at each stage t, we need to identify an equilibrium operating point, defined as *HEV equilibrium operating point*, among the subsystems, i.e., engine and motor, that will ensure maximization of the overall HEV efficiency. To compute the HEV equilibrium operating point, we formulate a multiobjective decision-making problem consisting of the engine's BSFC, f_{BSFC}, and the motor's efficiency, η_{mot}. Given the engine and motor speed $X_{t(1:2)}$, the objective is to find the optimal control action $U_{t(1:2)}$ (engine and motor torque) that minimizes a multiobjective function reflecting both the engine's fuel consumption and the motor's efficiency. To avoid dominance of one objective function over the other, both functions are normalized with respect to their maximum value. Furthermore, since we formulate a minimization problem, we consider the inverse of the motor efficiency. Using a weighted combination of the different objectives is scale dependent.

The BSFC of the engine is a function of the engine speed N_{eng}, and torque, T_{eng}. Similarly, the efficiency of the motor is a function of the motor speed N_{mot} and torque, T_{mot}. Hence, the normalized BSFC of the engine is $f_1(N_{eng}, T_{eng}) = \frac{f_{BSFC}(N_{eng},T_{eng})}{\| f_{BSFC}\|_\infty}$, and the normalized inverse of the motor's efficiency is $f_2(N_{mot}, T_{mot}) = \frac{\frac{1}{\eta_{mot}(N_{mot},T_{mot})}}{\|\frac{1}{\eta_{mot}}\|_\infty}$.

The multiobjective optimization problem is formulated as

$$\min_{U_t} k(X_{t(1:2)}, U_{t(1:2)}) =$$
$$\min_{U_t} \big(\alpha \cdot f_1(X_{t(1)}, U_{t(1)}) + (1-\alpha)\cdot f_2(X_{t(2)}, U_{t(2)})\big)$$
$$\text{s.t.} \sum_{i=1}^{2} U_{t(i)} = \sum_{i=1}^{2} W_{t(i)} = T_{driver} + T_{SOC}, \tag{6}$$

where α is a scalar that takes values in [0,1], $X_{t(1:2)} = (X_{t(1)}, X_{t(2)}))^T = (N_{eng}, N_{mot})^T \in \mathcal{S}$, $U_{t(1:2)} = (U_{t(1)}, U_{t(2)}))^T = (T_{eng}, T_{mot})^T \in \mathcal{U}$ is the vector of engine and motor torque, and T_{SOC} is the torque corresponding to the power required by the battery, P_{SOC}, to reach its target value. Since P_{SOC} is provided exclusively by the engine, T_{SOC} is computed by dividing P_{SOC} by the engine speed N_{eng}. The multiobjective optimization problem in (6) yields the Pareto efficiency set between the engine and the motor by varying α from 0 to 1 at any given state of the HEV.

C. Pareto Efficient Power Management Control

In a Pareto efficiency allocation among agents, no one can be made better off without making at least one other agent worse. The following is a formal definition.

Definition 3.2 [21]: A solution $u^o \in \mathcal{U}$ is called Pareto optimal if there is no $u \in \mathcal{U}$ such that $k(x,u) \leq k(x,u^o)$. If u^o is Pareto optimal, $k(x,u^o)$ is called efficient. If $u^1, u^2 \in \mathcal{U}$ and $k(x,u^1) < k(x,u^2)$, we say u^1 dominates u^2 and $k(x,u^1)$ dominates $k(x,u^2)$. The set of all Pareto optimal solutions $u^o \in \mathcal{U}$ is the Pareto set, $\mathcal{U}_{Pareto}$. The set of all efficient points $k(x,u^o) \in \mathcal{Y}$ where $u^o \in \mathcal{U}_{Pareto}$, is $\mathcal{Y}_{eff}$ the efficient set.

The question that arises is whether the Pareto efficiency in (6) exists. The following result provides the conditions for its existence.

Proposition 3.3 [21]: Let $\Gamma \in \mathbb{R}^l$ be a nonempty and compact set and each component $k(X_{t(q)}, U_{t(q)}) \colon \Gamma \to \mathbb{R}$ be lower semicontinuous for all $q = 1, ..., N, N \in \mathbb{N}$. Then the Pareto efficiency is not empty.

In our problem, the set of admissible state/action pairs, Γ, is a nonempty compact space (Definition 2.1). Furthermore, the engine's normalized BSFC, $k(X_{t(1)}, U_{t(1)})$, and the inverse of the motor's efficiency, $k(X_{t(2)}, U_{t(2)})$, are both continuous functions. Consequently, the Pareto efficiency exists, and the Pareto optimal solution can yield the HEV equilibrium operating point between the engine and motor.

Definition 3.4 : We define the Pareto control policy π^o the policy that selects a control action which is Pareto optimal.

In the problem considered here, the Pareto control policy is derived as follows. For each state, $i \in \mathcal{S}$, and for any different torque demand, $T_{driver}+T_{SOC}$, ranging from 0 to a maximum value, we solve (6) with α taken values from 0 to 1. The control action, $u^o_{(1:2)} = \mu(i)$, associated with the Pareto control policy is the one that yields the minimum one-stage expected cost in (6) among all values corresponding to different α, namely

$$u^o_{(1:2)} = \operatorname*{argmin}_{U_t} \big\{k_{\alpha_1}(i, u^{\alpha_1}_{(1:2)}), \ldots, k_{\alpha_r}(i, u^{\alpha_r}_{(1:2)})\big\}, r \in \mathbb{N}, \tag{7}$$

where $u^{\alpha_r}_{(1:2)}$ is the solution of (6) when the scalar is α_r, and $k_{\alpha_r}(i, u^{\alpha_r}_{(1:2)})$ is the corresponding minimum one-stage expected cost for the state $i \in \mathcal{S}$ for α_r. Thus, for each state of the HEV and torque demand we derive the Pareto optimal solution that minimizes (6), and store it in a table. If there are multiple solutions, then one of these solutions is selected randomly since all of them will yield the same one-stage expected cost. The Pareto control policy is implemented online using this table as follows. For any combination of vehicle speed, thus engine and motor speed, and torque demand, the Pareto control policy interpolates the control values of the table corresponding to the Pareto optimal solution, $u^o_{(1:2)} = \mu(i)$ that minimizes one-stage expected cost (6).

Theorem 3.5 [22]: The Pareto control policy π^o is the optimal control policy π^* that minimizes the average cost criterion (2).

D. Dynamic Programming Simulation-Based Solution for the Average Cost Criterion

To compare the control policy that yields the Pareto optimal solution with the optimal control policy of DP from Bellman's equation, we need to solve $|\mathcal{S}|$ linear equations, where $|\mathcal{S}|$ is the cardinality of the state space. It is well known that under Assumption 3.1 the minimum average cost J^* has a common value for all initial states, denoted by λ^*, $J^*(i) = \lambda^*, i \in \mathcal{S}$. Moreover, λ^* in conjunction with a differential cost vector $h = (h(1), ..., h(|\mathcal{S}|))$ satisfies Bellman's equation [23]

$$h(i) + \lambda^* = \min_{\pi \in \Pi} \sum_{j=1}^{|\mathcal{S}|} \left[P(j|i, \mu(i)) \cdot h(j) + k(i, \mu(i)) \right]. \quad (8)$$

To solve (8) we need to know the cost function, $k(X_{t(1:2)} = i, U_{t(1:2)} = \mu(X_{t(1:2)}))$ or $k(i, \mu(i))$ for simplicity, and the transition probabilities, $P(X_{t+1(1:2)} = j | X_{t(1:2)} = i, U_{t(1:2)})$, of the HEV, which are not available a priori in our problem as they depend on the driver's power demand. However, we can simulate the HEV model because the state space and control space are known. So at each stage t and for a given state $X_{t(1:2)} = i \in \mathcal{S}$, the controller can select a control action $U_{t(1:2)} = \mu(i)$, and based on the uncertainty, $W_{t(1:2)}$, the system will transit to a new state $X_{t+1(1:2)} = j \in \mathcal{S}$ as imposed by the system's dynamics, (1), and thus generate a corresponding transition cost $k(i, \mu(i))$. It is then possible to use repeated simulation to calculate (at least approximately) the transition probabilities of the system and the expected one-stage costs by averaging and then solve the $|\mathcal{S}|$ linear equations, (8). However for large and complex systems, e.g., HEVs, a more attractive method to derive the optimal control policy is to learn the optimal control policy rather than estimating explicitly the transition probabilities and stage costs using the Q-learning method. This method is analogous to value iteration and has the advantage that it can be used directly in the case of multiple policies. Instead of approximating the cost function of a particular policy, it updates directly the factors associated with an optimal policy, thereby avoiding the multiple policy evaluation steps of he policy iteration method.

It can be seen [24] that the Q-learning method for solving (8) is the following equation

$$Q^{t+1}(i, \mu(i)) = \sum_{j=1}^{|\mathcal{S}|} P(j|i, \mu(i)) \cdot \left(k(i, \mu(i)) + \min_{\mu(j) \in \mathcal{C}(j)} Q^t(j, \mu(j)) \right) - Q^t(i_o, \mu(i_o)), \forall i \in \mathcal{S}, \quad (9)$$

where $i_o \in \mathcal{S}$ is an arbitrary but fixed state. The aim of the Q-learning algorithm is to learn the Q-factors when the transition probabilities, $P(\cdot|\cdot, \cdot)$, are not known but there is access to a simulation device, e.g., simulating an HEV model over a given driving cycle, that can generate them by simulating the system. This can be achieved by simulating the HEV model over a given driving cycle repeatedly until the Q-factors converge. Then the optimal control policy can be extracted by (9). The resulting solution corresponds to the optimal control policy that minimizes the long-run expected average cost criterion [24].

IV. Simulation Results

To validate the effectiveness of the power management controller using the Pareto control policy, we used Autonomie. A vehicle model representing a heavy duty parallel HEV specified by the sponsor was developed in Autonomie and used in this study. The HEV model was simulated over the city-suburban heavy vehicle route (CSHVR). To derive the Pareto control policy, the multiobjective optimization problem (6) was solved offline for different combinations of vehicle speeds, e.g., 0-80 km/h (discretized in 5 km/h), and driver's torque requests, e.g., 0-18, 800 Nm (discretized in 100 Nm). For each vehicle speed-torque combination, the Pareto optimal solution that minimizes (7) was computed and stored in a table. Fig. 2 shows the Pareto efficiency set at 21 km/h when the torque demand is $6,975$ Nm and the transmission gear ratio is 1.5. Fig. 3 illustrates the Pareto set (set of all Pareto optimal solutions), $U^o_{t(1:2)} \in \mathcal{U}_{Pareto}$ corresponding to engine and motor torque (discretized in 100 Nm). In this particular case, the Pareto control policy selects control actions, e.g., $T_{eng} = 310$ Nm and $T_{mot} = 620$ Nm, yielding the normalized engine BSFC, $f_1(N_{eng}, T_{eng}) = 0.3272$, and the normalized inverse of the motor efficiency, $f_2(N_{mot}, T_{mot}) = 0.3095$, since it minimizes (7). The Pareto optimal solution is the HEV equilibrium operating point and can be implemented easily online by interpolating the table.

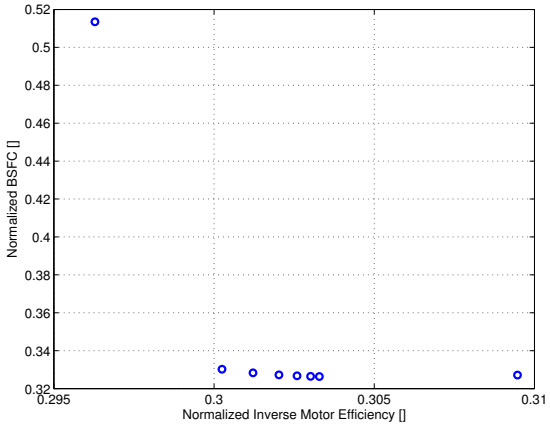

Fig. 2. The Pareto efficiency set, $\mathcal{Y}_{eff}$, corresponding to 21 km/h vehicle speed and 6,975 N·m driver's torque demand.

The Pareto control policy was evaluated over the CSHVR driving cycle and compared to the DP control policy corresponding to the long-run average cost per unit time optimization criterion (8). The DP control policy was derived though simulation by using Q-learning (9). The HEV model was simulated repeatedly over the same driving cycle until the Q-factors in (9) convergence. For the CSHVR driving cycle, Q-learning ran repeatedly 58 times until convergence. The one-on-one correlation, shown in Fig. 1, between SOC and the power added to the driver's power request aimed at maintaining SOC at the target value, i.e., 70% in this case. Both control policies achieved the same cumulative fuel consumption (Table I), which illustrates that the control policy yielding the Pareto optimal solution is an optimal control policy that can be implemented online.

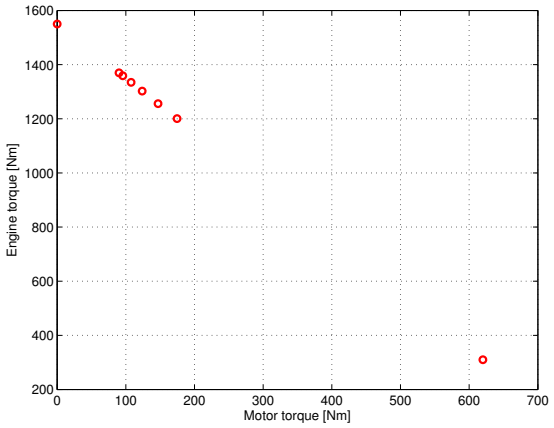

Fig. 3. The Pareto set, $\mathcal{U}_{Pareto}$, corresponding to 21 km/h vehicle speed and 6,975 N·m driver's torque demand.

TABLE I. SIMULATION RESULTS USING DYNAMIC PROGRAMMING AND THE PARETO CONTROL POLICY

Initial State of Charge of the Battery	
Dynamic Programming	70%
Pareto Control Policy	70%
Final State of Charge of the Battery	
Dynamic Programming	70%
Pareto Control Policy	70%
Cumulative Fuel Consumption	
Dynamic Programming	1.24 kg
Pareto Control Policy	1.24 Kg

V. CONCLUDING REMARKS

In the research reported here, we developed the analytical formulation for modeling HEV operation as a controlled Markov chain and presented the solution of the stochastic optimal control problem using the long-run expected average cost criterion. Then we formulated a multiobjective optimization framework and showed that the Pareto control policy minimizes the average cost per unit time criterion.

The effectiveness of the efficiency of the Pareto control policy was validated through simulation of an HEV model for the CSHVR driving cycle, and it was compared to the control policy derived with DP. Both control policies achieved the same cumulative fuel consumption, demonstrating that the Pareto control policy is an optimal control policy that minimizes the long-run expected average cost criterion online.

REFERENCES

[1] 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.

[2] C.-C. Lin, H. Peng, and J. W. Grizzle, "A stochastic control strategy for hybrid electric vehicles," in *Proceedings of the 2004 American Control Conference*, vol. 5, 2004, pp. 4710–4715.

[3] E. D. Tate, J. W. Grizzle, and H. Peng, "SP-SDP for Fuel Consumption and Tailpipe Emissions Minimization in an EVT Hybrid," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 3, pp. 1–16, 2010.

[4] 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 *Proceedings of the 2001 American Control Conference*, vol. 6, 2001, pp. 5064–5069.

[5] A. Sciarretta, M. Back, and L. Guzzella, "Optimal Control of Parallel Hybrid Electric Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 12, no. 3, pp. 352–363, 2004.

[6] C. Musardo, G. Rizzoni, and B. Staccia, "A-ECMS: An Adaptive Algorithm for Hybrid Electric Vehicle Energy Management," in *Proceedings of the 44th IEEE Conference on Decision and Control, and the European Control Conference*, 2005, pp. 1816–1823.

[7] 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.

[8] A. A. Malikopoulos, "Stochastic optimal control for series hybrid electric vehicles," in *Proceedings of the 2013 American Control Conference*, 2013, pp. 1189–1194.

[9] 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.

[10] D. Ambuhl and L. Guzzella, "Predictive Reference Signal Generator for Hybrid Electric Vehicles," *IEEE Transactions on Vehicular Technology*, vol. 58, no. 9, pp. 4730–4740, 2009.

[11] A. A. Malikopoulos and J. P. Aguilar, "Optimization of driving styles for fuel economy improvement," in *2012 15th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, 2012, pp. 194–199.

[12] ——, "An optimization framework for driver feedback systems," *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 2, pp. 955–964, 2013.

[13] X. Huang, Y. Tan, and X. He, "An intelligent multifeature statistical approach for the discrimination of driving conditions of a hybrid electric vehicle," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 2, pp. 453–465, 2011.

[14] A. A. Malikopoulos, "Supervisory Power Management Control Algorithms for Hybrid Electric Vehicles: A Survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 1869–1885, 2014.

[15] A. A. Malikopoulos, P. Y. Papalambros, and D. N. Assanis, "Online identification and stochastic control for autonomous internal combustion engines," *J. Dyn. Sys., Meas., Control*, vol. 132, no. 2, pp. 024 504–9, 2010.

[16] ——, "A state-space representation model and learning algorithm for real-time decision-making under uncertainty," in *ASME International Mechanical Engineering Congress and Exposition, IMECE 2007, November 11, 2007 - November 15, 2007*, vol. 9 PART A. Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI 48109, United States: American Society of Mechanical Engineers, 2008, pp. 575–584.

[17] A. A. Malikopoulos, P. Papalambros, and D. Assanis, "A real-time computational learning model for sequential decision-making problems under uncertainty," *J. Dyn. Syst. Meas. Control*, vol. 4, pp. 0 410 101–0 410 108, 2009.

[18] A. A. Malikopoulos, "Convergence properties of a computational learning model for unknown markov chains," *J. Dyn. Syst. Meas. Control*, vol. 131, no. 4, pp. 0 410 111–0 410 117, 2009.

[19] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. Oxford University Press, August 2001.

[20] S. M. Ross, *Stochastic Processes*, 2nd ed. Wiley, January 1995.

[21] M. Ehrgott, *Multicriteria Optimization*. Springer, 2nd edition, 2005.

[22] A. A. Malikopoulos, "A multiobjective optimization framework for online stochastic optimal control in hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, 2015.

[23] D. P. Bertsekas, "A new value iteration method for the average cost dynamic programming problem," *SIAM Journal on Control and Optimization*, vol. 36, no. 2, pp. 742–759, 1998.

[24] J. Abounadi, D. Bertsekas, and V. S. Borkar, "Learning algorithms for markov decision processes with average cost," *SIAM Journal on Control and Optimization*, vol. 40, no. 3, pp. 681–698, 2001.