A Multiobjective Optimization Framework for Online Stochastic Optimal Control in Hybrid Electric Vehicles

Andreas A. Malikopoulos, *Member, IEEE*

Abstract—The increasing urgency to extract additional efficiency from hybrid propulsion systems has led to the development of advanced power management control algorithms. 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 and show that the control policy yielding the Pareto optimal solution minimizes online the long-run expected average cost per unit time criterion. The effectiveness of the proposed solution is validated through simulation and compared with the solution derived with dynamic programming using the average cost criterion. Both solutions achieved the same cumulative fuel consumption demonstrating that the online Pareto control policy is an optimal control policy.

Index Terms—Hybrid electric vehicles (HEVs), multiobjective optimization, Pareto control policy, power management control, stochastic optimal control.

I. INTRODUCTION

A. Motivation

THE necessity for environmentally friendly vehicles, in conjunction with increasing concerns regarding climate change and U.S. dependency on foreign oil, has led to significant investment in enhancing the propulsion portfolio with new technologies. 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 [1].

B. Related Work

In the late 1990s, Kolmanovsky *et al.* [2] reviewed some emerging approaches at that time for the energy management

of advanced powertrain configurations and presented a case study including a parallel HEV with a turbocharged diesel engine. The objective was to optimize fuel consumption and emissions (e.g., NO_x, CO, HC, and PM). The resulting control policy operated the engine at higher speeds and loads only, where the engine exhibited higher efficiency. Since then, significant research efforts have focused on optimizing the power management control in parallel HEVs. He and Hodgson [3], [4] presented one of the first models for simulation of parallel HEVs with a specific rule-based control strategy aimed at increasing battery state-of-charge (SOC) recovery.

A significant amount of work has been proposed on optimizing the power management control in parallel HEVs using the deterministic formulation of dynamic programming (DP), thus deriving an optimal control policy for a given driving cycle. Lin *et al.* [5] used DP to compute the optimal control policy in a parallel HEV to minimize fuel consumption and selected emission species over a given driving cycle. The derived control policy was implemented online through the power split ratio $\mathrm{PSR} = P_{\mathrm{eng}}/P_{\mathrm{req}}$, where P_{eng} is the engine power and P_{req} is the power demanded from the driver. Four operating modes were defined: 1) motor-only ($\mathrm{PSR} = 0$); 2) engine-only ($\mathrm{PSR} = 1$); 3) power-assist ($0 < \mathrm{PSR} < 1$); and 4) recharging modes ($\mathrm{PSR} > 1$).

The deterministic formulation of DP has been used to benchmark the fuel economy of HEVs by providing the maximum theoretical efficiency over a given driving cycle. DP has been extended to the stochastic problem formulation by considering a family of driving cycles. Lin *et al.* [6] proposed a stochastic DP (SDP) approach using the discounted cost criterion where the one-stage cost was the weighted sum of fuel consumption, NO_x, and particulate matter emissions, with a penalty for SOC deviation. The optimal control policy was derived offline using the policy iteration method for seven different driving cycles. It was shown that SDP achieves better performance most of the time than the rule-based implementation of the control policy derived using DP for each particular driving cycle.

The first attempt to use the shortest path formulation of the power management control problem using SDP (SP-SDP) was in [7]. The method was illustrated on a parallel HEV truck model, and it was shown that there are two advantages of the SP-SDP compared with the discounted cost criterion.

1) A single tuning parameter is needed to trade off fuel economy and emissions versus battery SOC deviation,

as compared with two parameters in the discounted cost criterion.

2) The policy derived using SP-SDP demonstrates better fuel and emission minimization while also achieving better SOC control when the vehicle is turned OFF.

Opila *et al.* [8] presented a method to account for drivability metrics in their proposed power management control algorithm also using the SP-SDP formulation. It was shown that this method yields up to 10% fuel economy improvement on a representative parallel HEV when compared with a simpler instantaneous optimization formulation. Subsequently, Tate *et al.* [9] used SP-SDP to address minimization of a weighted sum of fuel consumption and tailpipe emissions for an HEV equipped with a dual-mode electrically variable transmission. The unique aspects of this paper included an electrically variable transmission and catalytic converter and a state-censoring technique to achieve short computation time. The optimal solution was derived offline by solving a linear program demonstrating more than 50% reduced tailpipe emissions compared with a baseline controller.

Although DP can provide the optimal solution in both the deterministic and stochastic formulations of the power management control problem, the computational burden associated with deriving the optimal control policy prohibits online derivation in vehicles. To address these issues, research efforts have been concentrated on developing online algorithms. 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.* [10] 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 *et al.* [11] proposed an ECMS algorithm in which EFC is evaluated under the assumption that every variation in SOC will be compensated in the future by the engine running at the current operating point. The simulation results illustrated that the proposed algorithm can keep deviations of SOC from the target value at a low level. Musardo *et al.* [12] presented an adaptive ECMS (A-ECMS) algorithm that periodically computes the equivalence factor and refreshes the control parameters based on the current driving conditions to maximize fuel economy in a parallel HEV. Pisu and Rizzoni [13] compared three algorithms that can be implemented online: a rule-based algorithm, an A-ECMS, and an $\mathcal{H}_\infty$ control. The simulation results showed that A-ECMS promises superior robustness and drivability, while it achieves better fuel economy results compared with the rule-based and $\mathcal{H}_\infty$ control algorithms.

There has also been a significant amount of work using model predictive control (MPC) to address this problem but mainly in power split HEVs [14] and series HEVs (see [15], and the references therein). Other recent efforts have focused on incorporating external information, e.g., destination route. Johannesson *et al.* [16] introduced a control algorithm enhanced with information supplied by the vehicle navigation system. Ambuhl and Guzzella [17] presented an

ECMS-based algorithm using information received from a global positioning system. To address variation in fuel consumption for different driving styles [18], [19], Huang *et al.* [20] recently developed a statistical approach to distinguish automatically driving styles in HEVs.

C. Contribution of This Paper

Although previous research reported in the literature has aimed at enhancing our understanding of power management control optimization in parallel HEVs, deriving online an optimal solution for different driving styles still remains a challenging control problem. This paper has two main objectives:

1) to provide a rigorous model of the power management control problem within a stochastic formulation;
2) to develop the theoretical framework that can yield an optimal solution online for any given driving style that minimizes the long-run expected average cost criterion.

A preliminary effort was reported in [21] where the potential of implementing a Pareto control policy was investigated.

The contributions of this paper are as follows:

1) the analytical formulation for modeling HEV operation as a controlled Markov chain;
2) the development of a multiobjective optimization framework that can be used to derive the optimal control policy;
3) the implementation of the Pareto control policy that minimizes the long-run expected average cost criterion and the conditions under which this policy exists.

D. Comparison With Related Work

We now discuss how the proposed solution is different from the other solutions reported in the literature. Some of the previous work has been focused on deriving an optimal control policy using DP for a given vehicle speed profile [5], [22] with respect to the total cost criterion. Then, the control policy is implemented online to account for other vehicle speed profiles through rules resulting in suboptimal performance. To address the latter, the stochastic formulation of DP is used to derive the optimal control policy offline using different optimality criteria. The discounted cost criterion and policy evaluation were used in [6] and linear programming in [9]. In all the cases, the control policy is derived offline and is optimal either for specific driving cycles or a family of them.

Some online approaches [10]–[13], on the other hand, address this problem through an instantaneous optimization problem. 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. These algorithms require some *a priori* knowledge of the vehicle speed profile. Using MPC, an optimization problem is formulated to derive the optimal control policy for a receding horizon. Wang and Boyd [23] determined that the shortcoming of MPC is that it can only be used in applications with slow dynamics, where the sample time is measured in seconds or minutes and they described a collection of methods for improving the speed of MPC using online optimization.

The proposed solution in this paper is different than the approaches above as follows. The evolution of the HEV state, rather than the driver's power demand, is shown to be a controlled Markov chain. Thus, the driver is considered as an unknown disturbance to the system, i.e., the future driver behavior is unknown, and it is assumed that the pedal position, e.g., acceleration or brake, is a sequence of independent random variables, which takes values in a given finite set as discussed in Section III-A. Based on this assumption, it can be shown [24, Lemma 6.6, p. 18] that the evolution of the system is a controlled Markov chain. This allows us to describe the evolution of the state of the system (HEV) by means of the one-step transition probability, and thus we can perform the following:

1) use standard results for the analysis of the long-run average cost criterion (see [25, p. 75]);
2) use DP to solve the stochastic control problem and compare it with our solution.

The solution of the proposed multiobjective optimization framework reveals an equilibrium operating point among the subsystems for all different values of the disturbance, i.e., for any driver's driving style, which is Pareto efficient (Lemma 11). If all the subsystems operate at this equilibrium point for each realization of the HEV state, then the long-run expected average cost is minimized (Theorem 12). The Pareto control policy that reveals the equilibrium operating point for each subsystem can be implemented online, and it is an optimal policy with respect to the average cost criterion for any different driver.

There are still open issues, however, with practical implications. First, the proposed solution optimizes the efficiency for any driver using the long-run expected average cost criterion. Namely, being able to derive the optimal control policy online for a specific trip (e.g., total cost criterion from point A to point B) still remains an open issue. Second, 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 and account for it as discussed in [26]–[28].

Multiobjective optimization has been used in the literature to address control problems with conflicting objectives. Logist *et al.* [29] proposed a generic solution strategy for multiple objective mixed-integer optimal control problems. To tackle the multiple objective functions, a scalarization method, similar to the one used here, was exploited to transform the original optimization problem into a series of parametric single objectives. Logist *et al.* [30] investigated efficient multiple objective strategies with fast deterministic approaches for dynamic optimization and exploited techniques aimed at efficiently and accurately generating the Pareto efficiency sets. The Pareto efficiency set in the proposed solution here is generated offline and stored in a lookup table.

E. Organization of This Paper

The remainder of this paper proceeds as follows. In Section II, we introduce our notation and formulate the problem. In Section III, we present the analytical offline solution to the stochastic optimal control problem, develop the multiobjective optimization framework, and introduce the Pareto control policy that minimizes the average cost criterion. In Section IV, we present the DP simulation-based solution for the average cost criterion and compare it with the solution of the Pareto control policy in a parallel HEV for different driving cycles. Finally, in Section V, we present concluding remarks.

II. PROBLEM FORMULATION

A. Notation

In our analysis, we denote random variables with upper-case letters and their space of realizations by script letters. Subscripts denote time, and subscripts in parentheses 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 $(X_{t(1)}, X_{t(2)}, \ldots, X_{t(N)})^T$.

$\mathbb{P}(\cdot)$ is the transition probability matrix, and $\mathbb{E}[\cdot]$ is the corresponding expectation of a random variable. For a control policy π, we use $\mathbb{P}^\pi(\cdot)$, $\mathbb{E}^\pi[\cdot]$, and β^π to denote that the transition probability matrix, expectation, and stationary distribution depend on the choice of the control policy π.

B. Power Management Control Problem

For this paper, we used a parallel HEV with a diesel engine and automatic transmission. The electric machine (motor/generator) is coupled to the output shaft of the engine through a clutch and gear ratio before the transmission (pretransmission configuration). 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}.

C. Modeling HEV Operation as a Controlled Markov Chain

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, $\mathcal{S} \subset \mathbb{R}^2$, where the engine and motor speed progress in a compact subset of $\mathbb{R}$. The control space $\mathcal{U}$ is the vector of engine and motor torque, $\mathcal{U} \subset \mathbb{R}^2$; however, it can be expanded to also include gear selection, depending on the HEV configuration.
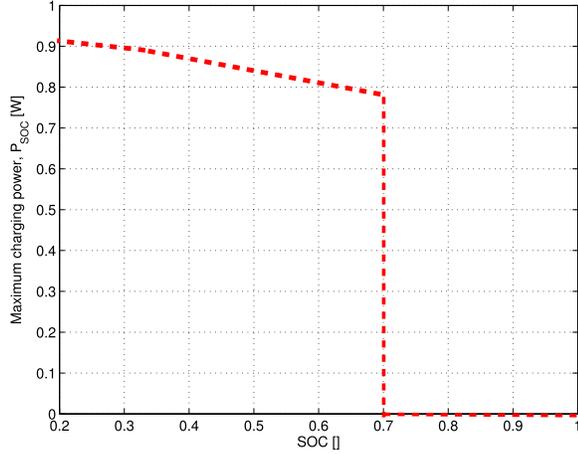
Fig. 1. One-on-one mapping indicating the maximum allowable charging power P_{SOC} to be provided to the battery until it reaches the target SOC (0.7 in this case).
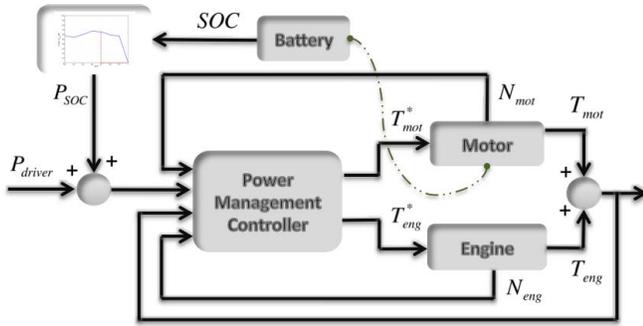
Fig. 2. Control scheme for the power management control of the HEV.

In [5]–[9] and [11] as 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. This mapping corresponds to the maximum allowable charging power (Fig. 1) of the battery with respect to the SOC as designated by the battery specifications. Thus, based on the current SOC, the mapping indicates the maximum allowable charging power P_{SOC} to be provided to the battery until it reaches the target SOC. If the current SOC, however, is above the target value (0.7 in this case), then P_{SOC} is assigned to be equal to zero. P_{SOC} is added to the driver's power demand P_{driver} (only when $P_{driver} > 0$) as shown in Fig. 2. However, the maximum allowable charging power will change over the lifecycle of the battery, which will have an impact on maintaining the SOC close to its target value. One potential approach to address this issue is to use related data from the manufacturer that can yield the change of the maximum allowable charging power with respect to the lifecycle of the battery and account for it.

The evolution of the state occurs at each of a sequence of stages $t = 0, 1, \ldots$, 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 1: The set of admissible state/action pairs is defined as

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

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

Let Π be the set of all the sequences $\pi = \{\mu(1), \mu(2), \ldots, \mu(|\mathcal{S}|)\}$. 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 power management 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 the 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 as designated by the pedal position, e.g., accelerator or brake. 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 expected cost $k(X_{t(1:2)}, U_{t(1:2)})$ is incurred corresponding to the engine's fuel consumption and motor's efficiency.

Assumption 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 pedal position) of the HEV at time t.

When we drive our vehicle, we press the accelerator or brake pedal at each time t based on what we wish at time t, which is conditionally independent on what we desired in the past, and it depends on what we encounter in the traffic at time t [31]. For example, we might have to accelerate at time t to pass a
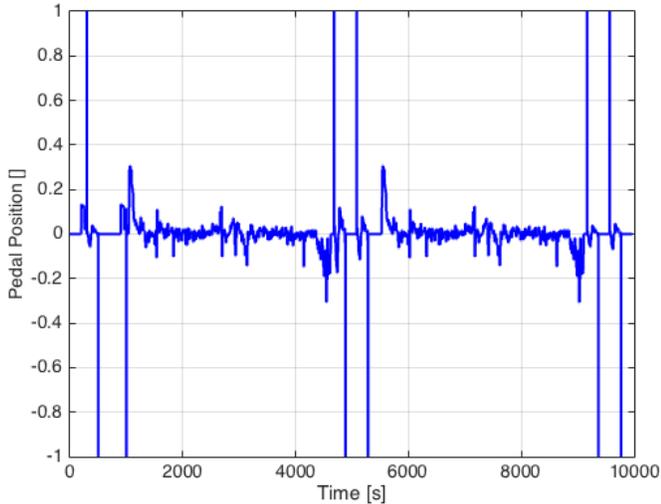
Fig. 3.   Driver's pedal position corresponding to different driving cycles.

Fig. 4.   Autocorrelation plot for the driver's pedal position.

vehicle or to brake to avoid collision, which are independent events of what we encountered at time $t-1$. Although this is a rather intuitive consideration, to justify it further, we use autocorrelation plots for the driver's pedal position, e.g., accelerator or brake. Autocorrelation plots [32, pp. 28–32] are widely used for evaluating randomness in a data set. This randomness is ascertained by computing autocorrelations for data values at different time lags. If the data are a sequence of independent random values, such autocorrelations should be near zero for any and all time-lag separations. If nonrandom, then one or more of the autocorrelations will be significantly nonzero.

To compute the autocorrelation plot for the driver's pedal position, we used various standard dynamometer driving schedules (DDSs) (or simply driving cycles), which are vehicle speed profiles established by the U.S. Environmental Protection Agency (EPA) for testing and measuring fuel economy and emissions. These driving cycles essentially represent situations in which the driver requests a particular vehicle speed profile deemed characteristic of her/his driving style. The pedal position shown in Fig. 3 corresponds to a combination of different portions of the federal test procedure (FTP), the US06, and the urban DDS (UDDS) driving cycles representing a typical urban and highway commute. Negatives values of the pedal position indicate brake, and positive values indicate acceleration. The autocorrelation plots corresponding to this pedal position are almost zero for all time-lag separations (Figs. 4 and 5 in zoomed-in view).

The 95% confidence limits in the autocorrelation plots are equal to $-1/N \pm 2/\sqrt{N}$ [32], where N is the length of the series, i.e., pedal position. For the pedal position shown in Fig. 3, the confidence limits are equal to ± 0.02. Thus, the sequence of the driver's pedal position is a random series. This observation leads to our next assumption.

Assumption 4: The driver's pedal position is a sequence of independent random variables, independent of the initial state $X_{0(1:2)}$.

Assumption 4 imposes a condition yielding that the state $X_{t+1(1:2)}$ depends only on $X_{t(1:2)}$ and $U_{t(1:2)}$ [24, Lemma 6.6, p. 18]. Namely, the evolution of the state can
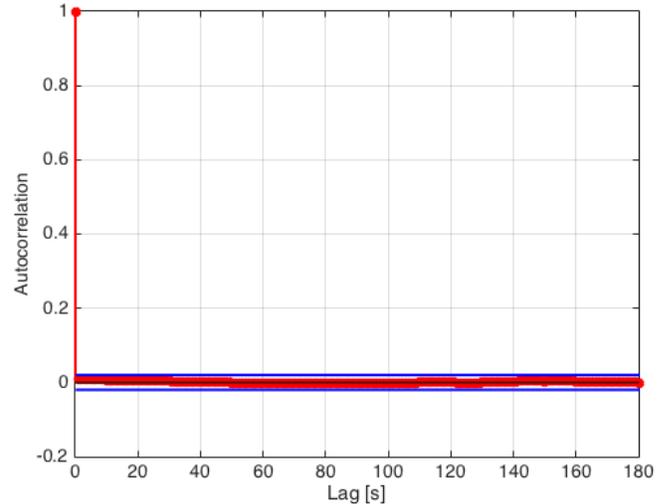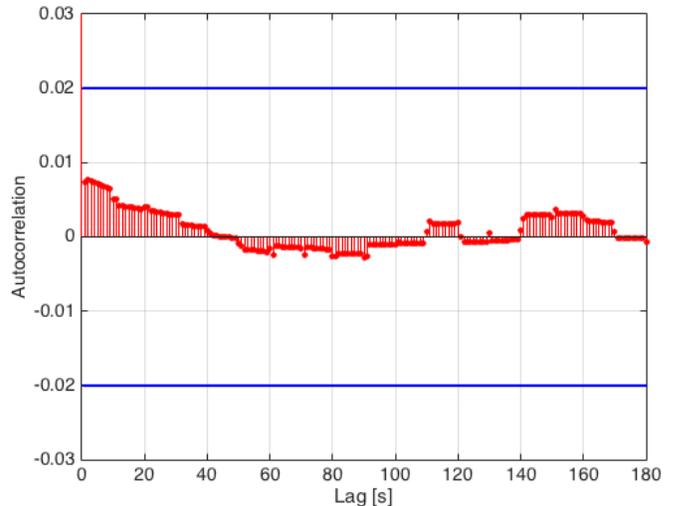
Fig. 5.   Autocorrelation plot for the driver's pedal position (zoomed-in view).

be modeled as a controlled Markov chain and 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 [33], as for example, in the HEV power management control problem where we are interested in optimizing HEV efficiency over the driver's commute.

III. MULTIOBJECTIVE OPTIMIZATION FRAMEWORK FOR THE SOLUTION TO THE POWER MANAGEMENT CONTROL PROBLEM

The power management controller is faced with the task of selecting control actions (engine and motor torque) in several

time steps to minimize the long-run expected average cost per unit time. In our approach, at each stage t, we seek to identify an equilibrium operating point, defined as *HEV equilibrium operating point*, among the subsystems, i.e., engine and motor, which will minimize the average cost criterion online.

A. 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 section. However, these problems require a more sophisticated analysis than the finite horizon problems because we need to analyze limiting behavior as the horizon tends to infinity. For the power management control problem formulated here, we select the average cost criterion as we wish to optimize HEV efficiency (minimize losses) with respect to any different driver and commute on average. Thus, we are concerned with deriving a stationary optimal control policy to minimize the long-run expected average cost per unit time

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

where $k(X_{t(1:2)}, U_{t(1:2)})$ is the one-stage cost of HEV. To guarantee that the limit in (2) exists, we impose the following assumption.

Assumption 5: For each stationary control policy $\pi = \{\mu(1), \mu(2), \ldots, \mu(|\mathcal{S}|)\}$, where $|\mathcal{S}|$ is the cardinality of the system's state space, the Markov chain $\{X_{t(1:2)}|t = 1, 2, \ldots\}$ has a single ergodic class.

Namely, for each stationary policy $\pi \in \Pi$, there is a unique probability distribution (row vector) $\beta^\pi = (\beta_1, \beta_2, \ldots, \beta_{|\mathcal{S}|})$ 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 [34, p. 227]. Under our assumption, it is known [35, p. 175] that

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

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

$$J^\pi = \beta^\pi \cdot k^\pi \quad (4)$$

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

$$J^* = \inf\{J^\pi | \pi \in \Pi\}. \quad (5)$$

Since we assume $\mathbb{P}^\pi$ to be continuous, it follows that β^π is continuous, and since k^π is also assumed continuous (Assumption 3), 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. The next result yields the solution to (5).

Theorem 6 [36]: Suppose that the Markov chain has a single ergodic class (Assumption 5) and that the column vector of the entire system's one-stage expected cost k belongs to the set of all bounded, continuous, and real-valued functions on $\mathcal{S}$ (Assumption 3). If $\pi \in \Pi$ is a control policy, then (C, J) is the solution to the following:

$$C + \mathbf{1} \cdot J = \mathbb{P}^\pi \cdot C + k^\pi \quad (6)$$

where $C \in \mathbb{R}^{|\mathcal{S}|}$ is a column vector of real numbers, J is the average cost, and $|\mathcal{S}|$ is the cardinality of the system's state space.

The following theorem yields the analytical solution for the minimum average cost J^* (Bellman equation).

Theorem 7 [37]: Let $\pi \in \Pi$. If there exist (C, J^*) such that

$$C + \mathbf{1} \cdot J^* = \min_{\pi\in\Pi} [\mathbb{P}^\pi \cdot C + k^\pi] \quad (7)$$

then π is the optimal control policy.

Equation (7) consists of $|\mathcal{S}|$ linear equations. Various methods can be used to solve (7) 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 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 the HEV efficiency. Using a myopic approach, namely, operating 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 the target value.

Consequently, at each stage t, we seek to identify an equilibrium operating point [38] among the subsystems, i.e., engine and motor, which 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.

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 $f_1(N_{\text{eng}}, T_{\text{eng}}) = (f_{\text{BSFC}}(N_{\text{eng}}, T_{\text{eng}})/\|f_{\text{BSFC}}\|_\infty)$, and the normalized inverse of the motor's efficiency is $f_2(N_{\text{mot}}, T_{\text{mot}}) = ((1/(\eta_{\text{mot}}(N_{\text{mot}}, T_{\text{mot}})))/(\|(1/\eta_{\text{mot}})\|_\infty))$.

The multiobjective optimization problem is formulated as

$$\min_{U_t} k(X_{t(1:2)}, U_{t(1:2)}) = \min_{U_t}(\alpha \cdot f_1(X_{t(1)}, U_{t(1)})$$
$$+ (1 - \alpha) \cdot f_2(X_{t(2)}, U_{t(2)}))$$
$$\text{s.t.} \sum_{i=1}^{2} U_{t(i)} = T_{\text{driver}} + T_{\text{SOC}} \qquad (8)$$

where α is a scalar that takes values in $[0, 1]$, $X_{t(1:2)} = (X_{t(1)}, X_{t(2)}))^T = (N_{\text{eng}}, N_{\text{mot}})^T \in \mathcal{S}$, $U_{t(1:2)} = (U_{t(1)}, U_{t(2)}))^T = (T_{\text{eng}}, T_{\text{mot}})^T \in \mathcal{U}$ 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 exclusively provided by the engine, T_{SOC} is computed by dividing P_{SOC} by the engine speed N_{eng}. The multiobjective optimization problem in (8) 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 8 [39]: 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 Pareto 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 $\mathcal{U}_{\text{Pareto}}$. The set of all efficient points $k(x, u^o) \in \mathcal{Y}$ where $u^o \in \mathcal{U}_{\text{Pareto}}$ is $\mathcal{Y}_{\text{eff}}$ the Pareto efficient set.

The question that arises is under what conditions the Pareto efficient set in (8) exists. The following result provides the conditions for its existence.

Proposition 9 [39]: Let $\Gamma \in \mathbb{R}^l$ be a nonempty and compact set and each component $k(X_{t(q)}, U_{t(q)})$: $\Gamma \to \mathbb{R}$ be lower semicontinuous for all $q = 1, \ldots, N$, $N \in \mathbb{N}$. Then, the Pareto efficient set is not empty.

In our problem, the set of admissible state/action pairs Γ is a nonempty compact space (Definition 1). Furthermore, the engine's normalized BSFC $f_1(X_{t(1)}, U_{t(1)})$ and the inverse of the motor's efficiency $f_2(X_{t(2)}, U_{t(2)})$ are both continuous functions. Consequently, the Pareto efficient set exists, and the Pareto optimal solution can yield the HEV equilibrium operating point between the engine and the motor.

Definition 10: The Pareto control policy π^o is defined as the policy that yields the Pareto efficient one-stage expected cost for each subsystem at each state $i \in \mathcal{S}$ of the system.

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 $T_{\text{driver}} + T_{\text{SOC}}$, we solve (8) 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 (8) among all values corresponding

to different α, namely

$$u^o_{(1:2)} = \underset{U_t}{\arg\min} \left\{ k_{\alpha_1}\left(i, u^{\alpha_1}_{(1:2)}\right), \ldots, k_{\alpha_r}\left(i, u^{\alpha_r}_{(1:2)}\right) \right\}, \; r \in \mathbb{N} \quad (9)$$

where $u^{\alpha_r}_{(1:2)}$ is the solution to (8) 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 (8) and store it in a table. If there are multiple solutions, then one of these solutions is randomly selected 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 (8).

D. Connection Between the Pareto Optimal Solution and the Average Cost Criterion

In this section, we show that the Pareto control policy is the optimal control policy that minimizes the average cost criterion (2).

Lemma 11: The Pareto control policy π^o minimizes the one-stage expected cost in (8).

Proof: For any state at time t, $X_{t(1:2)} = i \in \mathcal{S}$, and let $k^{\pi'}(i, u_{(1:2)})$ and $k^{\pi^o}(i, u^o_{(1:2)})$ be the one-stage expected costs at this state corresponding to any control policy π' and the Pareto control policy π^o, respectively. By Definition 10, at each state, the Pareto control policy π^o yields the Pareto optimal solution. By Definition 8 of the Pareto optimal solution, there is no $u_{(1:2)} \in \mathcal{U}$ such that $k^{\pi'}(i, u_{(1:2)}) \leq k^{\pi^o}(i, u^o_{(1:2)})$ for all $\pi' \in \Pi$. Thus, at each realization of the random variable $X_{t(1:2)}$, the Pareto control policy π^o minimizes the one-stage expected cost in (8). $\qquad\square$

Theorem 12: The Pareto control policy π^o is the optimal control policy π^* that minimizes the average cost criterion (2).

Proof: Let π^o be the Pareto control policy. From Lemma 11, we have that for each realization of the state $X_{t(1:N)} = i$, $k^{\pi^o}(i, U_{t(1:2)}) \leq k^{\pi'}(i, U_{t(1:2)})$ for any control policy $\pi' \in \Pi$. Since the system's one-stage cost is bounded (Assumption 3), taking the expected average sum from $t = 0$ up to a finite time $T \in \mathbb{N}$ is well defined and finite. Thus

$$\frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{t=0}^{T} k^{\pi^o}(X_{t(1:2)}, U_{t(1:2)}) \right]$$
$$\leq \frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{t=1}^{T} k^{\pi'}(X_{t(1:2)}, U_{t(1:2)}) \right]. \quad (10)$$

Taking the lim inf as T goes to infinity

$$\liminf_{T \to \infty} \frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{t=0}^{T} k^{\pi^o}(X_{t(1:2)}, U_{t(1:2)}) \right]$$
$$\leq \liminf_{T \to \infty} \frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{t=1}^{T} k^{\pi'}(X_{t(1:2)}, U_{t(1:2)}) \right]. \quad (11)$$

Since each stationary control policy has a single ergodic class (Assumption 5), the limit in (11) is well defined; hence

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

$$\leq J^{\pi'} = \lim_{T \to \infty} \frac{1}{T+1} \mathbb{E}^\pi \left[\sum_{t=1}^T k^{\pi'}(X_{t(1:2)}, U_{t(1:2)}) \right] \quad \forall \pi' \in \Pi.$$

$$(12)$$

□

IV. SUPERVISORY POWER MANAGEMENT CONTROL USING THE PARETO EFFICIENT SOLUTION

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

To compare the Pareto control policy with the optimal control policy of DP from Bellman's equation (7), we need to solve $|\mathcal{S}|$ linear equations, where $|\mathcal{S}|$ is the cardinality of the state space.

Under Assumption 5, the minimum average cost J^* has a common value for all the initial states [40], denoted by λ^*, and $J^*(i) = \lambda^*, i \in \mathcal{S}$. Moreover, λ^* in conjunction with a differential cost vector $h = (h(1), \ldots, h(|\mathcal{S}|))$ satisfies Bellman's equation [40]

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

$$(13)$$

To solve (13), 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*. However, we can simulate the HEV model because the state space and control space are known. Thus, 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 (13). 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 explicitly estimating 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 directly updates the factors associated with an optimal policy, thereby avoiding the multiple policy evaluation steps of the policy iteration method.

It can be observed [41] that the Q-learning method for solving (13) is the following:

$$Q^{t+1}(i, \mu(i))$$
$$= \sum_{j=1}^{|\mathcal{S}|} P(j|i, \mu(i)) \cdot \left(k(i, \mu(i)) + \min_{\mu(j) \in \mathcal{C}(j)} Q^t(j, \mu(j)) \right)$$
$$- Q^t(i_o, \mu(i_o)) \quad \forall i \in \mathcal{S} \qquad (14)$$

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, which 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 (14). The resulting solution corresponds to the optimal control policy that minimizes the long-run expected average cost criterion [41].

B. Simulation Results

To validate the effectiveness of the power management controller using the Pareto control policy and compare it with DP, we used Autonomie [42]. Autonomie is a MATLAB/Simulink simulation package for powertrain and vehicle model development developed by the Argonne National Laboratory. A vehicle model representing a heavy-duty parallel HEV was used in this paper. The model consists of a diesel engine with a maximum power of 374 kW, an electric machine with a continuous power of 200 kW and a peak power of 360 kW, and a 12 V battery with 40-Ah energy capacity. The gear ratio between the engine and the output shaft is 3, whereas the gear ratio between the motor and the output shaft is 6.

The HEV model was simulated over standard driving cycles, established by the U.S. EPA for testing and measuring fuel economy and emissions. The following driving cycles were used: 1) the city–suburban heavy vehicle route (CSHVR); 2) the elementary urban driving cycle; 3) the extra urban driving cycle; 4) the FTP; 5) the Japanese 10-mode cycle; 6) the Japanese 15-mode cycle; 7) the New York city cycle; and 8) the UDDS.

To ensure that the Pareto efficiency exists in our study, we computed the normalized BSFC with respect to the engine torque for different engine speeds (Fig. 6). Similarly, we computed the normalized inverse of the motor efficiency with respect to the motor torque for different motor speeds (Fig. 7). From these plots, given a vehicle speed, and thus engine and motor speed, a different combination of engine and motor torque can yield different values of the normalized BSFC and inverse motor efficiencies. Applying these values to (8), we can observe that the Pareto efficiency set exists.

To derive the Pareto control policy, the multiobjective optimization problem (8) 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 N·m (discretized in 100 N·m). For each vehicle speed–torque combination, the Pareto optimal solution that minimizes (9)
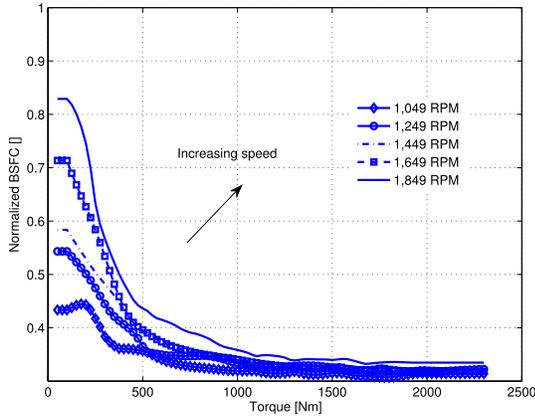
Fig. 6.   Normalized BSFC for different engine speeds with respect to engine torque.
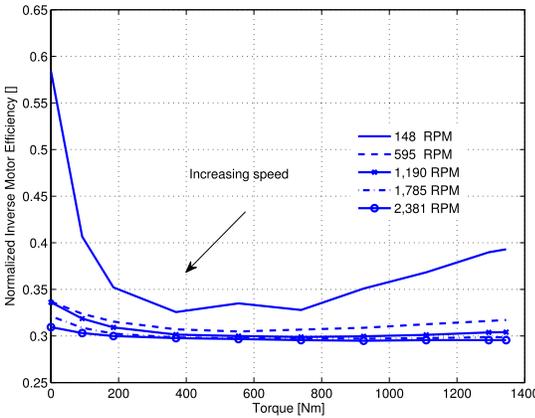
Fig. 7.   Normalized inverse of the motor efficiency for different motor speeds with respect to motor torque.
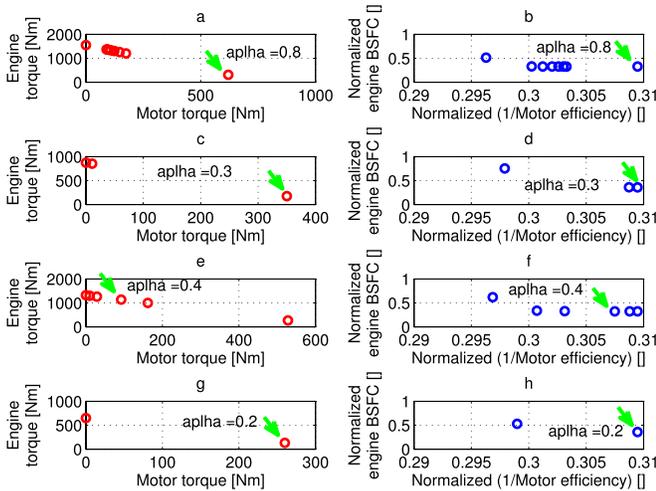
Fig. 8.   Pareto sets (on the left) $\mathcal{U}_{\text{Pareto}}$ and the Pareto efficiency sets (on the right) $\mathcal{Y}_{\text{eff}}$ corresponding to four different vehicle speed–torque combinations. (a) and (b) 21 km/h and 6975 N·m, (c) and (d) 59.4 km/h and 3937 N·m, (e) and (f) 48.6 km/h and 5940 N·m, and (g) and (h) 37.8 km/h and 2925 N·m. The green arrow indicates the solution selected by the Pareto control policy and the label the corresponding value of α.

was computed and stored in a table. Fig. 8 shows the Pareto sets $\mathcal{U}_{\text{Pareto}}$ and the Pareto efficiency sets $\mathcal{Y}_{\text{eff}}$ corresponding to different vehicle speed–torque combinations:
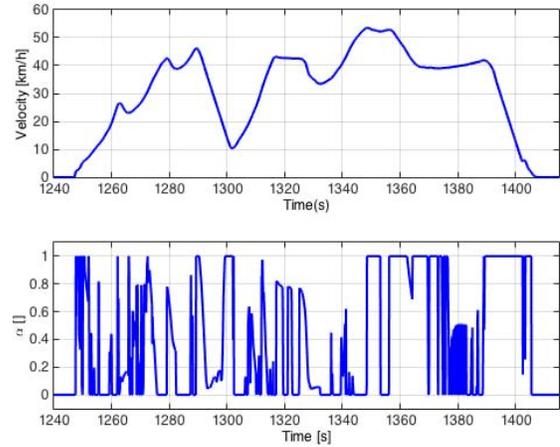
Fig. 9.   Trajectory of α over a portion of the CSHVR driving cycle.
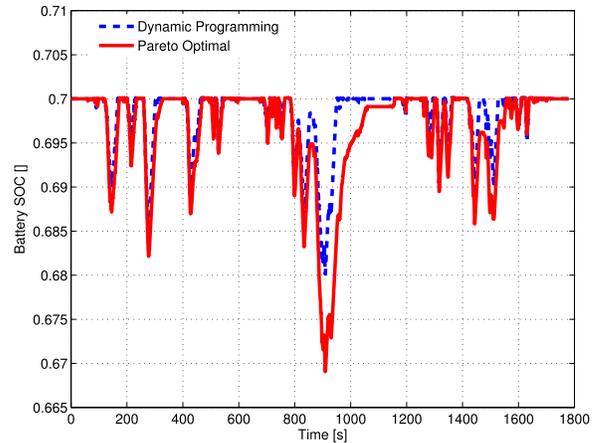
Fig. 10.   SOC of the battery using DP and the Pareto control policy over the CSHVR driving cycle.

21 km/h and 6975 N·m [Fig. 8(a) and (b)], 59.4 km/h and 3937 N·m [Fig. 8(c) and (d)], 48.6 km/h and 5940 N·m [Fig. 8(e) and (f)], and 37.8 km/h and 2925 N·m [Fig. 8(g) and (h)]. For each combination of vehicle speed and torque demand, the Pareto control policy selects the value from the Pareto set that minimizes (8). The green arrow in Fig. 8 indicates the solution selected by the Pareto control policy and the label the corresponding value of α.

As the driver drives the vehicle, the Pareto control policy is implemented by interpolating the values stored in the table. These values minimize (8) and correspond to different α derived from (9). Fig. 9 shows the trajectory of α over a portion of the CSHVR driving cycle, which essentially demonstrates how the driver's torque request is distributed between the engine and the motor. When $\alpha = 0$, only the motor powers the vehicle, whereas when $\alpha = 1$, only the engine powers the vehicle. For $0 < \alpha < 1$, both the engine and the motor power the vehicle.

The Pareto control policy was evaluated over eight driving cycles, and the effectiveness of its efficiency was compared with the DP control policy corresponding to the long-run average cost per unit time optimization criterion (13). The DP control policy was derived though simulation using
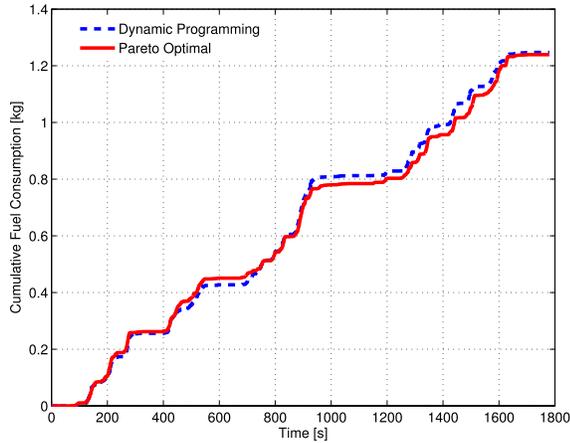
Fig. 11. Cumulative fuel consumption using DP and the Pareto control policy over the CSHVR driving cycle.

TABLE I
DIFFERENT DRIVING CYCLES

	INITIAL SOC [%]	FINAL SOC [%]	FUEL CONSUMPTION WITH DP [kg]	FUEL CONSUMPTION WITH PARETO [kg]
CSHVR[1]	70	70	1.24	1.24
ECE[2]	70	70	0.08	0.08
EUDC[3]	70	66	0.44	0.44
FTP[4]	70	66	1.99	1.99
JAPAN10[5]	70	70	0.06	0.06
JAPAN15[6]	70	70	0.25	0.25
NYCC[7]	70	69	0.35	0.35
UDDS[8]	70	70	1.34	1.34

[1] The city-suburban heavy vehicle route.
[2] The elementary urban driving cycle.
[3] The extra urban driving cycle.
[4] The federal test procedure.
[5] The Japanese 10 mode cycle.
[6] The Japanese 15 mode cycle.
[7] The New York city cycle.
[8] The urban dynamometer driving schedule.

Q-learning (14). The HEV model was repeatedly simulated over the same driving cycle until the Q-factors in (14) convergence. Fig. 10 shows SOC of the HEV battery using the DP and the Pareto control policies over the CSHVR driving cycle. For this driving cycle, Q-learning repeatedly ran 58 times until it convergences. The one-on-one correlation, as shown in Fig. 1, between SOC and the power added to the driver's power demand aimed at maintaining SOC at the target value, i.e., 70% in this case. Both control policies achieved the same cumulative fuel consumption (Fig. 11), which illustrates that the Pareto control policy is an optimal control policy that can be implemented online. The simulation results corresponding to the other driving cycles are summarized in Table I.

V. CONCLUSION

In this paper, 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 the simulation of an HEV model for different driving cycles, and it was compared with the DP control policy. 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.

This work has been extended [43], [44], and the proposed multiobjective optimization framework has considered the battery in the problem formulation in addition to the engine's BSFC and motor's efficiency. The extended optimization framework aims to enhance our understanding of the associated tradeoffs among the HEV subsystems, e.g., the engine, the motor, and the battery, and investigates the related implications for fuel consumption and battery capacity and lifetime. Addressing this problem can provide insights on how to prioritize these objectives based on consumers' needs and preferences.

REFERENCES

[1] A. A. Malikopoulos, "Supervisory power management control algorithms for hybrid electric vehicles: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1869–1885, Oct. 2014.
[2] I. Kolmanovsky, M. Van Nieuwstadt, and J. Sun, "Optimization of complex powertrain systems for fuel economy and emissions," in *Proc. IEEE Int. Conf. Control Appl.*, vol. 1. Aug. 1999, pp. 833–839.
[3] X. He and J. W. Hodgson, "Modeling and simulation for hybrid electric vehicles. I. Modeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 235–243, Dec. 2002.
[4] X. He and J. W. Hodgson, "Modeling and simulation for hybrid electric vehicles. II. Simulation," *IEEE Trans. Intell. Transp. Syst.*, vol. 3, no. 4, pp. 244–251, Dec. 2002.
[5] C.-C. Lin, H. Peng, J. W. Grizzle, and J.-M. Kang, "Power management strategy for a parallel hybrid electric truck," *IEEE Trans. Control Syst. Technol.*, vol. 11, no. 6, pp. 839–849, Nov. 2003.
[6] C.-C. Lin, H. Peng, and J. W. Grizzle, "A stochastic control strategy for hybrid electric vehicles," in *Proc. Amer. Control Conf.*, vol. 5. Jun./Jul. 2004, pp. 4710–4715.
[7] E. D. Tate, Jr., J. W. Grizzle, and H. Peng, "Shortest path stochastic control for hybrid electric vehicles," *Int. J. Robust Nonlinear Control*, vol. 18, no. 14, pp. 1409–1429, 2008.
[8] D. F. Opila, D. Aswani, R. McGee, J. A. Cook, and J. W. Grizzle, "Incorporating drivability metrics into optimal energy management strategies for hybrid vehicles," in *Proc. 47th IEEE Conf. Decision Control*, Dec. 2008, pp. 4382–4389.
[9] E. D. Tate, J. W. Grizzle, and H. Peng, "SP-SDP for fuel consumption and tailpipe emissions minimization in an EVT hybrid," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 3, pp. 673–687, May 2010.
[10] G. Paganelli, M. Tateno, A. Brahma, G. Rizzoni, and Y. Guezennec, "Control development for a hybrid-electric sport-utility vehicle: Strategy, implementation and field test results," in *Proc. Amer. Control Conf.*, vol. 6. Jun. 2001, pp. 5064–5069.
[11] A. Sciarretta, M. Back, and L. Guzzella, "Optimal control of parallel hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 12, no. 3, pp. 352–363, May 2004.
[12] C. Musardo, G. Rizzoni, and B. Staccia, "A-ECMS: An adaptive algorithm for hybrid electric vehicle energy management," in *Proc. 44th IEEE Conf. Decision Control, Eur. Control Conf.*, Dec. 2005, pp. 1816–1823.
[13] P. Pisu and G. Rizzoni, "A comparative study of supervisory control strategies for hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 15, no. 3, pp. 506–518, May 2007.

[14] 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 Trans. Control Syst. Technol.*, vol. 20, no. 3, pp. 593–603, May 2012.

[15] A. A. Malikopoulos, "Stochastic optimal control for series hybrid electric vehicles," in *Proc. Amer. Control Conf.*, Jun. 2013, pp. 1189–1194.

[16] L. Johannesson, M. Åsbogård, and B. Egardt, "Assessing the potential of predictive control for hybrid vehicle powertrains using stochastic dynamic programming," *IEEE Trans. Intell. Transp. Syst.*, vol. 8, no. 1, pp. 71–83, Mar. 2007.

[17] D. Ambuhl and L. Guzzella, "Predictive reference signal generator for hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 58, no. 9, pp. 4730–4740, Nov. 2009.

[18] A. A. Malikopoulos and J. P. Aguilar, "Optimization of driving styles for fuel economy improvement," in *Proc. 15th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2012, pp. 194–199.

[19] A. A. Malikopoulos and J. P. Aguilar, "An optimization framework for driver feedback systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 2, pp. 955–964, Jun. 2013.

[20] X. Huang, Y. Tan, and X. He, "An intelligent multifeature statistical approach for the discrimination of driving conditions of a hybrid electric vehicle," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 2, pp. 453–465, Jun. 2011.

[21] A. A. Malikopoulos, "Pareto efficient policy for supervisory power management control," in *Proc. IEEE 18th Int. Conf. Intell. Transp. Syst.*, Sep. 2015.

[22] O. Sundstrom, L. Guzzella, and P. Soltic, "Torque-assist hybrid electric powertrain sizing: From optimal control towards a sizing law," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 4, pp. 837–849, Jul. 2010.

[23] Y. Wang and S. Boyd, "Fast model predictive control using online optimization," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 2, pp. 267–278, Mar. 2010.

[24] P. R. Kumar and P. Varaiya, *Stochastic Systems*. Englewood Cliffs, NJ, USA: Prentice-Hall, Jun. 1986.

[25] O. Hernandez-Lerma and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*. New York, NY, USA: Springer-Verlag, 1996.

[26] A. A. Malikopoulos, P. Y. Papalambros, and D. N. Assanis, "A real-time computational learning model for sequential decision-making problems under uncertainty," *J. Dyn. Syst., Meas., Control*, vol. 131, no. 4, pp. 041010-1–041010-8, 2009.

[27] A. A. Malikopoulos, "Convergence properties of a computational learning model for unknown Markov chains," *J. Dyn. Syst., Meas., Control*, vol. 131, no. 4, pp. 041011-1–041011-7, 2009.

[28] A. A. Malikopoulos, P. Y. Papalambros, and D. N. Assanis, "Online identification and stochastic control for autonomous internal combustion engines," *J. Dyn. Syst., Meas., Control*, vol. 132, no. 2, pp. 024504-1–024504-6, 2010.

[29] F. Logist, S. Sager, C. Kirches, and J. F. Van Impe, "Efficient multiple objective optimal control of dynamic systems with integer controls," *J. Process Control*, vol. 20, no. 7, pp. 810–822, Aug. 2010.

[30] F. Logist, B. Houska, M. Diehl, and J. Van Impe, "Fast Pareto set generation for nonlinear optimal control problems with multiple objectives," *Struct. Multidiscip. Optim.*, vol. 42, no. 4, pp. 591–603, Oct. 2010.

[31] A. A. Malikopoulos, *Real-Time, Self-Learning Identification and Stochastic Optimal Control of Advanced Powertrain Systems*. Ann Arbor, MI, USA: ProQuest, Sep. 2011.

[32] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. New York, NY, USA: Wiley, 2008.

[33] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 4th ed. Belmont, MA, USA: Athena Scientific, 2007.

[34] G. R. Grimmett and D. R. Stirzaker, *Probability and Random Processes*, 3rd ed. New York, NY, USA: Oxford Univ. Press, Aug. 2001.

[35] S. M. Ross, *Stochastic Processes*, 2nd ed. New York, NY, USA: Wiley, Jan. 1995.

[36] H. J. Kushner, *Introduction to Stochastic Control*. New York, NY, USA: Holt, Rinehart and Winston, 1971.

[37] P. Varaiya, "Optimal and suboptimal stationary controls for Markov chains," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 388–394, Jan. 1978.

[38] A. A. Malikopoulos, V. Maroulas, and J. Xiong, "A multiobjective optimization framework for stochastic control of complex systems," in *Proc. Amer. Control Conf.*, Jul. 2015.

[39] M. Ehrgott, *Multicriteria Optimization*, 2nd ed. New York, NY, USA: Springer-Verlag, 2005.

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

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

[42] *Autonomie*. [Online]. Available: http://www.autonomie.net/, accessed Jul. 1, 2015.

[43] M. L. Shaltout, A. A. Malikopoulos, S. Pannala, and D. Chen, "Multidisciplinary decision making and optimization for hybrid electric propulsion systems," in *Proc. IEEE Int. Electr. Veh. Conf.*, Dec. 2014, pp. 1–6.

[44] M. L. Shaltout, A. A. Malikopoulos, S. Pannala, and D. Chen, "A consumer-oriented control framework for performance analysis in hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 4, pp. 1451–1464, Jul. 2015.

Andreas A. Malikopoulos (M'06) received the Diploma degree in mechanical engineering from the National Technical University of Athens, Athens, Greece, in 2000, and the M.S. and Ph.D. degrees from the Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA, in 2004 and 2008, respectively.

He was a Senior Researcher with General Motors Global Research and Development, Warren, MI, USA, conducting research in the area of stochastic optimization and control of advanced propulsion systems. He is currently the Deputy Director of the Urban Dynamics Institute and an Alvin M. Weinberg Fellow with the Energy and Transportation Science Division, Oak Ridge National Laboratory (ORNL), Oak Ridge, TN, USA. His research interests at ORNL span several fields, including analysis, optimization, and control of complex systems; stochastic control; decentralized stochastic systems; and stochastic scheduling and resource allocation problems. The emphasis is on applications related to energy, transportation, and operations research.