Centralized Stochastic Optimal Control of Complex Systems

Andreas A. Malikopoulos, *Member, IEEE*

***Abstract*— We consider the problem of minimizing the long-run expected average cost of a complex system consisting of interactive subsystems with an emphasis on advanced propulsion systems, e.g., hybrid electric vehicles. We formulate a multiobjective optimization problem of the one-stage expected costs of the subsystems and provide a framework to prove that the control policy yielding the Pareto optimal solution minimizes the average cost criterion of the system.**

I. Introduction

Complex systems consist of interdependent, diverse entities, that are connected with each other and can adapt, i.e., they can respond to their local and global environment. Complex systems encountered in virtually all transportation- and energy-related applications. Hybrid electric vehicles (HEVs) and plug-in HEVs (PHEVs) are such complex systems. A typical HEV powertrain configuration (Fig. 1) consists of various interdependent subsystems, e.g., the internal combustion engine, the electric machines (motor and generator), and the energy storage system (battery), that interact with each other and adapt appropriately to provide the power demanded by the driver. In cruising, the power demanded by the driver P_{driver} is expressed by a positive amount of torque. In braking, HEVs/PHEVs can recover energy, and thus, recharge the energy storage system. The motor acts as a generator and absorbs the maximum possible amount as imposed by the system's physical constraints. If a residual amount of braking remains, the friction brakes handle this.

Thus the main advantage of HEVs and PHEVs is the existence of these individual subsystems that can power the vehicle either separately or in combination. The supervisory power management control algorithm determines the control policy that designates how to distribute the power demanded by the driver to these subsystems to optimize vehicle's efficiency. The problem is formulated as sequential decision making under uncertainty. A key aspect of this problem is that each decision may influence the circumstances under which future decisions will be made. Thus, the decision maker must balance the desire for low present cost to avoid future situations where high cost is inevitable. For example, the supervisory controller may need to save the energy available in the battery and use it later when excessive stop-and-go driving is happening that the engine is inefficient for.

Deriving the optimal control policy online and for different drivers constitutes a challenging control problem and has been the object of intense study since 1998. Dynamic programming (DP) has been widely used as the principal method for deriving the optimal control policy in HEVs/PHEVs [1]–[3] for a given vehicle speed profile (i.e., commute from point A to point B). DP has been extended to the stochastic problem formulation in which the optimal control policy is derived also offline for a family of vehicle speed profiles [4]–[9].

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. Nonetheless, DP is the only general approach for sequential decision making problems under uncertainty, and even when it is computationally prohibitive, it serves as the basis for other, suboptimal approaches. To address these issues, research efforts have been concentrated on developing online power management algorithms [10]–[12]. A detailed survey of the supervisory power management control algorithms that have been reported in the literature to date can be found in [13].

In this paper, and motivated by previous work [14], we seek to establish a framework for the analysis and stochastic optimization of complex systems with constraints and properties in line with HEVs/PHEVs. We generalize this framework, however, so that it can be applied to other complex systems consisting of subsystems that interact with each other and their environment with constraints consistent to those studied here. The contribution of this paper is the formulation and solution of a multiobjective optimization problem of the one-stage expected costs of all interactive subsystems yielding an operating point that minimizes the long-run expected average cost of the system.

The remainder of the paper proceeds as follows. In Section II, we introduce our notation and formulate the problem. In Section III, we develop a multiobjective optimization framework to address the problem and we show that the control policy yielding the Pareto optimal solution minimizes the long-run expected average cost criterion. Finally, we present an illustrative example in Section IV and concluding remarks in Section V.

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.

A.A. Malikopoulos is with the Energy & Transportation Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (phone: 865-946-1529; fax: 865-946-1354; e-mail: andreas@ornl.gov).

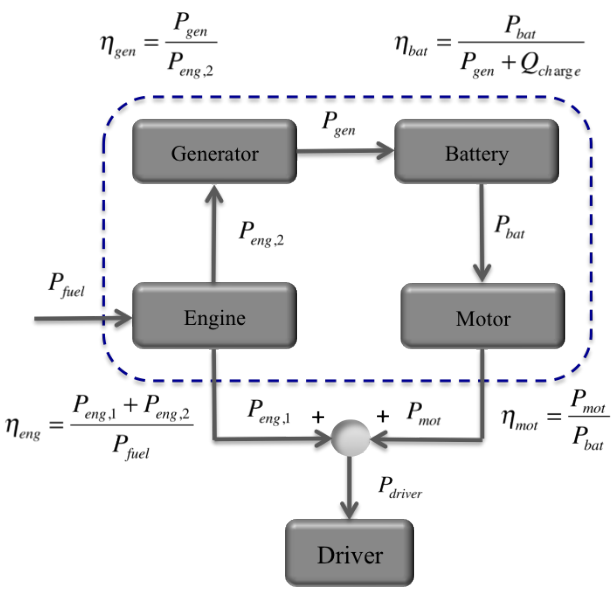

Fig. 1. A hybrid electric vehicle configuration: $\eta_{eng}, \eta_{gen}, \eta_{bat}$, and η_{mot} are the efficiencies of the engine, generator, battery, and motor; P_{driver} is the power demanded by the driver; P_{fuel} is the heating value per unit time of the fuel; $P_{eng,1}$ and $P_{eng,2}$ are the amounts of engine power provided to the driver and generator respectively; P_{gen} is the electrical power created from the generator; P_{bat} is the electrical power drawn from the battery to power the motor; Q_{charge} is the initial capacity of the battery; and P_{mot} is the mechanical power that the motor provides to the driver.

II. Problem Formulation

A. Notation

We denote random variables with upper case letters, and their realization with lower case letters, e.g., for a random variable X, x denotes its realization. Subscripts denote time, and subscripts in parentheses denote subsystems; for example, $X_{t(i)}$ denotes the random variable of the subsystem i at time t, and $x_{(i)}$ its realization. The shorthand notation $X_{t(1:N)}$ denotes the vector of random variables $\big(X_{t(1)}, X_{t(2)}, \cdots, X_{t(N)}\big)$ and $x_{(1:N)}$ denotes the vector of their realization $\big(x_{(1)}, x_{(2)}, \cdots, x_{(N)}\big)$. $\mathbb{P}(\cdot)$ 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. The System Model

We consider a system consisting of N subsystems. The subsystems interact with each other and their environment. At time $t, t = 1, 2, \cdots, T$, the state of each subsystem $i, X_{t(i)}$, takes values in a finite state space $\mathcal{S}_{(i)}$, which is a metric space. For each subsystem i, we also consider a finite control space $\mathcal{U}_{(i)}$, which is also a metric space, from which control actions, $U_{t(i)}$, are chosen.

The initial state of the system $X_{0(1:N)}$ is a random variable taking values in the system's state space, $\mathcal{S} = \prod_{i=1}^{N} \mathcal{S}_{(i)}$. The evolution of the state is imposed by the discrete-time equation

$$X_{t+1(1:N)} = f\big(X_{t(1:N)}, U_{t(1:N)}, W_{t(1:N)}\big), \tag{1}$$

where $W_{t(1:N)}$ is the input from the environment. In a HEV, for example, $W_{t(1:N)}$ corresponds to the driver's power demand. The state of the system can be completely observed.

In our formulation, a state-dependent constraint is incorporated; that is, for each realization of the state of the subsystem i, $X_{t(i)} = x_{(i)}$, there is a nonempty and closed set $\mathcal{C}(x_{(i)}) := \big\{u_{(i)} | X_{t(i)} = x_{(i)}\big\} \subset \mathcal{U}_{(i)}$ of feasible control actions when the system is in state $x_{(i)}$. For each subsystem i, we denote the set of admissible state/action pairs

$$\Gamma_{(i)} :\, = \{(x_{(i)}, u_{(i)}) | x_{(i)} \in \mathcal{S}_{(i)} \text{ and } u_{(i)} \in \mathcal{C}(x_{(i)})\} \tag{2}$$

such that it is a measurable subset of $\mathcal{S}_{(i)} \times \mathcal{U}_{(i)}$, so it is closed with respect to the induced topology on $\mathcal{S}_{(i)} \times \mathcal{U}_{(i)}$, and thus it is compact. The set of admissible state/action pairs for the system is

$$\Gamma :\, = \prod_{i=1}^{N} \Gamma_{(i)} = \{(x_{(1:N)}, u_{(1:N)})) | x_{(1:N)} \in \mathcal{S} \text{ and } u_{(1:N)} \in \mathcal{C}(x_{(1:N)})\}, \tag{3}$$

where $\mathcal{C}(x_{(1:N)}) = \prod_{i=1}^{N} \mathcal{C}_{(i)}(x_{(i)})$.

For each state of the system $X_{t(1:N)} = x_{(1:N)}$, we define the Borel measurable functions $\mu : \mathcal{S} \to \mathcal{U}$, where $\mathcal{U} = \prod_{i=1}^{N} \mathcal{U}_{(i)}$, that map the state space to the control action space defined as the control law. When the system is at state $X_{t(1:N)} = x_{(1:N)}$, the controller chooses action according to the control law $u_{(1:N)} = \mu\big(x_{(1:N)}\big)$.

Definition 1: Each sequence of the measurable functions μ is defined as a stationary control policy of the system

$$\pi :\, = \big(\mu(1), \mu(2), \cdots, \mu(|\mathcal{S}|)\big), \tag{4}$$

where $|\mathcal{S}|$ is the cardinality of the system's state space $\mathcal{S}$.

Let Π denote the set of the collection of the stationary control policies

$$\Pi :\, = \Big\{\pi | \pi = \big(\mu(1), \mu(2), \cdots, \mu(|\mathcal{S}|)\big)\Big\}. \tag{5}$$

The stationary control policy π operates as follows. Associated with each state $X_{t(1:N)} = x_{(1:N)}$ is the function $\mu\big(x_{(1:N)}\big) \in \mathcal{C}(x_{(1:N)})$. If at any time the controller finds the system in state $x_{(1:N)}$, then the controller always chooses the action based on the function $\mu\big(x_{(1:N)}\big)$. In a HEV, for instance, based on the current HEV state, the controller chooses the amount of power for each subsystem that should be either delivered to the driver or to other subsystems. A stationary policy depends on the history of the process only through the current state, and thus to implement it, the controller only needs to know the current state of the system. The advantages for implementation of a stationary policy are apparent as it requires the storage of less information than required to implement a general policy. Thus a stationary policy is attractive in automotive-related applications where computational and storage power is limited onboard a vehicle.

At each stage t, the controller observes the state of the system, $X_{t(1:N)} = x_{(1:N)} \in \mathcal{S}$, and an action, $u_{t(1:N)} = \mu(X_{t(1:N)})$, is realized from the feasible set of actions at that state. At the same stage t, an uncertainty, $W_{t(1:N)}$, is incorporated in the system. At the next stage, $t+1$, the system transits to the state $X_{t+1(1:N)} = x'_{(1:N)} \in \mathcal{S}$ and a transition cost for each subsystem i , $c_{t(i)}\big(X_{t+1(i)}|X_{t(i)}, U_{t(i)}\big)$, where $c_{t(i)} : \mathcal{S}_{(i)} \times \mathcal{C}(x_{(i)}) \times \mathcal{S}_{(i)} \to \mathbb{R}$, and for the system, $c_t\big(X_{t+1(1:N)}|X_{t(1:N)}, U_{t(1:N)}\big)$, where $c_t : \mathcal{S} \times \mathcal{C}(x_{(1:N)}) \times \mathcal{S} \to \mathbb{R}$, are incurred.

C. Assumptions

In the model described above, we consider the following assumptions:

(A1) The set Γ contains the graph of all Borel measurable functions $\big(\mu(1), \mu(2), \cdots, \mu(|\mathcal{S}|)\big)$.

(A2) The input from the uncertainty $W_{t(1:N)}$ is a sequence of independent random variables, independent of the initial state $X_{0(1:N)}$, and take values in the finite sets $\mathcal{W}$.

(A3) For each stationary control policy π, the Markov chain $\big\{X_{t(1:N)}|t = 1, 2, \cdots\big\}$ has a stationary probability distribution.

(A4) The one-stage expected cost of the system, $k_t^{\pi} : \Gamma \to \mathbb{R}$, is a function of the one-stage costs of the subsystems and it is bounded.

(A5) We relax the relationship imposed in [14] between the one-stage expected costs of the subsystems, and we assume that the one-stage cost of each subsystem is a monotonic decreasing function with respect to the one-stage expected cost of the other subsystems. The implicit belief here is that there are associated tradeoffs between the subsystems. In HEV/PHEV, for example, from Fig. 1 it can be inferred that the efficiency of the generator is a decreasing function of the efficiencies of the battery and engine

$$\eta_{gen} = \frac{P_{bat} - \eta_{bat} \cdot Q_{charge}}{\eta_{bat} \cdot \big(\eta_{eng} \cdot P_{fuel} - P_{eng,1}\big)}. \tag{6}$$

Thus, by operating the battery and the engine in a way to maximize their efficiency, the efficiency of the generator is decreasing. Similar tradeoffs appear also between the other subsystems in HEVs/PHEVs.

We are concerned with deriving a stationary optimal control policy π to minimize the long-run expected average cost of the system

$$J(\pi) = \lim_{T\to\infty} \frac{1}{T+1} \mathbb{E}^{\pi} \left[\sum_{0}^{T} k_t^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big) \right], \tag{7}$$

where $k_t^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big)$ is the one-stage expected cost of the system.

III. Multiobjective Optimization Analysis

A. Pareto Control Policy

Various methods can be used to solve (7) offline. In this paper, we seek the theoretical framework that will yield the optimal control policy online while the subsystems interact with each other. At each stage t, we need to identify an operating point among the subsystems that will minimize the average cost of the system.

Let's consider the function $f : \Gamma \to \mathbb{R}^N$,

$$f = \Bigg(k_{t(1)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big), k_{t(2)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big), ..., k_{t(N)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big) \Bigg), \tag{8}$$

where $k_{t(i)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big)$ is the one-stage expected cost for each subsystem i and the following multiobjective optimization problem

$$\min_{U_{t(1:N)} \in \mathcal{C}(x_{(1:N)})} \Bigg(k_{t(1)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big), k_{t(2)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big), ..., k_{t(N)}^{\pi}\big(X_{t(1:N)}, U_{t(1:N)}\big) \Bigg). \tag{9}$$

The result of the optimization problem (9) is called Pareto efficiency. In a Pareto efficiency allocation among agents, no one can be made better without making at least one other agent worse. The following is the a formal definition of the Pareto efficiency or Pareto efficient set from [15] adapted to the problem formulation (9).

Definition 2 [15]: A solution $U_{t(1:N)} = u_{(1:N)}^{o} \in \mathcal{U}$ is called Pareto optimal if, for each realization of the state $X_{t(1:N)} = x_{(1:N)}$, there is no $u_{(1:N)} \in \mathcal{U}$ such that $k_{t(i)}(x_{(1:N)}, u_{(1:N)}) \leq k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{o})$, for $i = 1, \ldots, N$. If $u_{(1:N)}^{o}$ is Pareto optimal, $k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{o})$ is called Pareto efficient. If $u_{(1:N)}^{1}, u_{(1:N)}^{2} \in \mathcal{U}$ and $k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{1}) < k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{2})$, we say $u_{(1:N)}^{1}$ dominates $u_{(1:N)}^{2}$ and $k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{1})$ dominates $k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{2})$. The set of all Pareto optimal solutions $u_{(1:N)}^{o} \in \mathcal{U}$ is the Pareto set, $\mathcal{U}_{Pareto}$. The set of all efficient points $k_{t(i)}(x_{(1:N)}, u_{(1:N)}^{o}) \in \mathcal{Y}$ where $u_{(1:N)}^{o} \in \mathcal{U}_{Pareto}$ is the Pareto efficient set $\mathcal{Y}_{eff}$.

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

Proposition 1 [15]: Let Γ be a nonempty and compact set, and the one-stage expected cost for each subsystem i, $k_{t(i)}^{\pi}(X_{t(i)}, U_{t(i)}) : \Gamma \to \mathbb{R}$, be lower semicontinuous for all $i = 1, \cdots, N, l \in \mathbb{N}$. Then the Pareto set is not empty.

In our problem, the set of admissible state/action pairs, Γ, is a nonempty compact set. Furthermore, the one-stage expected cost for each subsystem i, $k_{t(i)}^{\pi}(X_{t(i)}, U_{t(i)})$, is a continuous function. Consequently, the Pareto efficient set exists.

Definition 3: The Pareto control policy π^o is defined as the policy that yields the Pareto efficient one-stage expected cost for each subsystem i, $k_{(i)}^{\pi^o}(X_{t(1:N)}, U_{t(1:N)}^{o})$, at each realization of the system state $X_{t(1:N)} = x_{(1:N)}$.

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

To simplify notation, in the rest of the paper the one-stage expected cost of each subsystem i,

$k^{\pi}_{t(i)}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, and the one-stage expected cost of the system, $k^{\pi}_{t}\big(X_{t(1:N)}, U_{t(1:N)}\big)$, incurred when the system operates under the control policy π, will be denoted by $k^{\pi}_{t(i)}$ and k^{π}_{t} respectively.

We've assumed (A5) that the one-stage expected cost of each subsystem i is a monotonic decreasing function, $\gamma_{(i)}$, with respect to the expected cost of the other subsystems. So, for each subsystem, when the system operates under the control policy π the one-stage expected cost is given as a function of the expected costs of the other subsystems

$$k^{\pi}_{t(i)} = \gamma_{(i)}\big(k^{\pi}_{t(1)}, k^{\pi}_{t(2)}, \cdots, k^{\pi}_{t(i-1)}, k^{\pi}_{t(i+1)}, \cdots, k^{\pi}_{t(N)}\big). \tag{10}$$

Thus for any other subsystem j and for any two control policies $\pi, \pi' \in \Pi$ such that $k^{\pi}_{t(j)} \leq k^{\pi'}_{t(j)}$, if we fix the one-stage cost of the other subsystems, then

$$k^{\pi}_{t(i)} = \gamma_{(i)}\big(\cdots, k^{\pi}_{t(j)}, \cdots\big) \geq k^{\pi'}_{t(i)} = \gamma_{(i)}\big(\cdots, k^{\pi'}_{t(j)}, \cdots\big). \tag{11}$$

1) Problem 1: We consider the special case where the one-stage expected cost of the system is considered to be a decreasing function, δ, with respect to the expected cost of each subsystem i. So, when the system operates under the control policy π, the one-stage expected cost is given as

$$k^{\pi}_{t} = \delta\big(k^{\pi}_{t(1)}, k^{\pi}_{t(2)}, \cdots, k^{\pi}_{t(N)}\big). \tag{12}$$

Thus for each subsystem i and for any two control policies $\pi, \pi' \in \Pi$ such that $k^{\pi}_{t(i)} \leq k^{\pi'}_{t(i)}$, then

$$k^{\pi}_{t} = \delta\big(\cdots, k^{\pi}_{t(i)}, \cdots\big) \geq k^{\pi'}_{t} = \delta\big(\cdots, k^{\pi'}_{t(i)}, \cdots\big). \tag{13}$$

Theorem 1: For the Problem 1, the Pareto control policy π^o yields the minimum one-stage expected cost of the system, $k^{*}_{t} = k^{\pi^o}_{t}$.

Proof: For the ease of notation, we first prove the result for 2 subsystems, and then we extend it for N subsystems.

From (10), the one-stage expected cost for the subsystem 1 is decreasing function with respect to subsystem 2

$$k^{\pi}_{t(1)} = \gamma_{(1)}\big(k^{\pi}_{t(2)}\big)\big(\text{or alternatively, } k^{\pi}_{t(2)} = \gamma_{(2)}\big(k^{\pi}_{t(1)}\big)\big), \tag{14}$$

and from (12), we have $k^{\pi}_{t} = \delta\big(k^{\pi}_{t(1)}, k^{\pi}_{t(2)}\big)$.

Suppose that there is another control policy π' that yields the minimum one-stage cost of the system at each realization of the state, namely

$$k^{\pi'}_{t} = \delta\big(k^{\pi'}_{t(1)}, k^{\pi'}_{t(2)}\big) < k^{\pi^o}_{t} = \delta\big(k^{\pi^o}_{t(1)}, k^{\pi^o}_{t(2)}\big), \tag{15}$$

and thus from (13), this implies $k^{\pi'}_{t(1)} > k^{\pi^o}_{t(1)}$ and $k^{\pi'}_{t(2)} > k^{\pi^o}_{t(2)}$.

However, from (14)

$$k^{\pi'}_{t(1)} = \gamma_{(1)}\big(k^{\pi'}_{t(2)}\big) > k^{\pi^o}_{t(1)} = \gamma_{(1)}\big(k^{\pi^o}_{t(2)}\big), \tag{16}$$

implies that

$$k^{\pi'}_{t(2)} < k^{\pi^o}_{t(2)}, \tag{17}$$

which contradicts the hypothesis.

Now, we extend the result for N subsystems following similar arguments. Suppose that there is another control policy π' that yields the minimum one-stage cost of the system at each realization of the state, namely

$$\begin{aligned} k^{\pi'}_{t} &= \delta\big(k^{\pi'}_{t(1)}, k^{\pi'}_{t(2)}, \cdots, k^{\pi'}_{t(N)}\big) < \\ k^{\pi^o}_{t} &= \delta\big(k^{\pi^o}_{t(1)}, k^{\pi^o}_{t(2)}, \cdots, k^{\pi^o}_{t(N)}\big), \end{aligned} \tag{18}$$

and thus from (13), for each subsystem i we have

$$k^{\pi'}_{t(i)} > k^{\pi^o}_{t(i)}. \tag{19}$$

However, from (11) we infer that there is a subsystem j whose expected cost is a monotonic decreasing function of the cost of subsystem i (A6), and thus since $k^{\pi'}_{t(i)} > k^{\pi^o}_{t(i)}$

$$\begin{aligned} k^{\pi'}_{t(j)} &= \gamma_{(j)}\big(\cdot, \cdots, k^{\pi'}_{t(i)}, \cdots, \cdot\big) < \\ k^{\pi^o}_{t(j)} &= \gamma_{(j)}\big(\cdot, \cdots, k^{\pi^o}_{t(i)}, \cdots, \cdot\big), \end{aligned} \tag{20}$$

which contradicts the hypothesis.

■

2) Problem 2: We consider the general case where the one-stage expected cost of the system is both a monotonic decreasing function with respect to the expected cost of a group of subsystems and a monotonic increasing function with respect to the expected cost of a group of the other subsystems.

Definition 4: The group of subsystems whose expected costs are a decreasing function with respect to the cost of the system is defined as the *minor* group.

Definition 5: The group of subsystems whose expected costs are an increasing function with respect to the cost of the system is defined as the *principal* group.

Without loss of generality, we assume that the minor group consists of the subsystems $1, 2, \cdots, m, m \in \mathbb{N}$, and the principal group consists of the subsystems $m+1, \cdots, N$. Thus, from Definition 4, for each subsystem i in the minor group and for any two control policies $\pi, \pi' \in \Pi$ such that $k^{\pi}_{t(i)} \leq k^{\pi'}_{t(i)}$, if we fix the one-stage cost of the other subsystems in both minor and principal groups we have

$$k^{\pi}_{t} = \delta\big(\cdots, k^{\pi}_{t(i)}, \cdots\big) \geq k^{\pi'}_{t} = \delta\big(\cdots, k^{\pi'}_{t(i)}, \cdots\big). \tag{21}$$

Similarly, from Definition 5, for each subsystem j in the principal group and for any two control policies $\pi, \pi' \in \Pi$ such that $k^{\pi}_{t(j)} \leq k^{\pi'}_{t(j)}$, if we fix the one-stage cost of the other subsystems in both minor and principal groups we have

$$k^{\pi}_{t} = \delta\big(\cdots, k^{\pi}_{t(j)}, \cdots,\big) \leq k^{\pi'}_{t} = \delta\big(\cdots, k^{\pi'}_{t(j)}, \cdots,\big). \tag{22}$$

Let $\bar{\pi}$ be the Pareto control policy of the minor group. Namely, from Definition 3 if the system operates under the control policy $\bar{\pi}$, at each state of the system, then the expected costs of the subsystems in the minor group are Pareto efficient. Namely, for any control policy $\pi \in \Pi$, if we ignore the one-stage expected cost of each subsystem in the principal group

$$k^{\bar{\pi}}_{t(i)} \leq k^{\pi}_{t(i)}, i = 1, \ldots, m. \tag{23}$$

Let $\bar{\pi^o}$ be the Pareto control policy of the principal group. Then, from Definition 3, if the system operates under the control policy $\bar{\pi^o}$, then the expected costs of the subsystems in the principal group are Pareto efficient at each state of the system. Namely, for any control policy $\pi \in \Pi$, if we ignore the one-stage expected cost of each subsystem in the minor group then we have

$$k_{t(j)}^{\bar{\pi}^o} \leq k_{t(j)}^{\pi}, j = m+1, \ldots, N. \tag{24}$$

Since the expected cost of the system exhibits different correlations with the minor and the principal group, we are interested in studying the impact of each group on the optimal control policy. In particular, we would like to investigate how the Pareto control policy of each group of subsystems is related to the optimal control policy for the system.

Theorem 2: In a complex system consisting of both minor and principal groups, the Pareto control policy $\bar{\pi}^o$ of the principal group yields the minimum one-stage expected cost of the system. Namely, for any control policy $\pi \in \Pi$, $k_t^{\bar{\pi}^o} \leq k_t^{\pi}$.

Proof: Let $\bar{\pi}$ and $\bar{\pi}^o$ be the Pareto control policies of the minor and principal group respectively. Then for the minor group and for any control policy $\pi \in \Pi$,

$$k_{t(i)}^{\bar{\pi}} \leq k_{t(i)}^{\pi}, i = 1, \ldots, m.$$

From definition 4 this implies

$$\begin{aligned} k_t^{\bar{\pi}} = \delta\big(k_{t(1)}^{\bar{\pi}}, k_{t(2)}^{\bar{\pi}}, \cdots, k_{t(m)}^{\bar{\pi}}, \cdots\big) \geq \\ k_t^{\pi} = \delta\big(k_{t(1)}^{\pi}, k_{t(2)}^{\pi}, \cdots, k_{t(m)}^{\pi}, \cdots\big). \end{aligned} \tag{25}$$

Since the last inequality is true for any control policy $\pi \in \Pi$, it is also true for the Pareto control policy of the principal group, $\bar{\pi}^o \in \Pi$, thus we have

$$\begin{aligned} k_t^{\bar{\pi}} = \delta\big(k_{t(1)}^{\bar{\pi}}, k_{t(2)}^{\bar{\pi}}, \cdots, k_{t(m)}^{\bar{\pi}}, \cdots, \cdot\big) \geq \\ k_t^{\bar{\pi}^o} = \delta\big(k_{t(1)}^{\bar{\pi}^o}, k_{t(2)}^{\bar{\pi}^o}, \cdots, k_{t(m)}^{\bar{\pi}^o}, \cdots, \cdot\big). \end{aligned} \tag{26}$$

Similarly, for the principal group and for any control policy $\pi \in \Pi$ we have

$$k_{t(j)}^{\bar{\pi}^o} \leq k_{t(j)}^{\pi}, j = m+1, \ldots, N$$

which implies

$$\begin{aligned} k_t^{\bar{\pi}^o} = \delta\big(\cdots, k_{t(m+1)}^{\bar{\pi}^o}, k_{t(m+2)}^{\bar{\pi}^o}, \cdots, k_{t(N)}^{\bar{\pi}^o}\big) \leq \\ k_t^{\pi} = \delta\big(\cdots, k_{t(m+1)}^{\pi}, k_{t(m+2)}^{\pi}, \cdots, k_{t(N)}^{\pi}\big). \end{aligned} \tag{27}$$

Since the last inequality is true for any control policy $\pi \in \Pi$, it is also true for the Pareto control policy of the minor group, $\bar{\pi} \in \Pi$, thus we have

$$\begin{aligned} k_t^{\bar{\pi}^o} = \delta\big(\cdot, \cdots, k_{t(m+1)}^{\bar{\pi}^o}, k_{t(m+2)}^{\bar{\pi}^o}, \cdots, k_{t(N)}^{\bar{\pi}^o}\big) \leq \\ k_t^{\bar{\pi}} = \delta\big(\cdot, \cdots, k_{t(m+1)}^{\bar{\pi}}, k_{t(m+2)}^{\bar{\pi}}, \cdots, k_{t(N)}^{\bar{\pi}}\big). \end{aligned} \tag{28}$$

Hence from (25) and (28)

$$\begin{aligned} k_t^{\bar{\pi}^o} = \delta\big(k_{t(1)}^{\bar{\pi}^o}, k_{t(2)}^{\bar{\pi}^o}, \cdots, k_{t(N)}^{\bar{\pi}^o}\big) \leq \\ k_t^{\pi} = \delta\big(k_{t(1)}^{\pi}, k_{t(2)}^{\pi}, \cdots, k_{t(N)}^{\pi}\big). \end{aligned} \tag{29}$$

■

Corollary 1: In a complex system consisting of both minor and principal groups with the assumptions and constraints consistent to those considered here, it is sufficient to focus only on the principal group.

Theorem 3 [16]: In a complex system consisting of both minor and principal groups, the Pareto control policy $\bar{\pi}^o$ of the principal group is the optimal control policy π^* that minimizes the long-run expected average cost criterion (7).

Proof: Let $\bar{\pi}^o$ be the Pareto control policy. From Theorem 2 we have that for each realization of the state $X_{t(1:N)}$, $k^{\bar{\pi}^o}\big(X_{t(1:N)}, U_{t(1:N)}\big) \leq k^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)$ for any control policy $\pi' \in \Pi$. Since the system's one-stage cost is bounded (A4), taking the expected average sum from $t = 0$ up to a finite time $T \in \mathbb{N}$ is well defined and finite. Thus

$$\begin{aligned} &\frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k^{\bar{\pi}^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] \\ &\leq \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \end{aligned} \tag{30}$$

Taking the liminf as T goes to infinity

$$\begin{aligned} &\liminf_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k^{\bar{\pi}^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] \\ &\leq \liminf_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \end{aligned} \tag{31}$$

Since each stationary control policy has a single ergodic class (A3) the limit in (31) is well defined; hence for all $\pi' \in \Pi$

$$\begin{aligned} J^{\bar{\pi}^o} = \lim_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=0}^{T} k^{\bar{\pi}^o}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right] \leq \\ J^{\pi'} = \lim_{T\to\infty} \frac{1}{T+1}\mathbb{E}^{\pi}\left[\sum_{t=1}^{T} k^{\pi'}\big(X_{t(1:N)}, U_{t(1:N)}\big)\right]. \end{aligned} \tag{32}$$

■

IV. Illustrative Examples

A. Power Management Control of a Hybrid Electric Vehicle: A System with Subsystems of a Principal Group

The results presented here have been used in the problem of optimizing online the power management control in a parallel HEV configuration [16] consisting of subsystems of a principal group. To compare the Pareto control policy with the optimal control policy of DP, we need to solve $|\mathcal{S}|$ linear equations, where $|\mathcal{S}|$ is the cardinality of the state space. However for complex systems such as HEVs a more attractive method to derive the optimal control policy is to learn the optimal control policy using the Q-learning method rather than estimating explicitly the transition probabilities and stage costs as it is shown in [16]. 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 the policy iteration method.

The Pareto control policy was validated through simulation in a parallel HEV, and it was compared with the control policy of DP for the long-run expected average cost. In the case of the Pareto control policy, the model ran just once. The DP control policy was derived by simulating the HEV model repeatedly over the same driving cycle until the Q-factors convergence. For the particular driving cycle the model ran repeatedly 58 times until convergence. Both control policies achieved the same cumulative fuel consumption as illustrated in Fig. 2, demonstrating that the Pareto control policy is the optimal control policy with respect to the average cost criterion.

There are still open issues, however, with practical implications. First, the proposed solution optimizes the efficiency in HEVs/PHEVs for any driver's commutes on average. Namely, being able to derive the optimal control policy online for any given trip, e.g., commute from point A to point B, remains still an open question. Second, the proposed solution uses the efficiency maps of each subsystem corresponding to their steady-state operation. Even though the supervisory controller in HEVs/PHEVs 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 could be to learn the transient operation [17], [18] associated with the driver's driving style and account for this.

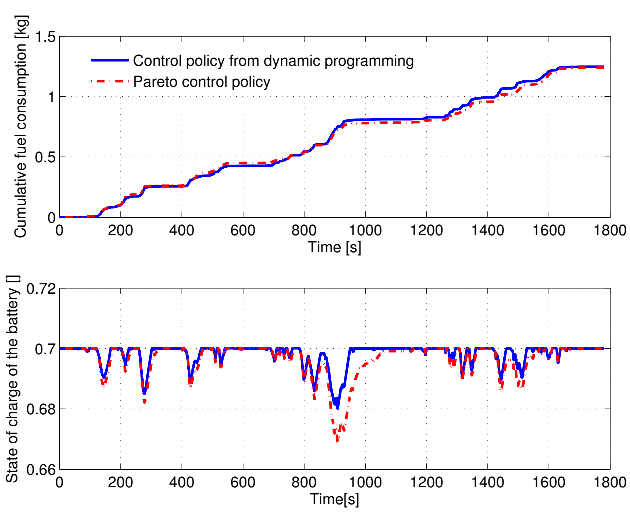

Fig. 2. Cumulative fuel consumption and state of charge of the battery for a parallel hybrid electric vehicle using the control policy derived from dynamic programming and the Pareto control policy over the city-suburban heavy duty vehicle route driving cycle.

V. Concluding Remarks

In this paper, we presented a framework for the analysis and stochastic optimization of the power management control problem in HEVs/PHEVs. We formulated the control problem as a multiobjective optimization problem of the one-stage expected costs of the subsystems and proved that the Pareto control policy minimizes the long-run expected average cost criterion of the system. The proposed solution reveals an operating point among the subsystems, e.g., engine, motor, generator, and battery, for all different values of the disturbance (driver), which is Pareto efficient. If all subsystems are operating at this operating point, then the long-run average cost of the HEV/PHEV is minimized.

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] M. Koot, J. Kessels, B. de Jager, W. Heemels, P. P. J. van den Bosch, and M. Steinbuch, "Energy management strategies for vehicular electric power systems," *IEEE Transactions on Vehicular Technology*, vol. 54, no. 3, pp. 771–782, May 2005.

[3] O. Sundstrom, L. Guzzella, and P. Soltic, "Torque-Assist Hybrid Electric Powertrain Sizing: From Optimal Control Towards a Sizing Law," *IEEE Transactions on Control Systems Technology*, vol. 18, no. 4, pp. 837–849, July 2010.

[4] 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.

[5] E. D. J. Tate, J. W. Grizzle, and H. Peng, "Shortest path stochastic control for hybrid electric vehicles," *International Journal of Robust and Nonlinear Control*, vol. 18, pp. 1409–1429, 2008.

[6] J. Liu and H. Peng, "Modeling and Control of a Power-Split Hybrid Vehicle," *IEEE Transactions on Control Systems Technology*, vol. 16, no. 6, pp. 1242–1251, 2008.

[7] 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.

[8] D. F. Opila, X. Wang, R. McGee, R. B. Gillespie, J. A. Cook, and J. W. Grizzle, "An Energy Management Controller to Optimally Trade Off Fuel Economy and Drivability for Hybrid Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 20, no. 6, pp. 1490–1505, 2011.

[9] S. J. Moura, H. K. Fathy, D. S. Callaway, and J. L. Stein, "A Stochastic Optimal Control Approach for Power Management in Plug-In Hybrid Electric Vehicles," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 3, pp. 545–555, 2011.

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

[11] 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.

[12] 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.

[13] 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.

[14] A. A. Malikopoulos, V. Maroulas, and J. Xiong, "Multiobjective optimization framework for stochastic control of complex systems," in *2015 American Control Conference*, 2015.

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

[16] A. A. Malikopoulos, "A multiobjective optimization framework for online optimal control in parallel hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, 2014.

[17] ——, *Real-Time, Self-Learning Identification and Stochastic Optimal Control of Advanced Powertrain Systems*. ProQuest, September 2011.

[18] 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.