IEEE | Transactions on Intelligent Transportation Systems

Ecological Adaptive Cruise Control and Energy Management Strategy for Hybrid Electric Vehicles based on Heuristic Dynamic Programming

Journal:	*Transactions on Intelligent Transportation Systems*
Manuscript ID	T-ITS-18-01-0097
Manuscript Type:	Regular Papers
Date Submitted by the Author:	31-Jan-2018
Complete List of Authors:	Li, Guoqiang; University of Kaiserslautern, Department of Electrical and Computer Engineering Görges, Daniel; University of Kaiserslautern, Department of Electrical and Computer Engineering
Keywords:	Automotive and Transportation Safety; Intelligent Systems; Neural Networks; ICC, Intelligent Transportation system (ITS), Learning control systems, Energy management, Hybrid Vehicles, Vehicle control; Driver assistance; Inter-vehicle communications; Automated highway systems
Abstract:	In this paper an ecological adaptive cruise controller (ECO-ACC) for hybrid electric vehicles (HEVs) in a car-following scenario is presented which jointly reduces the fuel consumption and improves the driving safety by maintaining a safe inter-vehicle distance. An adaptive cruise controller (ACC) based on action dependent heuristic dynamic programming (ADHDP) is proposed to realize an active distance control for safe and comfortable driving. ADHDP is able to adapt internal parameters online and can thus deal with systems with disturbances. Furthermore an online intelligent energy management strategy (IEMS) is introduced to control the gear shift and power split for fuel consumption optimization. The gear shift command is designed by enumeration and the power distribution between the engine and the electric motor is performed by ADHDP. The ACC and IEMS are finally combined to an ECO-ACC to achieve a multi-objective optimization. Only the current velocity and acceleration of the preceding vehicle are used while knowledge about the future velocity is not needed. Simulations for different driving cycles indicate that the ECO-ACC can not only ensure the desired inter-vehicle distance for safe and comfortable driving, but also lead to near-optimal fuel economy.

Ecological Adaptive Cruise Control and Energy Management Strategy for Hybrid Electric Vehicles based on Heuristic Dynamic Programming

Guoqiang Li and Daniel Görges, *Member*, *IEEE*

***Abstract*—In this paper an ecological adaptive cruise controller (ECO-ACC) for hybrid electric vehicles (HEVs) in a car-following scenario is presented which jointly reduces the fuel consumption and improves the driving safety by maintaining a safe inter-vehicle distance. An adaptive cruise controller (ACC) based on action dependent heuristic dynamic programming (ADHDP) is proposed to realize an active distance control for safe and comfortable driving. ADHDP is able to adapt internal parameters online and can thus deal with systems with disturbances. Furthermore an online intelligent energy management strategy (IEMS) is introduced to control the gear shift and power split for fuel consumption optimization. The gear shift command is designed by enumeration and the power distribution between the engine and the electric motor is performed by ADHDP. The ACC and IEMS are finally combined to an ECO-ACC to achieve a multi-objective optimization. Only the current velocity and acceleration of the preceding vehicle are used while knowledge about the future velocity is not needed. Simulations for different driving cycles indicate that the ECO-ACC can not only ensure the desired inter-vehicle distance for safe and comfortable driving, but also lead to near-optimal fuel economy.**

***Index Terms*—Adaptive cruise control, energy management, reinforcement learning**

I. Introduction

In recent years safe and comfortable driving has received considerable attention both in academia and industry. Studies indicate that human perception errors contribute to 90% of the traffic accidents [1] and are thus the most important factor for a safe driving. Advanced driving assistant systems (ADAS) can substantially reduce human perception errors and in this way improve the driving safety and also the driving comfort. Adaptive cruise control (ACC) is one of the most significant ADAS technologies that have been studied in recent years [2]. The ACC aims to maintain a host vehicle behind a preceding vehicle at a certain distance and in this way to improve the safety and also the comfort. Besides the road safety and driving comfort, the fuel consumption and pollutant emissions have received significant attention during the last years. Hybrid electric vehicles (HEVs) have a high potential to reduce the consumption and emissions through mutual usage of fuel and electricity [3]. The implementation of ACC in HEVs is promising to simultaneously improve the overall fuel consumption, the vehicle safety, and the driving comfort.

The first author is supported by the China Scholarship Council (CSC).

The authors are with the Juniorprofessorship for Electromobility, Dept. of Electrical and Computer Engineering, University of Kaiserslautern, 67663 Kaiserslautern, Germany (e-mail:gli@eit.uni-kl.de; goerges@eit.uni-kl.de).

A. Adaptive cruise control

Several control methods have been proposed for ACC in conventional vehicles to address the distance control and fuel economy, partly with adaptability to individual driver characteristics. In order to satisfy multiple objectives, model predictive control is proposed in [4] for optimal control in terms of fuel economy and tracking capability. A conditional linear Gaussian model for future velocity prediction of the preceding vehicle is used in stochastic model predictive control in [5]. A periodic servo-loop longitudinal control algorithm is presented in [6] to minimize the fuel consumption in a car-following scenario. Most of the control methods reported in the literature so far rely on a precise longitudinal vehicle model. Furthermore the optimal gear shift control for fuel economy is not covered yet. Recently learning-based control methods have been studied for ACC without relying on knowledge of the vehicle model and future velocity. A reinforcement learning-based cooperative adaptive cruise control method is proposed in [7] to maintain a safe distance in a stop-and-go scenario. In [8] a supervised adaptive dynamic programming algorithm for full-range adaptive cruise control is introduced. Both methods utilize a discrete control input and reward during learning which can deteriorate the performance in a real continuous driving environment.

B. Energy management in HEVs

Various control algorithms have been developed for the energy management in HEVs with focus on fuel consumption optimization [9]. Dynamic programming (DP) has been successfully applied to design a power split distribution for an optimal fuel economy under a given driving cycle and is usually regarded as the test bench [10], [11]. The equivalent consumption minimization strategy (ECMS) and Pontryagin's minimum principle (PMP) are proposed to control the power distribution online for overall fuel economy of HEVs [12], [13]. Model predictive control is used to reduce the calculation burden for real-time optimal energy management [14]. Convex optimization have been studied to jointly optimize the component size and fuel economy [15], [16]. The control methods indicated above strongly rely on a system model with accurate parameters, particularly on a precise approximation of the nonlinear efficiency maps of the engine and the electric motor. The approximations and parameter inaccuracies may cause a deviation from the optimal solution in a real system. Several intelligent control approaches such as neural

networks and Q-learning have been proposed to design the power split online [17], [18]. The neural network controller as a supervised learning method requires extensive offline training. An online adaption in an uncertain environment is difficult. The Q-learning method is suitable for problems with discrete variables, but has limitations for power split control in HEVs with continuous variables. Due to the discrete input for gear shift control, an optimal online shift strategy is still not available. Most approaches use a predetermined gear shift map instead. The gear position determines the instantaneous power transfer to the driveline and thus influences the fuel consumption of the engine. For online gear shift control in HEVs, in [16] an energy management strategy combining a gear shift design based on DP and a power distribution with convex optimization is proposed. A similar control algorithm integrating DP and PMP is presented in [19] to control the gear shift and power split between the engine and electric motor. However, both methods utilize an independent control of the gear shift with DP to reduce the calculation burden. Therefore a prior knowledge of the driving cycle is required which hinders a real-time adaptation.

C. Implementation of ACC in HEVs

Currently most of the studies focus on ACC in conventional vehicles. Only few papers address HEVs equipped with ACC for both fuel economy and driving safety. A multi-objective decoupling architecture for ACC in HEVs is proposed in [20] to improve the fuel economy and ride comfort based on a rule-based energy management algorithm. The resulting rule-based strategy can, however, not guarantee the global fuel optimality and the gear shift control is predetermined without online optimization. A multistep offline dynamic programming with an online lookup table is presented for ACC in HEVs in [21]. The optimization must be performed in advance which prevents an adaptation to the driving condition. In [22] a multi-objective optimal control method for ACC in HEVs is studied. An integration of ACC and an optimal energy management strategy particularly with respect to the online optimal gear shift control for fuel economy is, however, not considered.

D. Contributions and organization

Overall an integration of the energy management strategy and the ACC in HEVs to simultaneously optimize the gear shift and power split for fuel economy and to adaptively control the inter-vehicle distance for safe and comfortable driving has not been studied yet. Optimal energy management strategies usually focus on a single vehicle without considering ACC. Moreover an online gear shift optimization for fuel economy which can be applied in real-time without driving cycle knowledge and large computation burden is still open. This paper contributes to closing these gaps.

The objective of this paper is to propose an ecological adaptive cruise controller (ECO-ACC) for HEVs in a car-following scenario to improve the fuel economy and driving safety. An adaptive cruise control method is derived based on action-dependent heuristic dynamic programming (ADHDP) to ensure the desired velocity tracking and the relative distance. The gear shift command is generated by enumeration over a finite set and the power split between the engine and the electric motor is performed using ADHDP. The ECO-ACC has two major properties. First, the ADHDP can adapt the parameters online which improves the velocity tracking performance in the car-following scenario under disturbances. Second, the online gear shift and power split control for optimal fuel economy of HEVs is realized without prior knowledge of the driving cycle and system model. A comparison with dynamic programming (DP) indicates that the ECO-ACC can achieve near-optimal fuel economy and does not impose much calculation and storage burden. Different driving cycles are tested to validate the robustness.

The paper is organized as follows. In Section II heuristic dynamic programming and the online learning algorithm are introduced. The ACC model and the control strategy are presented in Section III. The online gear shift and energy management in HEVs are discussed in Section IV. Section V addresses the control algorithm for ECO-ACC and provides simulation results. Conclusion are finally given in Section VI.

II. Heuristic Dynamic Programming

A. Heuristic dynamic programming

Heuristic dynamic programming, also denoted as approximate dynamic programming (ADP) or reinforcement learning (RL), is an effective method to let an agent learn how to make optimal decisions by trial-and-error interactions with a dynamic environment [23]. The agent is naive at the beginning and it learns to make decisions to maximize or minimize a cumulative reward from the environment over time. Consider a nonlinear discrete-time system described by

$$\boldsymbol{x}_{k+1} = f(\boldsymbol{x}_k, \boldsymbol{u}_k), \quad k = 0, 1, 2, ... \tag{1}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state vector, $\boldsymbol{u} \in \mathbb{R}^m$ is the input vector and $f(\boldsymbol{x}_k, \boldsymbol{u}_k)$ is the system dynamics function. At each time step k, the agent receives an instantaneous reward $r(\boldsymbol{x}_k, \boldsymbol{u}_k)$ depending on the state and input and generates an action to maximize or minimize a value function. The agent can adjust its parameters over time to optimize the performance. Note that the system model (1) is not required.

The value function for a discrete-time system over an infinite horizon can be defined as

$$\begin{aligned} J(\boldsymbol{x}_k) =& \sum_{i=k}^{\infty} \alpha^{i-k} r(\boldsymbol{x}_i, \boldsymbol{u}_i) \\ =& r(\boldsymbol{x}_k, \boldsymbol{u}_k) + \alpha \sum_{i=k+1}^{\infty} \alpha^{i-(k+1)} r(\boldsymbol{x}_i, \boldsymbol{u}_i) \\ =& r(\boldsymbol{x}_k, \boldsymbol{u}_k) + \alpha J(\boldsymbol{x}_{k+1}) \end{aligned} \tag{2}$$

where α is a discount factor with $0 < \alpha < 1$ and $r(\boldsymbol{x}_i, \boldsymbol{u}_i)$ is the instantaneous reward depending on the state vector $\boldsymbol{x}_i$ and input vector $\boldsymbol{u}_i$.

The optimal control policy $\boldsymbol{u}_k^* = h(\boldsymbol{x}_k)$ for the value function (2) can be derived by solving the equation

$$J^*(\boldsymbol{x}_k) = \min_{h(.)} \left[\alpha J^*(\boldsymbol{x}_{k+1}) + r(\boldsymbol{x}_k, \boldsymbol{u}_k^*)\right] \tag{3}$$

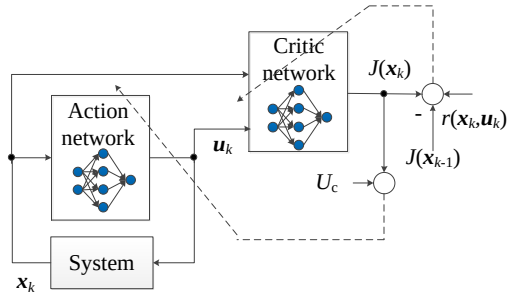


Fig. 1. Actor-critic structure

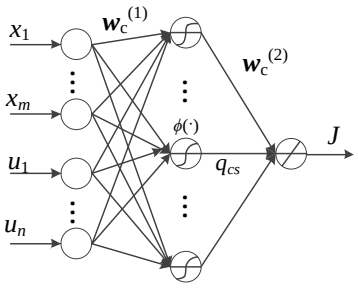


Fig. 2. Structure of the critic network

where $J^*(\boldsymbol{x}_k)$ is the optimal value function for the current state $\boldsymbol{x}_k$, $J^*(\boldsymbol{x}_{k+1})$ is the optimal value function for the next state $\boldsymbol{x}_{k+1}$, and $r(\boldsymbol{x}_k, \boldsymbol{u}_k^*)$ is the immediate cost for $\boldsymbol{u}_k^*$ from $\boldsymbol{x}_k$ to $\boldsymbol{x}_{k+1}$.

This difference equation is denoted as Bellman optimality equation or Hamilton-Jacobi-Bellman (HJB) equation and can be solved by dynamic programming, usually after discretizing the state and input space. A discrete state and input space results in an exponential growth of the computation time and memory demand known as the curse of dimensionality, preventing a real-time application. The Bellman optimality equation can be solved approximately based on approximate value or policy iteration, avoiding the curse of dimensionality and thus allowing a real-time implementation. In the following parts the policy iteration using an actor-critic structure for solving the Bellman optimality equation will be discussed.

B. Actor-critic structure

For continuous-state problems, the exact value of the cost $J^*(\boldsymbol{x}_k)$ in (3) can be approximated by the output of a critic neural network and the control policy can be derived from an action neural network. The learning process is realized with the actor-critic structure shown in Fig. 1, which is also referred to as action dependent heuristic dynamic programming (ADHDP). Through trial-and-error adaptation with the dynamic environment, the optimal control policy and the optimal value function can be learned with the action and critic network. Widely applied value function approximations in practical applications are neural networks, linear basis functions and piecewise linear functions [24]. In this work, a multi-layer network is adopted as function approximation.

C. Critic network and online learning

The critic network is used to approximate the value function which is a nonlinear mapping of the input variable consisting of the m-dimensional state vector $\boldsymbol{x}_k$, the n-dimensional input (action) vector $\boldsymbol{u}_k$ and the reward $r(\boldsymbol{x}_k, \boldsymbol{u}_k)$ resulting from the current action as the reinforcement signal. In Fig. 2 the structure of the critic network for ADHDP is shown consisting of a nonlinear multilayer feedforward network with $m+n$ input neurons, N_{ch} neurons in the hidden layer and one output neuron. The output of the critic network can be derived from

$$\begin{aligned}
\hat{J}(\boldsymbol{x}_k) &= \boldsymbol{w}_{\mathrm{c}}^{(2)}(k)\boldsymbol{q}_{\mathrm{c}}(k) = \sum_{s=1}^{N_{\mathrm{ch}}} w_{\mathrm{cs}}^{(2)}(k) q_{cs}(k), \\
q_{cs}(k) &= \phi(\sigma_{cs}(k)) = \frac{1-e^{-\sigma_{cs}(k)}}{1+e^{-\sigma_{cs}(k)}}, s = 1, 2, \ldots, N_{\mathrm{ch}} \\
\sigma_{cs}(k) &= \sum_{i=1}^{m} w_{csi}^{(1)}(k) x_{ik} + \sum_{j=1}^{n} w_{csj}^{(1)}(k) u_{jk}.
\end{aligned} \tag{4}$$

where $\boldsymbol{w}_{\mathrm{c}}^{(1)}(k)$ is the weighting matrix from the input to the hidden neurons, $\sigma_{cs}(k)$ is the weighted sum of all inputs to one hidden neuron s, $q_{cs}(k)$ is the output value of each hidden neuron resulting from the hyperbolic tangent transfer function $\phi(\sigma_{cs}(k))$ and the weighting vector from the hidden layer to the output neuron is denoted as $\boldsymbol{w}_{\mathrm{c}}^{(2)}$.

The error function of the critic network can be defined as the temporal difference between the approximation and the actual value as

$$e_{\mathrm{c}}(k) = \hat{J}(\boldsymbol{x}_k) - \left[\hat{J}(\boldsymbol{x}_{k-1}) - r(\boldsymbol{x}_k, \boldsymbol{u}_k)\right] \tag{5}$$

In order to satisfy (3) for optimal control, the error function $e_{\mathrm{c}}(k)$ should be brought to 0. The learning objective for the critic network is therefore to minimize the error $e_{\mathrm{c}}(k)$ by updating the parameters $\boldsymbol{w}_{\mathrm{c}}$, i.e.

$$\min_{\boldsymbol{w}_c} E_c(k) = \min_{\boldsymbol{w}_c} \frac{1}{2} e_c^2(k). \tag{6}$$

A gradient descent adaptation algorithm with chain derivation rule is used to update the weights. The adaption of weights results from

$$\begin{aligned}
\boldsymbol{w}_{\mathrm{c}}(k+1) &= \boldsymbol{w}_{\mathrm{c}}(k) + \Delta\boldsymbol{w}_{\mathrm{c}}(k) \\
\Delta\boldsymbol{w}_{\mathrm{c}}(k) &= \eta_{\mathrm{c}}(k)\left[-\frac{\partial E_{\mathrm{c}}(k)}{\partial \boldsymbol{w}_{\mathrm{c}}(k)}\right] \\
\frac{\partial E_{\mathrm{c}}(k)}{\partial \boldsymbol{w}_{\mathrm{c}}(k)} &= \frac{\partial E_{\mathrm{c}}(k)}{\partial e_{\mathrm{c}}(k)} \frac{\partial e_{\mathrm{c}}(k)}{\partial \hat{J}(\boldsymbol{x}_k)} \frac{\partial \hat{J}(\boldsymbol{x}_k)}{\partial \boldsymbol{w}_{\mathrm{c}}(k)}
\end{aligned} \tag{7}$$

where $\eta_{\mathrm{c}}(k)$ is the learning rate of the critic network.

D. Action network and online learning

In this paper a three-layer network is used for the action network whose input is the state vector $\boldsymbol{x}_k$ and whose output is the control action $\boldsymbol{u}_k$ as shown in Fig. 3. The hidden layer contains N_{ah} neurons with hyperbolic tangent transfer

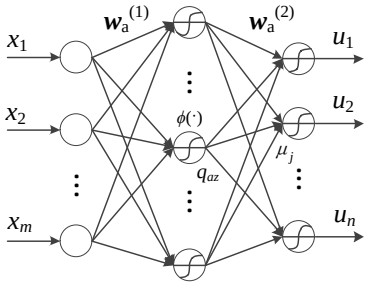


Fig. 3. Structure of the action network

function. The equation of the action network can be stated as

$$\begin{aligned} u_j(k) &= \frac{1-e^{-\mu_j(k)}}{1+e^{-\mu_j(k)}}, j=1,2,\ldots,n \\ \mu_j(k) &= \boldsymbol{w}_{\mathrm{a}}^{(2)}(k)\boldsymbol{q}_{\mathrm{a}}(k) = \sum_{z=1}^{N_{\mathrm{ah}}} w_{\mathrm{a}z}^{(2)}(k) q_{\mathrm{a}z}(k) \\ q_{\mathrm{a}z}(k) &= \phi(\sigma_{\mathrm{a}z}(k)) = \frac{1-e^{-\sigma_{\mathrm{a}z}(k)}}{1+e^{-\sigma_{\mathrm{a}z}(k)}}, z=1,2,\ldots,N_{\mathrm{ah}} \\ \sigma_{\mathrm{a}z}(k) &= \sum_{i=1}^{m} w_{\mathrm{a}i}^{(1)}(k) x_{ik} \end{aligned} \tag{8}$$

where $\mu_j(k)$ is the input to each action node, $q_{\mathrm{a}z}(k)$ is the output of the hidden neuron, $\sigma_{\mathrm{a}z}(k)$ is the input value to each hidden neuron, $\boldsymbol{w}_{\mathrm{a}}^{(1)}(k)$ is the weighting matrix from the input neurons to the hidden neurons, and the weighting vector from the hidden neurons to the output neurons is $\boldsymbol{w}_{\mathrm{a}}^{(2)}(k)$.

The action network generates the approximated optimal policy $\boldsymbol{u}_k^* = h(\boldsymbol{x}_k)$ by minimizing the cost $\hat{J}(\boldsymbol{x}_k)$, i.e.

$$\boldsymbol{u}^* = \arg\min_{h(.)} \hat{J}(\boldsymbol{x}_k). \tag{9}$$

With the system states acting as the input variables, the optimal control action can be generated by the action network through adapting the weights $\boldsymbol{w}_{\mathrm{a}}$ so as to minimize $\hat{J}$. The learning objective can thus be formulated as

$$\min_{\boldsymbol{w}_{\mathrm{a}}} E_{\mathrm{a}}(k) = \min_{\boldsymbol{w}_{\mathrm{a}}} \hat{J}(\boldsymbol{x}_k). \tag{10}$$

The training of the action network is similar to the one used in the citric network. With a gradient descent rule, the weights can be adapted using

$$\begin{aligned} \boldsymbol{w}_{\mathrm{a}}(k+1) &= \boldsymbol{w}_{\mathrm{a}}(k) + \Delta\boldsymbol{w}_{\mathrm{a}}(k) \\ \Delta\boldsymbol{w}_{\mathrm{a}}(k) &= \eta_{\mathrm{a}}(k)\left[-\frac{\partial E_{\mathrm{a}}(k)}{\partial \boldsymbol{w}_{\mathrm{a}}(k)}\right] \\ \frac{\partial E_{\mathrm{a}}(k)}{\partial \boldsymbol{w}_{\mathrm{a}}(k)} &= \frac{\partial \hat{J}(\boldsymbol{x}_k)}{\partial \boldsymbol{u}_k}\frac{\partial \boldsymbol{u}_k}{\partial \boldsymbol{\mu}_k}\frac{\partial \boldsymbol{\mu}_k}{\partial \boldsymbol{w}_{\mathrm{a}}(k)} \end{aligned} \tag{11}$$

where $\eta_{\mathrm{a}}(k)$ is the learning rate of the action network.

III. ACC Model and Control Design

In this section, an adaptive cruise controller for maintaining a desired inter-vehicle distance by adjusting the velocity of the host vehicle for any given driving condition is presented.

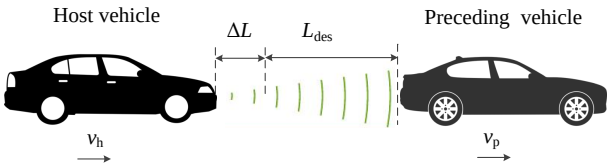


Fig. 4. Car-following scenario

A. Dynamic car-following model

A typical car-following scenario consisting of a host vehicle and a preceding vehicle is presented in Fig. 4. x_{h}, v_{h}, and a_{h} are the position, velocity, and acceleration of the host vehicle. x_{p}, v_{p}, and a_{p} are the position, velocity, and acceleration of the preceding vehicle. The nonlinear longitudinal dynamics of the host vehicle is described by

$$\begin{aligned} \dot{x}_{\mathrm{h}} &= v_{\mathrm{h}} \\ \dot{v}_{\mathrm{h}} &= a_{\mathrm{h}} \\ a_{\mathrm{h}} &= \frac{F_{\mathrm{trac}}}{m} - \frac{\rho A c_{\mathrm{d}} v_{\mathrm{h}}^2}{2m} - gf\cos\beta + g\sin\beta \end{aligned} \tag{12}$$

where F_{trac} is the traction force at the wheels, ρ is the air density, A is the equivalent area of the vehicle body, c_{d} is the aerodynamic resistance coefficient, m is the vehicle mass, f is the rolling resistance coefficient, and β is the road angle. The resistance forces are simplified to the aerodynamic, rolling and grading resistance force.

The relative distance deviation is $\Delta L = x_{\mathrm{p}} - x_{\mathrm{h}} - L_{\mathrm{des}}$ where L_{des} denotes the driver's desired inter-vehicle distance. The relative velocity deviation results from $\Delta v = v_{\mathrm{p}} - v_{\mathrm{h}}$. To ensure safe driving, the constant time headway spacing policy

$$L_{\mathrm{des}} = \tau_{\mathrm{h}} v_{\mathrm{h}} + d_0 \tag{13}$$

is employed where τ_{h} is the nominal time headway and d_0 is the standstill distance. It can be seen that L_{des} changes with v_{h}, indicating that a long relative distance is required to guarantee driving safety at a high velocity of the host vehicle while during low-speed driving a short inter-vehicle distance is desirable to improve the roadway capacity.

The dynamic models of the inter-vehicle distance deviation ΔL and relative velocity deviation Δv are given by

$$\begin{aligned} \dot{\Delta} L &= v_{\mathrm{p}} - v_{\mathrm{h}} - \tau_{\mathrm{h}} a_{\mathrm{h}} \\ \dot{\Delta} v &= a_{\mathrm{p}} - a_{\mathrm{h}} \end{aligned} \tag{14}$$

B. Control problem formulation for ACC

The objective of the ACC is to control the velocity of the host vehicle by tracking the preceding vehicle with a desired inter-vehicle distance for safe and comfortable driving. The tracking performance can be evaluated based on the distance deviation ΔL and velocity deviation Δv. By tracking the velocity of the preceding vehicle, frequent vehicle cut-ins from adjacent lanes can be avoided and the probability of traffic accidents can be reduced. Hence the tracking error ΔL and Δv should be minimized for driving safety. In this paper driving safety is understood as keeping a desired relative distance

to the preceding vehicle in normal driving situations while emergency braking conditions are not considered. Emergency braking conditions must, as in conventional ACC, be handled by an assistive emergency braking system.

The fuel consumption is mainly determined by the vehicle traction force as discussed in [4]. In a car-following situation, the fuel consumption increases during acceleration. Any unnecessary acceleration will cause an unnecessary braking to maintain the desired inter-vehicle distance for safety and in turn lead to a higher fuel cost. For HEVs the braking energy can only be partly regenerated because of the power loss in the transmission, the generator and the battery. For conventional vehicles the braking force applied on the wheels generates friction heat and is completely dissipated into the surrounding without regeneration. Therefore the braking power should be kept as low as possible. A penalty on the acceleration and in turn on traction force can thus have an effect on the fuel consumption.

From this perspective the optimization problem is to minimize ΔL and Δv through controlling the traction force of the host vehicle, yielding the cost

$$r = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x} + Ru^2, \tag{15}$$

where $\boldsymbol{x} = [\Delta L, \Delta v]^{\mathrm{T}}$ is the state vector, $u = F_{\mathrm{trac}}/m$ is the control input corresponding to the traction or braking force applied on the wheels per unit mass, and $\boldsymbol{Q}$ and R are weighting factors. In order to ensure a safe and comfortable driving experience, the following tracking constraints are introduced:

$$\begin{aligned} a_{\min} \le a_{\mathrm{h}} \le a_{\max} \\ \Delta L_{\min} \le \Delta L \le \Delta L_{\max} \\ \Delta v_{\min} \le \Delta v \le \Delta v_{\max}. \end{aligned} \tag{16}$$

C. Control algorithm and simulation

ADHDP is adopted to realize the online adaptive cruise control. The state vector $\boldsymbol{x}$ is the input of the action network. Its output is the control variable u. The critic network approximates the accumulated cost during tracking. The weighting parameters of the critic and action network are initialized randomly from [0 0.1]. The learning rates are defined as $\eta_{\mathrm{a}} = 5 \cdot 10^{-5}$ and $\eta_{\mathrm{c}} = 1 \cdot 10^{-3}$. The numbers of hidden layers of the critic and action network are $N_{\mathrm{ah}} = N_{\mathrm{ch}} = 20$. In each time step the optimal control policy is updated by adapting the weights of the critic and action network using (7) and (11) based on the reward r defined in (15). The adaption of the weighting parameters is performed with two stopping criteria: the maximum iteration numbers $n_{\mathrm{a}} = n_{\mathrm{c}} = 20$ and the approximation error tolerance $T_{\mathrm{a}} = 10^{-8}, T_{\mathrm{c}} = 10^{-6}$. If any of the two criteria is satisfied, the weight adaptation is stopped and the optimal control policy is obtained from the action network. The sampling interval is chosen as $0.1\,\mathrm{s}$.

The proposed control method is applied for the Worldwide Harmonized Light Vehicle Test Cycle (WLTC). The simulation results are shown in Fig. 5. The host vehicle can follow the velocity of the preceding vehicle quite well. The acceleration trajectory of the host vehicle is very close with that of the preceding vehicle. The very minor difference of the accelerations does not influence the velocity tracking performance and

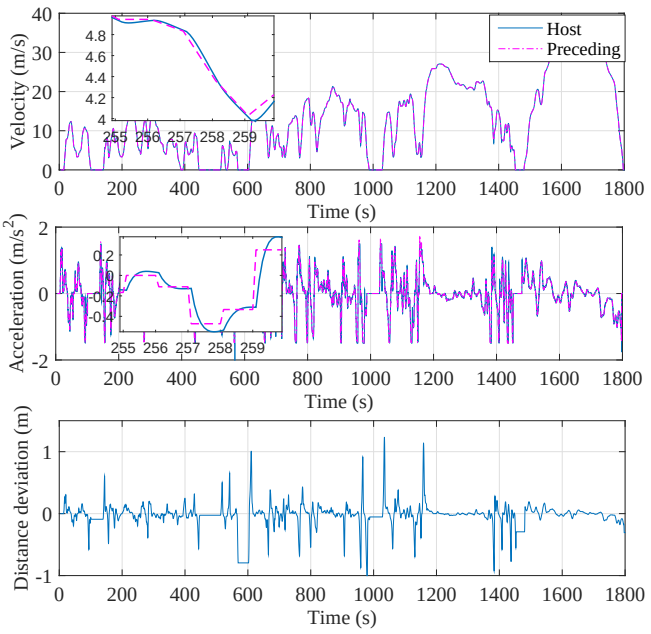


Fig. 5. Simulation results for WLTC

the distance deviation always remains in a narrow band. The magnitude of the acceleration is under $2\,\mathrm{m/s^2}$, underlining good driving comfort. Furthermore, the inter-vehicle distance deviation does not exceed 1.5 m which indicates a safe driving.

Since the control policy from the action network is derived based on actor-critic reinforcement learning, a system model is not required and the control policy can be continuously adapted, which yields robustness with respect to uncertainties and disturbances. In practice the measurement of the velocity and inter-vehicle distance can be disturbed by sensor noise. Additionally the vehicle parameters and driving situations may change over time. For further analysis it is assumed that all disturbances and sensor noise are lumped into a stochastic noise. Simulation results for a random noise acting on the velocity deviation measurement with 5% uniform distribution, i.e. $\Delta v := (1+0.05\mathcal{U}(0,1))\Delta v$, are shown in Fig. 6. The ACC control enables the host vehicle to track the preceding vehicle with the similar velocity profile. The distance deviation is restricted within $\pm 2\,\mathrm{m}$ and the range of the velocity deviation is between -4 m/s and 4 m/s, underlining the safe driving. Simulation results for a random noise with 5% uniform distribution acting on the inter-vehicle distance deviation measurement, i.e. $\Delta L := (1+0.05\mathcal{U}(0,1))\Delta L$ are given in Fig. 7. Similarly the velocity profile of the host vehicle is close to that of the preceding vehicle. The maximum distance deviation is 1.2 m while the velocity deviation varies from -5 m/s to 5 m/s. Therefore the ADHDP based ACC is adaptive in terms of sensor noise and guarantee the driving safety.

IV. Energy Management Strategy for HEVs

The host vehicle studied in this paper is a parallel HEV consisting of a diesel engine, an electric motor and a 5-speed gearbox. Main parameters are listed in Table I. In this section the online gear shift and power split control are presented. The

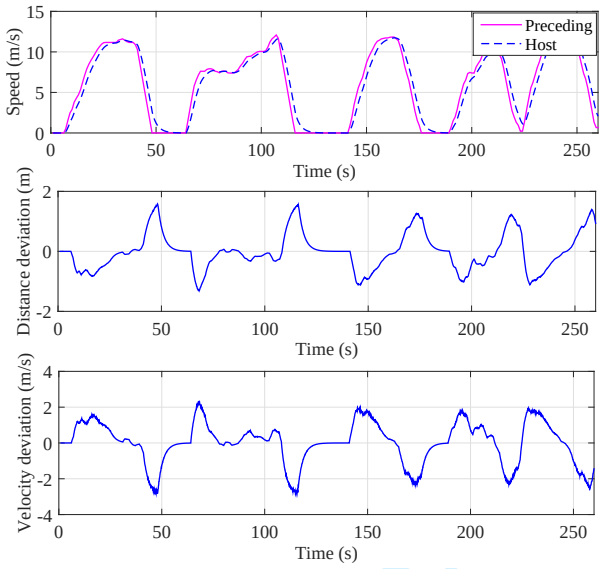


Fig. 6. Simulation results for velocity noise

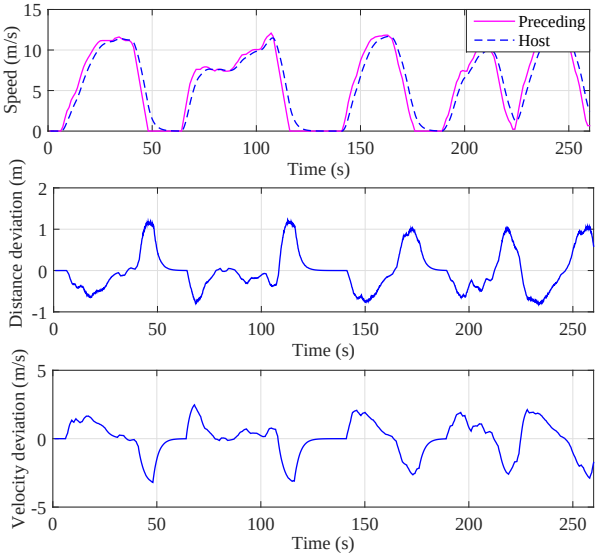


Fig. 7. Simulations results for inter-vehicle distance noise

sampling interval is chosen as 1 s assuming that the gear shift finishes within this sampling interval for smoothness.

A. Modeling of the HEV

The power for driving is provided by the engine and the electric motor and therefore described by

$$P_{\text{in}} = P_{\text{e}} + P_{\text{m}} \tag{17}$$

where P_{in} is the power delivered into the transmission, P_{e} is the power produced by the engine, and P_{m} is the power generated by the electric motor. The power demand P_{dem} during driving can be calculated from

$$P_{\text{dem}} = F_{\text{load}} v_{\text{h}} = \left(\frac{\rho A c_{\text{d}} v_{\text{h}}^2}{2} + mgf\cos\beta + mg\sin\beta + ma_{\text{h}}\right) v_{\text{h}} \tag{18}$$

TABLE I
PARAMETERS OF THE PARALLEL HEV

Vehicle mass	1500 kg
Maximum engine power	60 kW
Maximum engine torque	187 Nm
Maximum motor power	25 kW
Maximum motor torque	160 Nm
Transmission ratios	9.64, 6.08, 4.21, 3.07, 2.33
Battery energy	1.485 kWh

where F_{load} is the load force at the wheels resulting from the aerodynamic, rolling, grading and inertial resistance force.

The power of the battery can be described with a equivalent circuit model by

$$P_{\text{b}} = V_{\text{oc}} I - I^2 R \tag{19}$$

where V_{oc} is the open circuit voltage, I is the battery current, and R is the battery internal resistance with

$$R = \begin{cases} R_{\text{ch}} & \text{for } I \geq 0 \\ R_{\text{dis}} & \text{for } I < 0 \end{cases} \tag{20}$$

where R_{ch} and R_{dis} are internal resistances during charging and discharging which have a nonlinear relationship with the state of charge (SOC).

With (19) the dynamic model of the SOC can be stated as

$$\dot{SOC} = -\frac{I}{Q} = -\frac{V_{\text{oc}} - \sqrt{V_{\text{oc}}^2 - 4RP_{\text{b}}}}{2RQ} \tag{21}$$

where Q is the battery capacity.

B. Control problem formulation

The gear position determines the torque T_{in} and rotational speed ω_{in} of the input shaft, i.e.

$$\begin{aligned} \omega_{\text{in}} &= \frac{v}{r_{\text{w}}} i_{\text{g}}(g) \\ T_{\text{in}} &= \frac{T_{\text{load}}}{i_{\text{g}}(g)} = \frac{F_{\text{load}} r_{\text{w}}}{i_{\text{g}}(g)} \end{aligned} \tag{22}$$

where i_{g} is the transmission ratio resulting from the gear position g and r_{w} is the wheel radius. According to the parallel HEV structure, the input shaft torque consists of the engine torque T_{e} and the electric motor torque T_{m}. Therefore the engine working points can be adjusted by the gear shift to influence the fuel consumption. Assuming that u_{p} is the power distribution coefficient of the input shaft torque with range of $-1 < u_{\text{p}} \leq 1$, the torque provided by the engine and the electric motor can be expressed as

$$\begin{aligned} T_{\text{e}} &= T_{\text{in}}\,(1 - u_{\text{p}}) \quad (T_{\text{in}} > 0) \\ T_{\text{m}} &= T_{\text{in}} u_{\text{p}} \end{aligned} \tag{23}$$

The gear shift strategy can be described based on the current gear position $g(k)$, previous gear position $g(k-1)$ and the gear shift command $u_{\text{g}}(k)$ by the dynamic model

$$g(k) = g(k-1) + u_{\text{g}}(k). \tag{24}$$

Assuming that only a sequential gear shift is allowed, u_{g} belongs to the set $\{-1, 0, 1\}$ where -1 means downshift, 1 represents upshift, and 0 corresponds to sustainment.

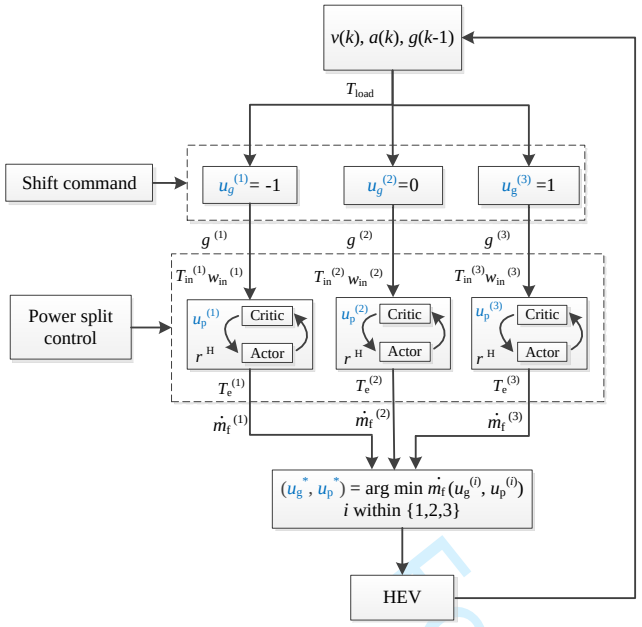


Fig. 8. Online control algorithm for gear shift and power split

The objective of the energy management strategy is to design an optimal control law $\boldsymbol{u} = [u_\mathrm{p}, u_\mathrm{g}]^\mathrm{T}$ for the power split and gear shift which reduces the total fuel consumption during the whole driving cycle, i.e. minimizes the cost function

$$J = \sum_{k=0}^{T_\mathrm{cyc}-1} \dot{m}_\mathrm{f}(k) \tag{25}$$

where $\dot{m}_\mathrm{f}(k)$ is the fuel mass flow rate in g/s which depends on the engine torque T_e and the rotational speed ω_e and T_cyc is the length of the driving cycle.

C. Online control algorithm

Since the control input of the HEV contains a discrete gear shift input and a continuous power split input, the optimal solution is usually derived with DP. However, for DP the state and control variables must be discretized which leads to an exponential growth of the computation and memory demand and prevents a real-time calculation. In this paper, an intelligent energy management strategy (IEMS) based on ADHDP is used to realize the gear shift and power split control. The control algorithm is shown in Fig. 8. It is assumed that the velocity $v(k)$ and acceleration $a(k)$ at the current time k are available. At each time step k the gear shift command $u_\mathrm{g}(k)$ is chosen from the set $\{-1, 0, 1\}$. The torque $T_\mathrm{in}(k)$ and the rotational speed $\omega_\mathrm{in}(k)$ of the input shaft for each gear shift command $u_\mathrm{g}(k)$ are obtained by combining (22) and (24) from

$$\begin{aligned} \omega_\mathrm{in}(u_\mathrm{g}(k)) &= \frac{v(k)}{r} i_\mathrm{g}\left(g(k-1) + u_\mathrm{g}(k)\right) \\ T_\mathrm{in}(u_\mathrm{g}(k)) &= \frac{F_\mathrm{load}(k) r_\mathrm{w}}{i_\mathrm{g}\left(g(k-1) + u_\mathrm{g}(k)\right)} \end{aligned} \tag{26}$$

The power split $u_\mathrm{p}(k)$ is then determined for each gear shift command $u_\mathrm{g}(k)$ from the actor-critic structure with ADHDP where the deviation between the actual $SOC(k)$ and the reference SOC_r is used as the state variable and the power split $u_\mathrm{p}(k)$ is employed as the control variable. The action network generates the power split between the engine torque $T_\mathrm{e}(u_\mathrm{g}(k))$ and the motor torque $T_\mathrm{m}(u_\mathrm{g}(k))$ as the optimal control input to the HEV and observes the resulting fuel mass flow rate $\dot{m}_\mathrm{f}\left(u_\mathrm{g}(k), u_\mathrm{p}(k)\right)$. The reinforcement signal is chosen as

$$r^\mathrm{H}(k) = \dot{m}_\mathrm{f}(k) + \gamma\left(SOC(k) - SOC_\mathrm{r}\right)^2 \tag{27}$$

to reduce the fuel consumption and ensure the charge sustainment where γ is a weighting coefficient. Finally the optimal gear shift command $u_\mathrm{g}^*(k)$ and power split command $u_\mathrm{p}^*(k)$ yielding minimum consumption are obtained from

$$\left(u_\mathrm{g}^*(k), u_\mathrm{p}^*(k)\right) = \arg \min_{u_\mathrm{g}^*(k)\in\{-1,0,1\}} \dot{m}_\mathrm{f}\left(u_\mathrm{g}(k), u_\mathrm{p}(k)\right). \tag{28}$$

The optimal power split generated from the action network is obtained by adapting the weighting parameters. The initial weighting parameters of the critic and action network are chosen randomly from [-0.2 0.2]. The learning rates are selected as $\eta_\mathrm{a}^\mathrm{H} = \eta_\mathrm{c}^\mathrm{H} = 0.03$. The numbers of neurons in the hidden layers are $N_\mathrm{ah}^\mathrm{H} = N_\mathrm{ch}^\mathrm{H} = 30$. At each time step, the weighting parameters of the critic and action networks are adapted with the backpropagation algorithm in (7) and (11). The stopping criteria for the weights adaption are parameterized by the maximum cycle numbers $n_\mathrm{a}^\mathrm{H} = 1500$, $n_\mathrm{c}^\mathrm{H} = 3000$ and the error tolerance $T_\mathrm{a}^\mathrm{H} = T_\mathrm{c}^\mathrm{H} = 10^{-6}$. If any of the two criteria is satisfied, the iteration is terminated and the optimal power split $u_p{}^*(k)$ is derived from the action network. The engine torque is then used to calculate the fuel consumption $\dot{m}_\mathrm{f}(u_p{}^*(k), u_g(k))$. By searching for the shift command $u_g(k)$ yielding the minimum fuel consumption, the optimal shift control $u_g{}^*(k)$ can be derived.

D. Simulation analysis

In this part the gear shift and power split control resulting from the IEMS for different driving cycles are studied to assess the optimality and robustness.

1) Analysis of optimality: The IEMS is first evaluated for the FTP-75 cycle. Fig. 9 and Table II contain the simulation results and a comparison with DP. Here DP is applied to calculate the optimal solution as a benchmark. It can be seen that the gear shift sequence is close to the one resulting from DP with only a few mismatches. Since DP is calculated with the whole driving cycle information, the gear shift sequence is adjusted to minimize the global fuel consumption. However, the gear shift command from the developed IEMS is derived based on the instantaneous fuel cost without the knowledge of the future driving information, yielding a sub-optimal solution. The power of the engine and the electric motor determined with IEMS and DP have the similar profiles. The final SOC value is regulated close to the initial value with the charge sustainment term in the cost function (27). The overall fuel consumption increases by 1% which underlines the near-optimality of the energy management strategy. The difference in the SOC trajectories indicates that there are different solutions yielding similar cost close to the minimum

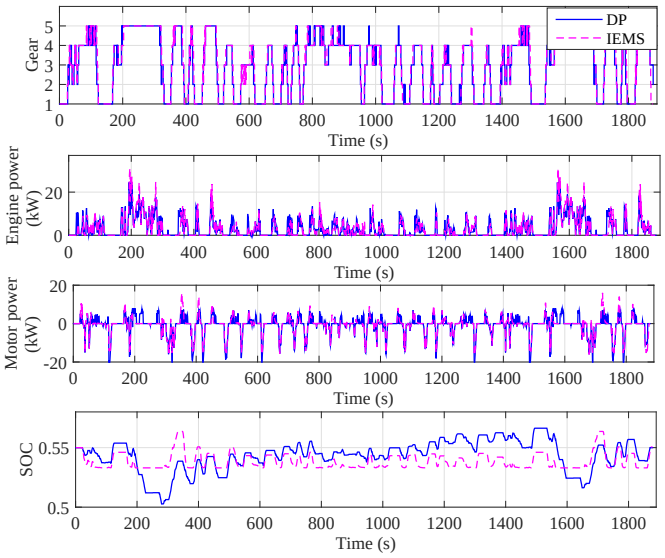


Fig. 9. Simulation results for FTP-75

TABLE II
SIMULATION RESULTS FOR FTP-75 (SOC(0) = 0.055)

Strategy	Fuel (g)	SOC(T_{cyc})	Change
IEMS	506.2	0.5506	+1%
DP	501.2	0.55	-

TABLE III
SIMULATION RESULTS FOR FURTHER DRIVING CYCLES (SOC(0) = 0.055)

Driving cycle	Strategy	Fuel (g)	SOC(T_{cyc})	Change	
WVUSUB	IEMS	322.88	0.5508	+ 2.2%	
	DP	315.8	0.55	0.55	-
LA92	IEMS	593.93	0.5509	+2.1%	
	DP	581.72	0.55	0.55	-
WVUINTER	IEMS	726.63	0.5497	+ 0.6%	
	DP	721.96	0.55	0.55	-

cost. Note that the developed IEMS utilizes only the current velocity $v(k)$ and acceleration $a(k)$ while the system model and the driving cycle are not required.

2) Analysis of robustness: In order to evaluate the robustness of the IEMS, the weights in ADHDP can be pre-trained offline for several driving cycles and then applied for the online power split and gear shift control for other driving cycles. Here the weights of the critic and action networks are pre-trained for a long trip resulting from concatenating four standard driving cycles (FTP-75, FTP-HIGHWAY, WVUCITY and WVUINTER) from [25] including city and highway driving patterns. The pre-trained weights are applied in the actor-critic reinforcement for the optimal gear shift and power split control. Simulation results for three different driving cycles are given in Table III. The IEMS can achieve charge sustainment with the similar initial and final SOC values. The fuel consumption with the IEMS increases from 0.6% to 2.2% compared to the results from DP, which again indicates that the developed control method is near-optimal and robust.

Detailed simulation result for the WVUSUB cycle are given in Fig. 10. The gear shift sequence from the IEMS follows the optimal one resulting from DP with only a few mismatches.

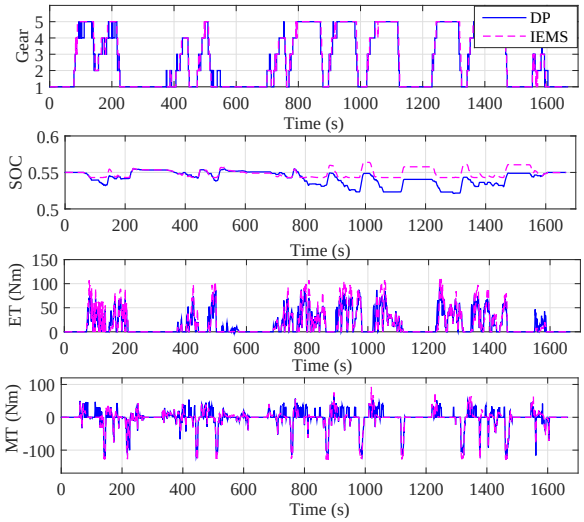


Fig. 10. Simulation results for WVUSUB

The engine torque (ET) and motor torque (MT) determined with DP and the IEMS follow similar profiles.

V. ECOLOGICAL ADAPTIVE CRUISE CONTROL FOR HEVS

In this section an ecological adaptive cruise control (ECO-ACC) is derived by integrating the ACC and the IEMS to achieve both a safe driving and optimal fuel economy in HEVs. For ACC the sampling interval is chosen as 0.1 s, allowing a fast reaction for safety. For the energy management in HEVs it is usually assumed that the gear shift happens within one second. Therefore different sampling intervals must be used for the ACC and the IEMS.

A. Control algorithm

The ECO-ACC algorithm for HEVs is illustrated in Fig. 11. At each time step, the velocity of the preceding vehicle v_{p} and the relative distance L are measured. Then the ACC determines the required acceleration a_{h} and the velocity v_{h} of the host vehicle. In order to get a smooth gear shift, the gear shift control is only performed with a sampling interval 1 s which is also the sampling interval of the IEMS. During the 1 s period, the gear position keeps unchanged for 10 steps of the 0.1 s period. When the time step t satisfies the condition $\mathrm{rem}(t,1) = 0$, the gear shift command is triggered and the IEMS derives the gear shift and power split jointly. Then T_{e} and T_{m} are applied to the HEV to follow the preceding vehicle. Here rem represents the remainder after the division of two variables. If t does not meet the condition $\mathrm{rem}(t,1) = 0$, the gear keeps the position at the previous time step and the power split between T_{e} and T_{m} is determined from the action network and applied to the HEV. The weights of the action network are not adapted during this process.

B. Simulation analysis

1) Analysis of velocity tracking: The adaptive cruise control enables the host vehicle to follow the velocity of the preceding vehicle. For comparison a single HEV whose velocity directly

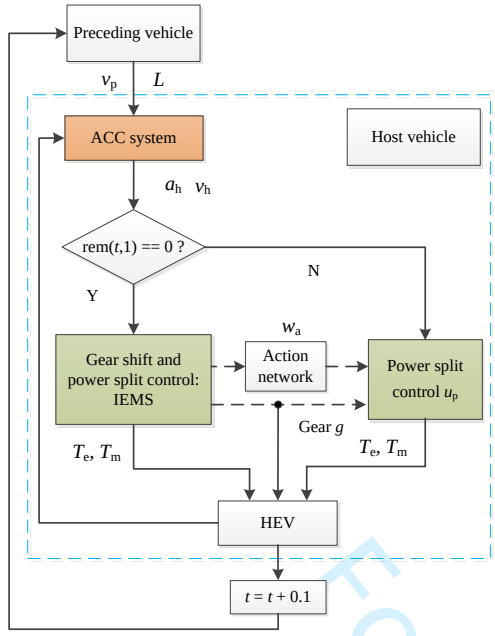


Fig. 11. Cntrol algorithm of ecological adaptive control of HEVs

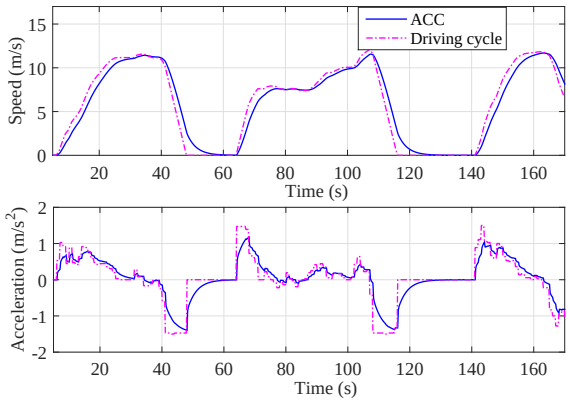


Fig. 12. Comparisons of the velocity and acceleration trajectories

from the driving cycle without ACC is used. The gear shift and power split control for the vehicle with and without ACC are both performed by the IEMS for fuel economy. In order to make clear comparison, the velocity and acceleration trajectories for UDDS cycle from [25] with 170 seconds are shown in Fig. 12. ACC can smoothen the velocity with smaller acceleration and deceleration value compared with that in the original driving cycle, which prevents fast speed increasing or decreasing. Simulation results on the fuel cost comparison is given in Table IV. The total fuel consumption decreases by 4.11% with ACC due to a smoother velocity trajectory. As discussed in Section III, the penalty on the traction force in the cost (15) can influence the fuel consumption through preventing unnecessary acceleration and deceleration. Therefore the velocity profile is optimized by ACC to reduce the fuel cost.

2) Robustness analysis: To validate the robustness of the ECO-ACC, different driving cycles are tested. For a fair comparison the velocity profile for the DP calculation is

TABLE IV
SIMULATION RESULTS WITH AND WITHOUT ACC (SOC(0) = 0.055)

Method	Fuel (g)	SOC(T_{cyc})	Change
with ACC	335.93	0.5506	- 4.11%
without ACC	350.34	0.5503	-

TABLE V
FUEL CONSUMPTION COMPARISON FOR ECO-ACC (SOC(0) = 0.055)

Driving cycle	Strategy	Fuel (g)	SOC(T_{cyc})	Change
WVUSUB	ECO-ACC	316.91	0.5498	+1%
	DP	313.83	0.55	-
FTP-75	ECO-ACC	504.08g	0.5496	+0.3%
	DP	502.57g	0.55	-
LA92	ECO-ACC	530.14	0.5496	+1.3%
	DP	523.15	0.55	-
UDDS	ECO-ACC	335.93	0.5506	+0.2%
	DP	335.13	0.55	-

provided by the ACC and the gear shift sequence is dependent on the velocity which is determined by IEMS. Therefore only the power split is calculated with DP to optimize the fuel consumption. The initial weights of the networks in IEMS are adopted from the pre-trained parameters for a long trip as detailed in Section IV-D2. Simulation results for different driving cycles are given in Table V. The total fuel cost with ECO-ACC increases by 0.2% to 1.3% compared to the one resulting from DP, indicating near-optimality and robustness.

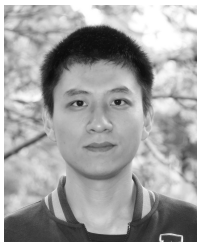
VI. CONCLUSIONS

In this paper, an ecological adaptive cruise control for HEVs in a car-following scenario to jointly optimize the driving safety and fuel consumption has been proposed. ADHDP is utilized in the ACC to maintain a desired inter-vehicle distance for safe and comfortable driving. The control policy can be learned online without a system model. The online gear shift control and power split control are realized by enumeration and ADHDP without prior knowledge of future driving information. The gear shift sequence, the engine torque, and the electric motor torque are close to the ones obtained from DP with the fuel consumption increasing by 0.6% to 2.2%. The ACC can reduce the fuel consumption by 4.11% by optimizing the velocity profile compared to the velocity trajectory given by the driving cycle. The fuel consumption obtained with the ecological adaptive cruise control increases by 0.2% to 1.3% compared to the one resulting from DP for different driving cycles, indicating the near-optimality and robustness.

REFERENCES

[1] L. Xiao and F. Gao, "A comprehensive review of the development of adaptive cruise control systems," *Vehicle System Dynamics*, vol. 48, no. 10, pp. 1167–1192, 2010.
[2] K. C. Dey, L. Yan, X. Wang, Y. Wang, H. Shen, M. Chowdhury, L. Yu, C. Qiu, and V. Soundararaj, "A review of communication, driver characteristics, and controls aspects of cooperative adaptive cruise control (CACC)," *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 2, pp. 491–509, 2016.
[3] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," *IEEE Control Systems Magazine*, vol. 27, no. 2, pp. 60–70, 2007.
[4] S. Li, K. Li, R. Rajamani, and J. Wang, "Model predictive multi-objective vehicular adaptive cruise control," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 3, pp. 556–566, 2011.

[5] D. Moser, R. Schmied, H. Waschl, and L. del Re, "Flexible spacing adaptive cruise control using stochastic model predictive control," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 1, pp. 114–127, 2017.
[6] S. E. Li, Q. Guo, L. Xin, B. Cheng, and K. Li, "Fuel-saving servo-loop control for an adaptive cruise control system of road vehicles with step-gear transmission," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2033–2043, 2017.
[7] C. Desjardins and B. Chaib-draa, "Cooperative adaptive cruise control: A reinforcement learning approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 12, no. 4, pp. 1248–1260, 2011.
[8] D. Zhao, Z. Hu, Z. Xia, C. Alippi, Y. Zhu, and D. Wang, "Full-range adaptive cruise control based on supervised adaptive dynamic programming," *Neurocomputing*, vol. 125, pp. 57–67, 2014.
[9] P. Pisu and G. Rizzoni, "A comparative study of supervisory control strategies for hybrid electric vehicles," *IEEE Transactions on Control Systems Technology*, vol. 15, no. 3, pp. 506–518, 2007.
[10] 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.
[11] O. Sundstrom and L. Guzzella, "A generic dynamic programming matlab function," in *Proceedings of the 18th IEEE Conference on Control Applications*, 2009, pp. 1625–1630.
[12] C. Sun, F. Sun, and H. He, "Investigating adaptive- ECMS with velocity forecast ability for hybrid electric vehicles," *Applied Energy*, vol. 185, pp. 1644–1653, 2017.
[13] N. Kim, S. Cha, and H. Peng, "Optimal control of hybrid electric vehicles based on pontryagin's minimum principle," *IEEE Transactions on Control Systems Technology*, vol. 19, no. 5, pp. 1279–1287, 2011.
[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 Transactions on Control Systems Technology*, vol. 20, no. 3, pp. 593–603, 2012.
[15] N. Murgovski, L. Johannesson, J. Sjöberg, and B. Egardt, "Component sizing of a plug-in hybrid electric powertrain via convex optimization," *Mechatronics*, vol. 22, no. 1, pp. 106–120, 2012.
[16] T. Nüesch, P. Elbert, M. Flankl, C. Onder, and L. Guzzella, "Convex optimization for the energy management of hybrid electric vehicles considering engine start and gearshift costs," *Energies*, vol. 7, no. 2, pp. 834–856, 2014.
[17] Y. L. Murphey, J. Park, L. Kiliaris, M. L. Kuang, M. A. Masrur, A. M. Phillips, and Q. Wang, "Intelligent hybrid vehicle power controlpart ii: Online intelligent energy management," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 1, pp. 69–79, 2013.
[18] T. Liu, Y. Zou, D. Liu, and F. Sun, "Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 12, pp. 7837–7846, 2015.
[19] V. Ngo, T. Hofman, M. Steinbuch, and A. Serrarens, "Optimal control of the gearshift command for hybrid electric vehicles," *IEEE Transactions on Vehicular Technology*, vol. 61, no. 8, pp. 3531–3543, 2012.
[20] Y. Luo, T. Chen, and K. Li, "Multi-objective decoupling algorithm for active distance control of intelligent hybrid electric vehicle," *Mechanical Systems and Signal Processing*, vol. 64, pp. 29–45, 2015.
[21] Y. Luo, T. Chen, S. Zhang, and K. Li, "Intelligent hybrid electric vehicle acc with coordinated control of tracking ability, fuel economy, and ride comfort," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2303–2308, 2015.
[22] L. Li, X. Wang, and J. Song, "Fuel consumption optimization for smart hybrid electric vehicle during a car-following process," *Mechanical Systems and Signal Processing*, vol. 87, pp. 17–29, 2017.
[23] J. Si and Y.-T. Wang, "Online learning control by association and reinforcement," *IEEE Transactions on Neural networks*, vol. 12, no. 2, pp. 264–276, 2001.
[24] F. L. Lewis and D. Liu, *Reinforcement learning and approximate dynamic programming for feedback control*. John Wiley & Sons, 2013.
[25] T. Markel, A. Brooker, T. Hendricks, V. Johnson, K. Kelly, and et al., "ADVISOR: a systems analysis tool for advanced vehicle modeling," *Journal of Power Sources*, vol. 110, no. 2, pp. 255–266, 2002.



Guoqiang Li received the B.S. and the M.S. degrees in mechanical engineering from Beijing Institute of Technology, Beijing, China, in 2011 and 2014 respectively. He is currently working toward his Ph.D. degree at the University of Kaiserslautern, Kaiserslautern, Germany. His research interests include the vehicle powertrain design and control, optimal control of hybrid electric vehicles, and learning control.

Daniel Görges (S'07–M'10) received the Dipl.-Ing. and the Dr.-Ing. degrees from the Department of Electrical and Computer Engineering, University of Kaiserslautern, Kaiserslautern, Germany, in 2005 and 2011. Since 2013 he is Juniorprofessor for Electromobility in the Department of Electrical and Computer Engineering, University of Kaiserslautern. His research interests include model predictive control, distributed control, networked control, learning control, hybrid systems, mechatronic systems, transportation systems, and power systems.