IEEE/ASME Transactions on Mechatronics, Volume: 22, Issue: 4, Aug. 2017 pp1497-1507
DOI: 10.1109/TMECH.2017.2707338

Reinforcement Learning Optimized Look-Ahead Energy Management of a Parallel Hybrid Electric Vehicle

Teng Liu, Xiaosong Hu, *Senior Member*, *IEEE*, Shengbo Eben Li, Dongpu Cao

***Abstract*—This paper presents a predictive energy management strategy for a parallel hybrid electric vehicle (HEV) based on velocity prediction and reinforcement learning (RL). The design procedure starts with modeling the parallel HEV as a systematic control-oriented model and defining a cost function. Fuzzy encoding and nearest neighbor approaches are proposed to achieve velocity prediction, and a finite-state Markov chain (MC) is exploited to learn transition probabilities of power demand. To determine the optimal control behaviors and power distribution between two energy sources, a novel RL-based energy management strategy is introduced. For comparison purposes, the two velocity prediction processes are examined by RL using the same realistic driving cycle. The look-ahead energy management strategy is contrasted with shortsighted and dynamic programming (DP)-based counterparts, and further validated by hardware-in-the-loop (HIL) test. The results demonstrate that the RL-optimized control is able to significantly reduce fuel consumption and computational time.**

***Index Terms*—Energy Management, Hybrid Electric Vehicle, Predictive Control, Markov Chain, Reinforcement Learning**

I. Introduction

HYBRID electric vehicles (HEVs) have been being greatly encouraged to overcome growing air pollution and oil consumption [1], [2]. HEVs of various configurations are increasing popular, as they can achieve great fuel economy and reduce emissions by multiple energy storage systems (ESSs) [3]. As one of the key technologies in HEVs, energy management affects the performance and cost effectiveness through governing power flow among multiple ESSs. The objective of energy management is to minimize a predifined cost function, such as harmful emissions, fuel economy, and running cost, while subjecting to necessary constraints [4].

Energy management strategies of HEVs can be mainly classified into two types: rule-based and optimization-based methods [5], [6]. The rule-based energy management strategies are widely applied in practice, since they can be easily exploited and are capable of operating steadily. Gao *et al.* proposed a novel rule-based energy management strategy that focuses on all charge depletion range and electric range operations [7]. The simulation results indicate that a significant amount of fuel can be displaced by electric energy in typical urban driving. Based on the state machine approach, a deterministic rule-based control strategy is proposed in [8], which has been successfully adopted in Toyota Prius and Honda Insight. Jalil *et al.* have devised a rule-based energy management strategy to set thresholds for determining power split between the engine and battery for a series HEV. Fuel economy exhibits an improvement of 11% in urban cycle and 6% in highway cycle [9]. All of these traditional rule-based schemes, however, are highly susceptible to heuristics and arbitrariness of design criterion and experience, thus losing a warranty of optimality [10].

Optimization-based energy management strategies can be further divided into global optimization and real-time optimization. Dynamic programming (DP) algorithm is a representative method to make a globally optimal control decision, as knowledge of driving cycle is presumably known in advance. In [11], Li *et al.* proposed a novel correctional DP-based controller to realize power split for a plug-in HEV, considering drivability and varying road slopes and loads. Based on a local linear approximation and a quadratic spline approximation, computational demand and memory storage requirements of DP algorithm are attenuated in [12]. Simulation results indicate that the computational time can be reduced by orders of magnitude with only a slight decline of fuel economy. Unfortunately, for practical applications, road topography is generally unknown, and thereby DP is inappropriate to real-time control [13]. Convex programming is another global optimization method that has been increasingly wielded for HEVs energy management [14], [15]. It arguably strikes a good balance between optimality and computational efficiency, via convex modeling and rapid solution search. In real-time optimization, equivalent consumption minimization strategy (ECMS) [16] and model predictive control (MPC) [17] are two most representative approaches. In order to derive an adaptive strategy, Rizzoni *et al.* added an on-the-fly algorithm to the ECMS framework to calculate the equivalent co-state

The work was in part supported by the EU-funded Marie Sklodowska-Curie Individual Fellowships (IF) Project under Grant 706253-pPHEV-H2020-MSCA-IF-2015. (T. Liu and X. Hu equally contributed to this research work, Corresponding author: X. Hu)

T. Liu is with the State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, China (email: tengliu17@gmail.com).

X. Hu is with the State Key Laboratory of Mechanical Transmission and Department of Automotive Engineering, Chongqing University, Chongqing 400044, China (email:xiaosonghu@ieee.org).

S. Li is with the State Key Laboratory of Automotive Safety and Energy, Department of Automotive Engineering, Tsinghua University, Beijing, 100084, China (email: lisb04@gmail.com).

D. Cao is with the Center for Automotive Engineering, Cranfield University, Bedford MK 43 0AL, UK (email: d.cao@cranfield.ac.uk).

according to driving conditions [18], [19]. In [20], the future speed is predicted periodically, and then a constant co-state in ECMS is evaluated backwards after each prediction. Nonetheless, the optimal co-state needs to be estimated offline, which strongly relies on the accuracy of velocity predictions [21]. For MPC, the controller settles an energy management problem via DP [22], quadratic programming [23], nonlinear programming [24], or Pontryagin's minimum principle (PMP) [25]. However, the performance of MPC control is highly determined by the precision of future velocity or power forecasts [26]. Numerous predictive control schemes were proposed, for example, Markov chain (MC) models, artificial neural networks (NNs), and radial basis function. In [27], Arsie *et al.* proposed a recurrent neural network (NN) to predict the future velocity profile in 20 seconds, based on the past and current speeds. After this operation, the global optimization problem is split into several local optimizations solved by DP. A Markov chain (MC) model is utilized for vehicular velocity prediction in [28], where a stochastic dynamic programming (SDP) is applied to optimize the energy mangement problem for a plug-in HEV.

Recently, two emerging methods, namely game theory (GT) [29] and reinforcement learning (RL) [30], [31], have been presented to implement real-time optimization feasible for HEVs. Chen *et al.* reported a game-theoretic approach based on a two-level single-leader multi-follower game in [32], where the vehicular fuel economy is close to the benchmarking optimal solution. In [33], a RL-based blended real-time energy management strategy is synthesized to address trade-off between real-time performance and optimality. Numerical analysis unveils that the RL-enabled strategy can achieve a near-optimal solution with 11.93% fuel savings, compared to a binary mode control strategy. We discussed adaptability and optimality of RL algorithm in [34], showing its advantages over SDP in fuel economy and computation time. Moreover, we also incorporated RL into a real-time control framework in [35]. The associated results indicate that the RL-based energy management strategy can considerably improve fuel efficiency and allows real-time implementation. To the best of our knowledge, combing RL with velocity forecasts indicative of future road information, nevertheless, has not been investigated. Furthermore, RL-based energy management of HEVs still lacks experimental verification.

In order to bridge the foregoing research gap, this article constructs a predictive energy management strategy for a parallel HEV via a synergy of velocity prediction and RL. First, the dynamics of the hybrid powertrain are modeled and formulated. Then, the nearest neighbor and fuzzy encoding approaches are compared, in terms of the performance of velocity prediction, meanwhile, a finite-state MC is exploited to learn transition probabilities of power demand. The *Q*-learing algorithm is harnessed to realize the predictive optimal control for increasing fuel economy and maintaining battery charge sustenance. Finally, the RL-based predictive energy management strategy is in contrast with the benchmarking DP to validate its effectiveness. In addition, the RL-driven strategy is verified through a hardware-in-the-loop (HIL) experiment. Three perspectives are contributed to the related literature: (1) two velocity prediction methods, i.e., nearest neighbor and fuzzy encoding using MC, are presented and validated via RL; (2) a comparison between the RL-based optimal control and DP-based one is illuminated; (3) an HIL experiment is carried out to evidence the performance of the proposed energy management strategy.

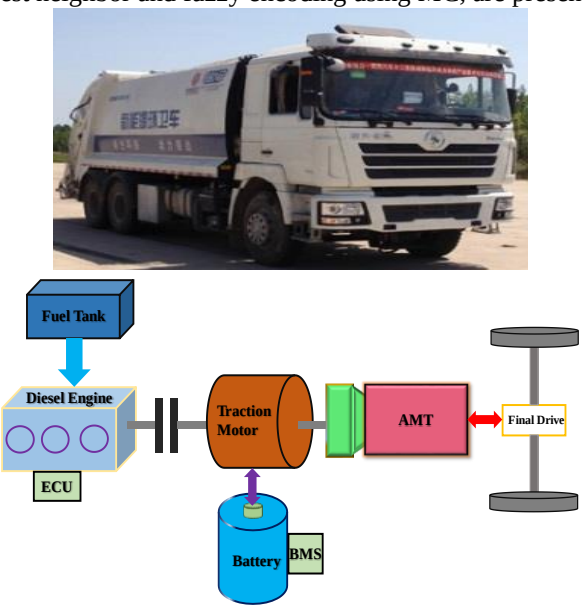


Fig. 1. Configuration of the parallel HEV powertrain.

TABLE I
MAIN PARAMETERS OF THE PARALLEL HEV SPECIFICATION

Symbol	Parameters	Values
m	Curb weight	16000 kg
A	Fronted area	1.8 m^2
C_d	Aerodynamic coefficient	0.55
η_T	Transmission axle efficiency	0.9
η_m	Traction motor efficiency	0.95
f	Rolling resistance coefficient	0.021
R	Tire radius	0.508 m

The remainder of this paper is organized as follows: Section II illustrates the configuration of the hybrid powertrain, where the optimal control problem is formulated as well; Section III describes the two velocity prediction approaches and the structure of the *Q*-learing algorithm; the comparative study between RL-based and DP-based optimization results is shown in Section IV; key takeaways are summarized in Section V.

II. CONFIGURATION OF THE PARALLEL HYBRID ELECTRIC VEHICLE AND PROBLEM FORMULATION

The vehicle studied is a commercial parallel HEV and its powertrain configuration is sketched in Fig. 1, which consists of a diesel engine, a battery pack, a traction motor, and an automated mechanical transmission (AMT). The rated power of the diesel engine is 155 kW at the speed of 2000 rpm, and the maximum torque is 900 Nm within the speed range from 1300 rpm to 1600 rpm. The traction motor has a maximum power of

90 kW, a maximum torque of 600 Nm, and a maximum speed of 2400 rpm. The battery pack is 60 Ah capacity with the nominal voltage of 312.5 V. The primary parameters of the parallel HEV are listed in Table I [36].

A. Power Demand Modeling

When the velocity profile is known a priori, the power demand to drive the vehicle is computed as follows:

$$P_{dem} = (F_i + F_a + F_f)v \tag{1}$$

where F_i is the inertial force, F_a is the aerodynamic drag, F_f is the rolling resistance, and v is the vehicle speed. The three types of environmental resistance can be calculated by

$$\begin{cases} F_i = \delta ma \\ F_a = (C_d A / 21.15)v^2 \\ F_f = mgf \end{cases} \tag{2}$$

where δ is the mass factor caused by the moment of inertia of four wheels and powertrain rotating components, m is the vehicle mass, a is the acceleration, and C_d is the aerodynamic coefficient. Furthermore, A is the fronted area, g is the gravity coefficient, and f is the coefficient of rolling resistance.

In order to maintain the energy balance of the vehicle, the power demand should be provided by the engine and battery together

$$P_{dem} = (P_{en} + P_{bat}\eta_m)\eta_T \tag{3}$$

where P_{en} is the output power from the engine, P_{bat} is the battery power, η_m is the traction motor efficiency, and η_T is the efficiency of the transmission and axle. The engine power is decided by the throttle variable, and then the battery power can be estimated from (3). In this paper, we set the throttle signal $th(t)$ to be the control variable of the energy management problem.

B. Engine Modeling

A quasi-static model is utilized to evaluate the fuel economy of engine [37]. The fuel consumption rate is defined as

$$\dot{m}_f = f(T_{en}, n_{en}) \tag{4}$$

where T_{en} is the engine output torque, and n_{en} is the engine speed. Then the total fuel consumption can be integrated as

$$Fuel = \int_0^T \dot{m}_f dt \tag{5}$$

where $t \in [0, T]$ is the specific time horizon.

C. Battery Modeling

The state of charge (SOC) in the battery is chosen as the state variable, which is calculated by

$$\dot{SOC} = -\frac{I_{bat}(t)}{Q_{bat}} \tag{6}$$

where I_{bat} is the battery current, and Q_{bat} is the battery nominal capacity. An internal resistance model is herein applied to reformulate the expression of SOC as [38]

$$\dot{SOC} = -\frac{V_{oc} - \sqrt{V_{oc}^2 - 4R_{\text{int}}P_{bat}}}{2R_{\text{int}}Q_{bat}} \tag{7}$$

where P_{bat} is the battery output power, V_{oc} is the battery open-circuit voltage, and R_{int} is the battery internal resistance. All of them are a function of SOC.

D. Energy Management Problem

In this work, the cost function is specified to minimize a trade-off between the fuel consumption and charge sustenance:

$$\begin{cases} J = \int_0^T [\dot{m}_f(t) + \beta\Delta_{SOC}^2]dt \\ \Delta_{SOC} = \begin{cases} SOC(t) - SOC_{ref} & SOC(t) < SOC_{ref} \\ 0 & SOC(t) \geq SOC_{ref} \end{cases} \end{cases} \tag{8}$$

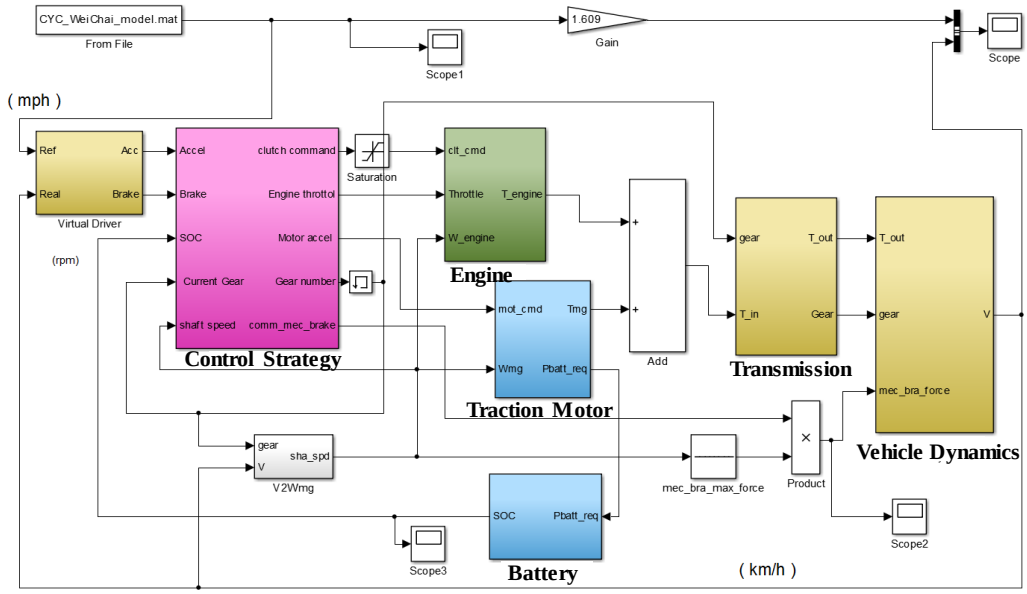


Fig. 2. Quasi-static parallel HEV model for HIL simulation.

where β is a positive weighting factor, and SOC_{ref} is a pre-assigned constant to maintain charge-sustaining constraints [39]. To ensure safety and reliability of the components, the following inequality constraints should be satisfied:

$$\begin{cases} 0 \le T_{en}(t) \le T_{en,\max} \\ SOC_{\min} \le SOC(t) \le SOC_{\max} \\ P_{bat,\min} \le P_{bat}(t) \le P_{bat,\max} \\ n_{en,\min} \le n_{en}(t) \le n_{en,\max} \\ I_{bat,\min} \le I_{bat} \le I_{bat,\max}. \end{cases} \tag{9}$$

Fig. 2 displays the overall quasi-static model of the parallel HEV in Matlab/Simulink. Since the emphasis in this paper is on discussing the RL-based predictive optimal control strategy, the implication of battery temperature change and aging is not considered, and the gear position is assumed to be appropriate at all times.

III. Velocity Prediction and Reinforcement Learning

A. Nearest Neighbor Velocity Predictor

In this paper, the vehicle velocity is modeled as a finite-state MC [40] and denoted as $V=\{v_j \mid j=1, \ldots, M\} \subset X$, where $X \subset R$ is bounded. The maximum likelihood estimator is used to estimate the transition probability of the vehicle velocity by

$$\begin{cases} p_{ij} = P(v^+ = v_j \,|\, v = v_i) = \dfrac{N_{ij}}{N_i} \\ N_i = \sum_{j=1}^{M} N_{ij} \end{cases} \tag{10}$$

where v and v^+ are the present and next one step-ahead velocity, respectively, and p_{ij} is the transition probability from v_i to v_j. Furthermore, N_{ij} indicates the transition counts from v_i to v_j, N_i is the total transition counts initiated from v_i, and the transition probability matrix (TPM) Π is filled with elements p_{ij}. The one step-ahead probability vector of v taking one of finite values v_j is linked as

$$(p^+)^T = p^T \Pi \tag{11}$$

and for $n>1$ steps ahead as

$$(p^{+n})^T = p^T \Pi^n. \tag{12}$$

In the nearest neighbor approach, X is divided into a finite set of disjoint intervals, I_j, $j=1, \ldots, M$, and each interval is assigned a Markov chain state, $v_j \in I_j$, which is typically the midpoint of the interval I_j. Based on this partitioning, a continuous state $v \in I_j$ corresponds to a discrete state v_j and may be associated with an M-dimensional probability vector $\alpha^T(v)=[0 \cdots 1 \cdots 0]$ with the j-th element is 1 and other elements equal to 0. Motivated by (11) and $\alpha(v)$, the probability vector of the next state is defined as

$$(\alpha^+(v))^T = (\alpha(v))^T \Pi = \Pi_j^T \tag{13}$$

where Π^T_j denotes the j-th row of the TPM Π. In the nearest neighbor predictor (NNP), the next one-step ahead velocity can be predicted as an expectation, according to the interval midpoints:

$$v^+ = \sum_{j=1}^{M} p_{ij} v_j \quad if\ v \in I_i. \tag{14}$$

B. Fuzzy Encoding Velocity Predictor

In the fuzzy encoding technique, the intervals I_j are replaced with fuzzy subsets Φ_j, $j=1, \ldots, M$. In fuzzy logic, the fuzzy

subset Φ_j is a pair $(X, \mu_j(\cdot))$, and $\mu_j(\cdot)$ is a Lebesgue measurable membership function that satisfies the property

$$\mu_j : X \to [0,1] s.t. \forall v \in X, \exists j, 1 \le j \le M, \mu_j(v) > 0 \tag{15}$$

where $\mu_j(v)$ reflects the degree of membership of $v \in X$ in μ_j. Unlike interval partitioning in NNP, a continuous state $v \in X$ in the fuzzy encoding may be associated with several states v_j of the underlying finite-state MC model [40].

The fuzzy encoding predictor (FEP) involves two transformations based on the theory of approximate reasoning [41]. The first transformation allocates an M-dimensional possibility (not probability) vector for each $v \in X$ as follow:

$$\tilde{O}^T(v) = \mu^T(v) = [\mu_1(v), \mu_2(v), \cdots, \mu_M(v)]. \tag{16}$$

Different from the probability vector $\alpha(v)$, the sum of the elements in the possibility vector $\tilde{O}(v)$ is unnecessary to equal 1. This transformation is named fuzzification and maps velocity in the space X to vector in M-dimensional possibility vector space $\tilde{X}$.

The second transformation is called the proportional possibility-to-probability transformation that converts the possibility vector $\tilde{O}(v)$ to a probability vector $O(v)$ by normalization:

$$O(v) = \tilde{O}(v) / \sum_{j=1}^{M} \tilde{O}_j(v) \tag{17}$$

where this transformation maps $\tilde{X}$ to an M-dimensional probability vector space, $\bar{X}$. Motivated by (13), the probability distribution of the next state in $\bar{X}$ is computed as

$$(O^+(v))^T = (O(v))^T \Pi \tag{18}$$

where the element p_{ij} in the TPM Π is interpreted as a transition probability between Φ_i and Φ_j. To decode vectors in $\bar{X}$ back to X, the probability distribution $O^+(v)$ is utilized to aggregate the membership function $\mu(v)$ to encode the probability vector of the next state in X [42]:

$$w^+(v) = (O^+(v))^T \mu(v) = (O(v))^T \Pi \mu(v). \tag{19}$$

Same as (14), the expected value over the possibility vector leads to the next one-step ahead velocity in FEP:

$$\begin{cases} v^+ = \int_X w^+(y) y dy / \int_X w^+(y) dy \\ \int_X w^+(y) y dy = \sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij} \int_X y \mu_j(y) dy \\ \int_X w^+(y) dy = \sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij} \int_X \mu_j(y) dy. \end{cases} \tag{20}$$

Note that the centroid and volume of the membership function $\mu_j(v)$ is expressed as

$$\begin{cases} \bar{c}_i = \int_X y \mu_j(y) dy \\ V_j = \int_X \mu_j(y) dy. \end{cases} \tag{21}$$

Thus, the expression (20) is rewritten as

$$\begin{cases} v^+ = \dfrac{\sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij} V_j \bar{c}_j}{\sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij} V_j}. \end{cases} \tag{22}$$

Assuming that membership functions have the same volume and using the fact $\sum_{j=1}^{M} p_{ij} = 1$ and $\sum_{i=1}^{M} O_i(v) = 1$, (22) is further simplified to

$$\begin{cases} v^+ = \dfrac{\sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij} c_j}{\sum_{i=1}^{M} O_i(v) \sum_{j=1}^{M} p_{ij}} = (O(v))^T \Pi \bar{c} \end{cases} \tag{23}$$

where (23) is the next one-step ahead velocity using FEP. It is noticed that the probability distribution and centroid in (23) is related to the membership functions. In this paper, these functions are taken as a Gaussian membership function [43] with the standard deviation $\sigma = 1$ as follows:

$$q_i = e^{\frac{-(x-2.5i+1.25)^2}{2 \cdot \sigma^2}}, \quad i = 1, \ldots 12. \tag{24}$$

C. Reinforcement Learning Algorithm

The interaction between the agent and environment in RL is modeled as a discrete discounted Markov decision process (MDP), as shown in Fig. 3. The MDP is a quintuple ($\boldsymbol{S}$, $\boldsymbol{A}$, $\boldsymbol{\Pi}$, $\boldsymbol{R}$, δ), where $\boldsymbol{S}$ and $\boldsymbol{A}$ are the set of states and actions, $\boldsymbol{\Pi}$ is the TPM, $\boldsymbol{R}$ is the reward function, and $\delta \in (0, 1)$ is a discount factor. Variables $p_{sa,s'}$ and $r(s, a)$ are denoted as the transition probability from state s to next state s' using action a and the reward of taking action a at state s, respectively.

The control policy π is the distribution over the control actions a, given the current state s. The optimal value function is exhibited as the finite expected discounted sum of the rewards [44]:

$$V^*(s) = \min_{\pi} E(\sum_{t=0}^{T} \delta^t r). \tag{25}$$

Because of the uniqueness, (24) can be reformulated as a recursion expression

$$V^*(s) = = \min_{a} (r(s,a) + \delta \sum_{s' \in S} p_{sa,s'} V^*(s')) \quad \forall s \in S. \tag{26}$$

Given the optimal value function, the optimal control policy is determined as follows:

$$\pi^*(s) = \arg\min_a (r(s,a) + \delta \sum_{s' \in S} p_{sa,s'} V^*(s')). \quad (27)$$

In addition, the action-value function $Q(s, a)$ and its optimal value $Q^*(s, a)$ are expressed as the following formula:

$$\begin{cases} Q(s,a) = r(s,a) + \delta \sum_{s' \in S} p_{s'a,s} Q(s', a') \\ Q^*(s,a) = r(s,a) + \delta \sum_{s' \in S} p_{s'a,s} \min_{a'} Q^*(s', a'). \end{cases} \quad (28)$$

The variable $V^*(s)$ is the value of s assuming that an optimal action is taken initially; therefore, $V^*(s) = Q^*(s, a)$ and $\pi^*(s) = arg\ \min_a Q^*(s, a)$. The updated rule of Q value for Q-learning algorithm is expressed by [45]

$$Q(s,a) \leftarrow Q(s,a) + \eta(r(s,a) + \delta \min_{a'} Q(s', a') - Q(s,a)) \quad (29)$$

where $\eta \in [0, 1]$ is a decaying factor of the Q-learning algorithm. As the vehicle velocity is predicted using NNP and FEP, (28) is used to acquire the RL-based predictive energy management strategy. The pseudo-code of the Q-learning algorithm is described in Table II.

TABLE II
PSEUDO-CODEOF THE Q-LEARNING ALGORITHM

Algorithm: Q-learning Algorithm
1. Initialize $Q(s, a)$, s, and number of iteration N
2. Repeat each step k=1, 2, 3…
3. Choose a, based on $Q(s, \cdot)$ (ε-greedy policy)
4. Taking action a, observe r, s'
5. Define a^*=arg $\max_a Q(s', a)$
6. $Q(s, a) \leftarrow Q(s, a) + \eta(r(s, a) + \delta \max_{a'} Q(s', a') - Q(s, a))$
7. $s \leftarrow s'$
8. until s is terminal

Specially, the energy management problem in this paper involves a set of state variables $\boldsymbol{S}=\{(SOC(t))|0.5 \le SOC(t) \le 0.8\}$, a set of actions $\boldsymbol{A}=\{th(t)|\ 0 \le th(t) \le 1\}$, and a reward function $\boldsymbol{R}=\{\dot{m}_f(s, a)\}$. In order to compare the performance of NNP and FEP for the energy management problem, two factors in the Q-learning algorithm are defined as the same value. The decaying factor η is correlated with the time step k and taken as $1/\sqrt{k+2}$, the discount factor δ is taken as 0.95, the number of iteration N is 10000, and the sample time is 1 second.

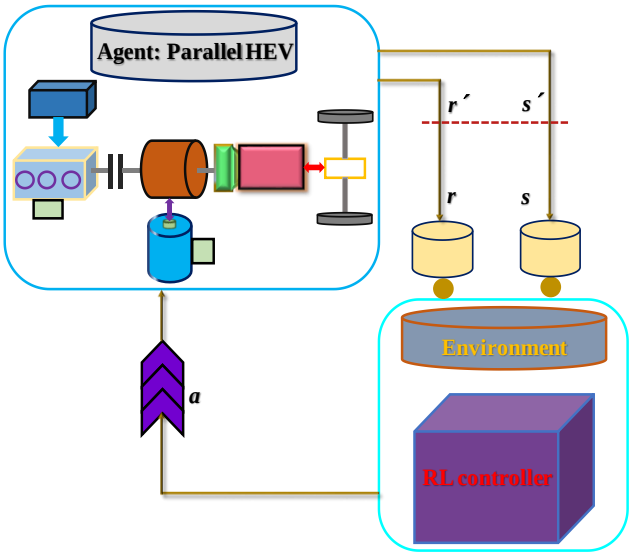


Fig. 3. Interaction between agent and environment in RL.

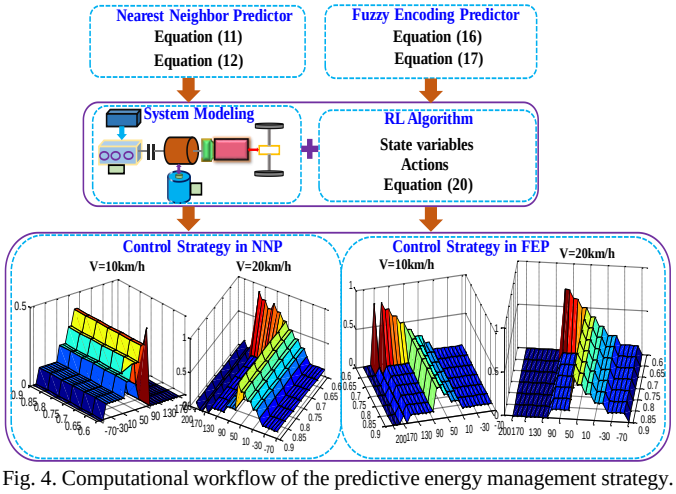


Fig. 4. Computational workflow of the predictive energy management strategy.

The computational process of the predictive energy management strategy is depicted in Fig. 4, which comprises two velocity predictors, system modeling, and the relevant control policy. The RL process is implemented in Matlab using the Markov decision process (MDP) introduced in [46]. The proposed control strategy can be utilized in real time, meanwhile, its optimality and robustness will be validated in Section IV.

IV. RESULTS AND ANALYSIS

The proposed predictive energy management strategy is compared with the DP-based and non-predictive ones in this section. First, two velocity predictors are evaluated in terms of mean square error (MSE). Subsequently, the non-predictive control strategy is derived from a RL algorithm according to a long driving schedule [36], and the DP-based control strategy is adopted as an optimality benchmark for the RL-based energy

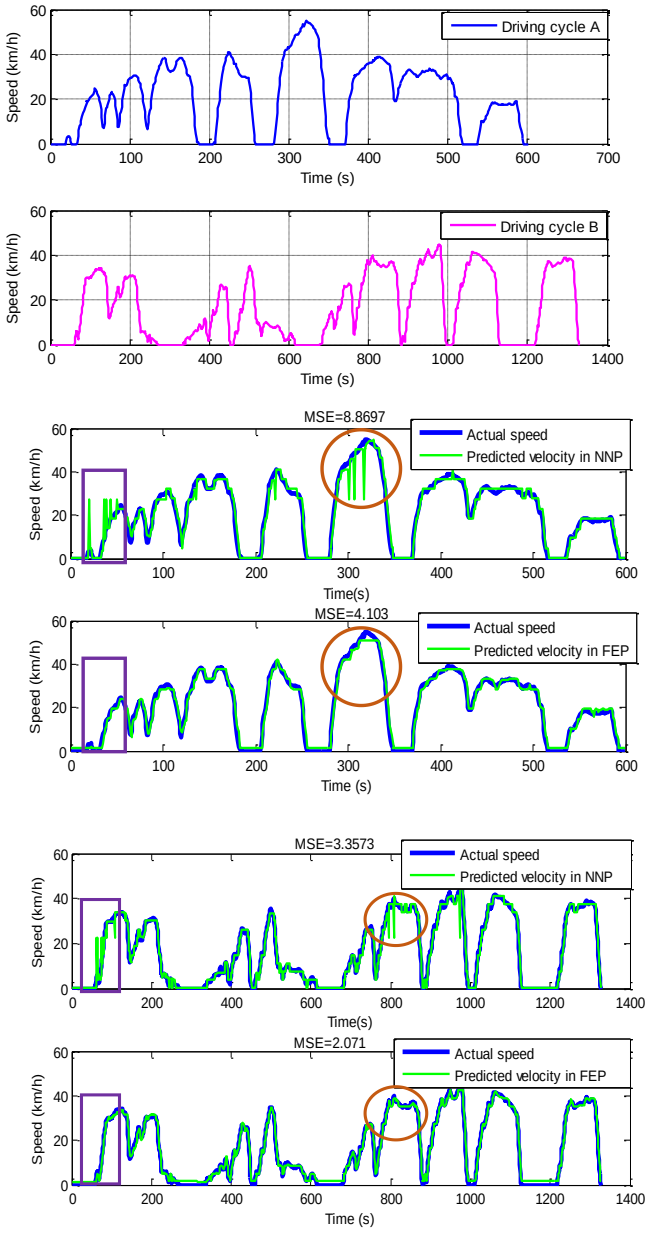


Fig. 5. One-step ahead velocity prediction for two realistic driving cycles.

management strategy. Ultimately, an HIL experimental validation is conducted.

A. Comparison of Two Velocity Predictors

The NNP and FEP are utilized to predict vehicle velocity at different step grades. Fig. 5 illustrates two realistic driving cycles and the one-step ahead velocity prediction for them. It is apparent that the FEP can achieve excellent accuracy, compared with the NNP, as the purple rectangles and orange ellipses highlight. The MSE in FEP (A=4.103, B=2.071) is less than that in NNP (A=8.8697, B=3.3573) for the two driving cycles.

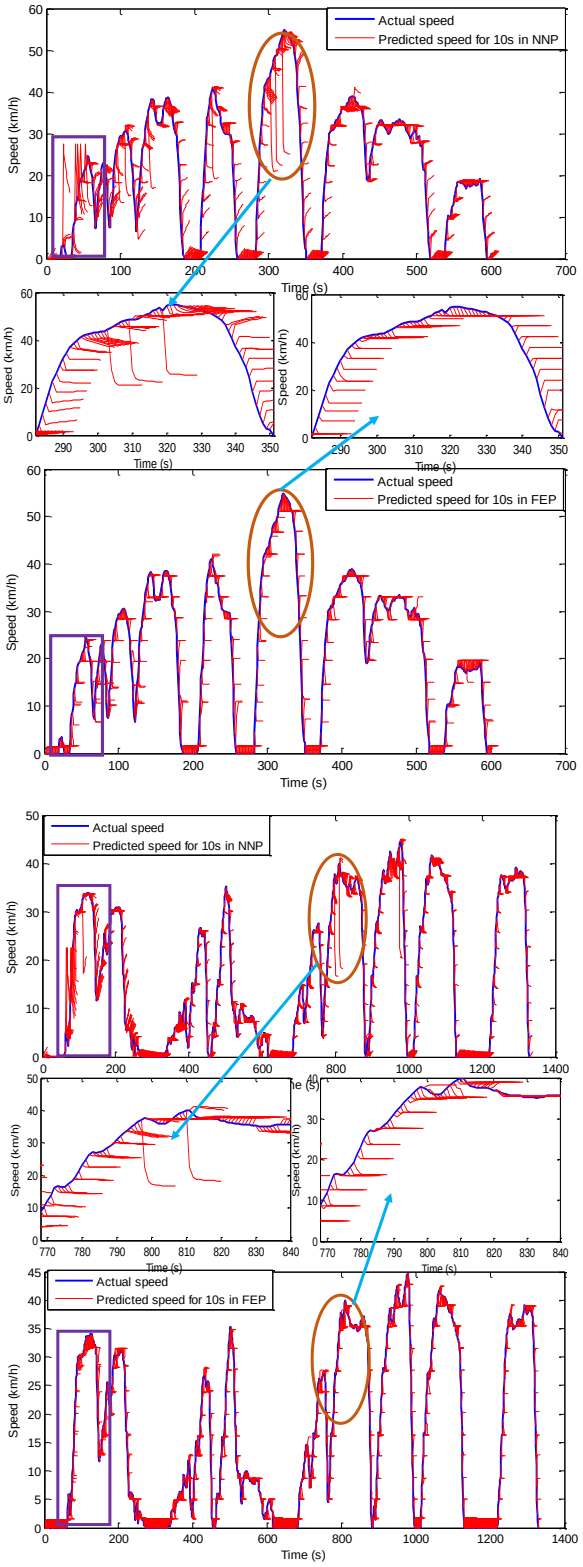


Fig. 6. 10-step ahead velocity prediction for two driving cycles.

Fig. 6 indicates the 10-step ahead prediction trajectories for the two driving cycles, based on NNP and FEP. The purple rectangles and orange ellipses underline that the FEP obtains superior prediction precision than NNP. The MSE for the FEP (A=3.626, B=3.516) is better than NNP (A=6.071, B=4.866) in the predicted availability.

B. *Comparison of Different Control Strategies*

The NNP and FEP based RL-enabled energy management strategies with 6-step ahead prediction are further compared with the DP-based one and a non-predictive control strategy.

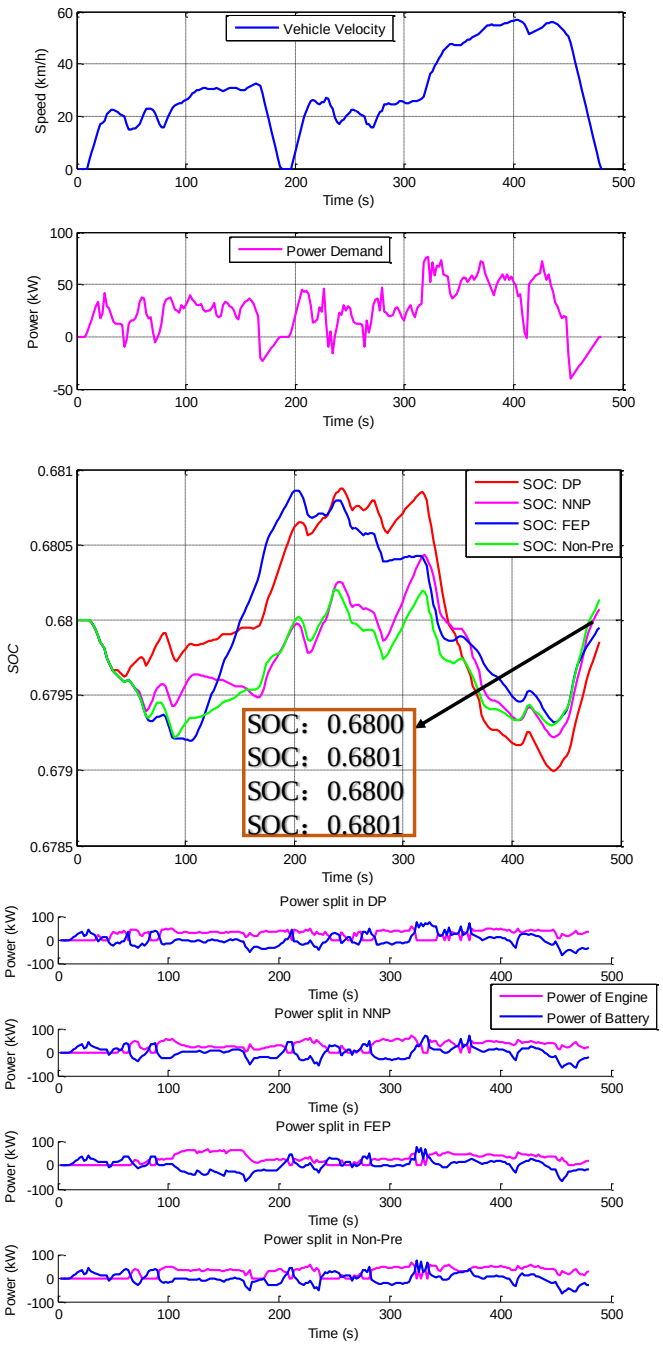


Fig. 7. SOC trajectories and power split for a simulation cycle with different control strategies.

Fig. 7 illustrates the SOC evolution and power split for a simulation cycle. It can be discerned that the SOC trajectory in the FEP based predictive control strategy is close to that of DP-based control strategy and clearly differs from those of the NNP-based and non-predictive controls. We can observe an analogous result in the power split trajectory.

TABLE III
THE FUEL CONSUMPTION AFTER SOC-CORRECTION IN DIFFERENT CONTROL STRATEGIES

Algorithms	Fuel consumption (g)	Relative increase (%)
DP	172	—
FEP	179	4.07
NNP	188	9.3
Non-Pre	196	13.95

TABLE IV
THE COMPUTATIONAL TIME IN DIFFERENT CONTROL STRATEGIES

Algorithms	Time[a] (min)	Relative increase (%)
Non-Pre	4.02	—
NNP	4.58	13.98
FEP	5.65	41.27
DP	8.21	104.23

[a] A 2.4 GHz microprocessor with 12 GB RAM was used.

The working points of the engine with the different control strategies are shown in Fig. 8. The engine working points under the predictive and DP-based control strategies locate in the lower fuel-consumption region more frequently, compared to the non-predictive control. Table III depicts the fuel consumption after SOC-correction for the four control strategies. Obviously, the fuel consumption under the FEP-based predictive control strategy is closest to that of the DP-based control, 9.8 % lower than that of the non-predictive control. The computational time of these control strategies is contrasted in Table IV. It is evident that both predictive controls are far faster than the DP-based control, which makes them online optimization feasible.

C. *Validation in the HIL Experiment*

An HIL experiment was conducted to assess the performance of the predictive RL-based energy management strategy. The rule-based control is adopted as the referential strategy that contains three modes, namely pure electric mode, hybrid mode, and charging mode. As an illustration, the hybrid mode implemented in Stateflow/Simulink is depicted in Fig. 9.

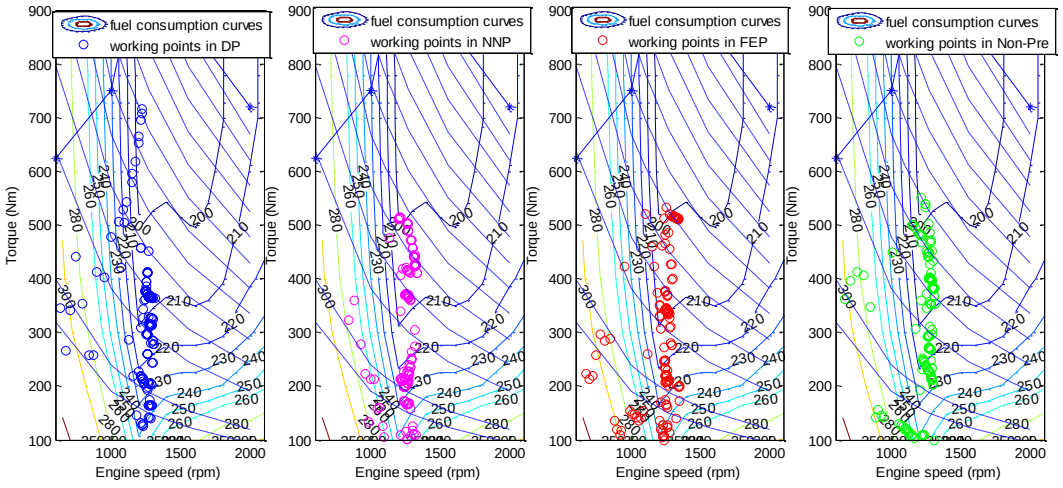


Fig. 8. Engine working area in different control strategies.

The experimental test setup includes control system development platform in MotoTron and vehicle model system development platform in RT-Lab, both of which are software-hardware development platforms on Matlab/Simulink providing C language rapid generating and online calibration functions [10].

```
Hybrid
  Acc
  du:
  comm_clt=1;
  req_torq=Accel*(mot_max_torq+eng_opt_torq);
  comm_eng=max(min(opt_throt,1),0);
  comm_mot=max(-1,min((req_torq-opt_throt*eng_opt_torq)/mot_max_torq,1));
  comm_shift=opt_gear;
  comm_mec_brake=0;
        [Accel>0 & Brake==0]          [Accel==0 | Brake>0]
  Bra
  du:
  comm_clt=2;
  comm_eng=0;
  req_torq=Brake*(mot_max_brake+F_max_mech*0.508/g_ratio/4.769);
  comm_mot=-max(mot_max_brake,req_torq)/mot_max_brake;
  comm_mec_brake=(-comm_mot*mot_max_brake-req_torq)
  /(-F_max_mech*0.508/g_ratio/4.769);
  comm_shift=cur_g;
```

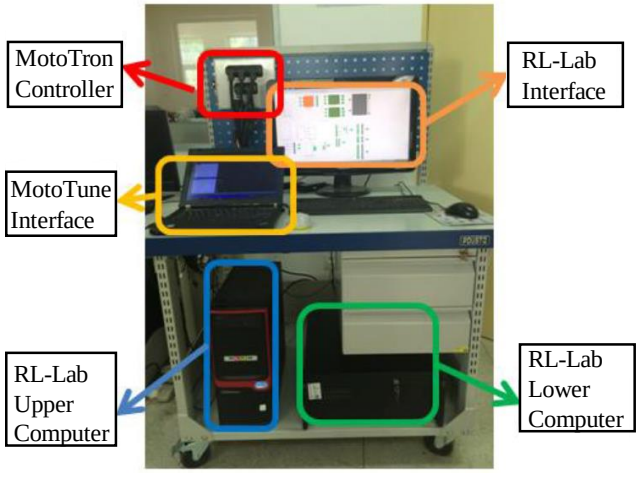


Fig. 9. HIL experimental bench.

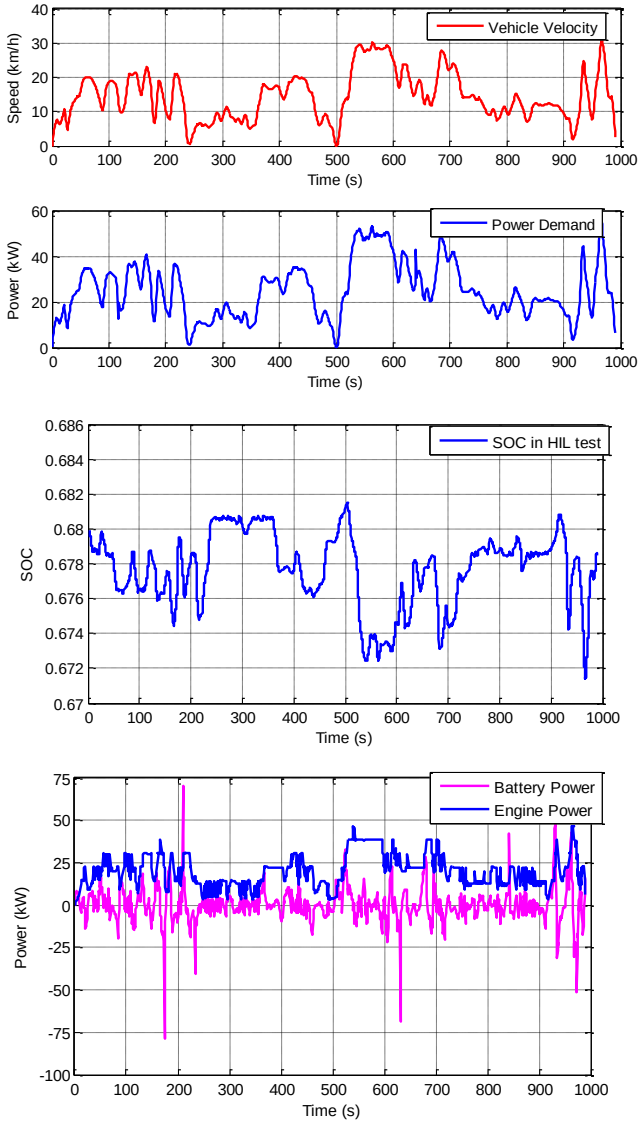


Fig. 10.SOC trajectories and power split for a real-world driving cycle in the HIL experiment.

The input/output interface of control strategy is set by MotoHawk toolkit, and the MotoTune software is employed to download the predictive control kernel into the controller hardware. The parallel HEV simulation model is established in the RL-Lab software, and the RL-Lab hardware is applied to download the vehicle model by Simulink automatic code generation technology. A Photo of the HIL experimental bench is also described in Fig. 9, which mainly consists of a controller, MotoTorn (hardware and software), and RL-Lab (hardware and software).

The predictive control was tested in the parallel HEV model environment over a real-world driving cycle. The simulation results are showcased in Fig. 10. Compared with the pre-existing rule-based control strategy, the engine is able to frequently work in the low fuel-consumption region in the predictive control. The fuel consumption of the predictive control (235 g) is 17.54% lower than that of the rule-based one (285 g). It can be concluded that relative to the rule-based scheme, the proposed predictive control strategy is more fuel-saving, while possessing real-time applicability.

V. Conclusion

This paper develops a reinforcement learning (RL) enabled predictive control strategy for a parallel HEV. First, a detailed control-oriented model for the parallel HEV is built. Then, two novel velocity predictors are presented to predict the future velocity profile in the RL control framework. Different driving cycles are applied to validate the performance of the two velocity predictors based on the *Q*-learning algorithm. The predictive control scheme is compared with non-predictive and DP-based ones, in order to demonstrate its optimality and potential in real-time control. The computational time of the DP-based control is considerably larger than that of the RL-based predictive control. The results in an HIL experiment substantiate that the predictive controller is real-time implementable and enables lower fuel consumption than do common counterparts, i.e., rule-based control solutions.

References

[1] H. Kim, and D. Kum, "Comprehensive design methodology of input-and output-split hybrid electric vehicles: In Search of Optimal Configuration," *IEEE/ASME Trans. Mechatronics*, 2016.

[2] Y. Chen, and J. Wang, "Design and experimental evaluations on energy efficient control allocation methods for overactuated electric vehicles: Longitudinal motion case," *IEEE/ASME Trans. Mechatronics*, vol.19, no.2, pp.538-548, 2014.

[3] C. M. Martinez, X. Hu, D. Cao, E. Velenis, B. Gao, and M. Wellers, "Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective," *IEEE Trans. Veh. Technol.*, 2016.

[4] J. P. F. Trovao, V. D. N. Santos, C. Henggeler Antunes, P. G. Pereirinha, and H. M. Jorge, "A real-time energy management architecture for multisource electric vehicles," *IEEE Trans. Ind. Electron.*, vol.62, no.5, pp. 3223-3233, May. 2015.

[5] T. H. Feng, L. Yang, Q. Gu, and Y. Q. Hu, "A supervisory control strategy for plug-in hybrid electric vehicles based on energy demand prediction and route preview," *IEEE Trans. Veh. Technol.*, vol.64, no.5, pp.1691-1700, May 2015.

[6] A. Sciarretta and L. Guzzella, "Control of hybrid electric vehicles," *IEEE Contr. Syst. Mag.*, vol.27, no.2, pp.60-70, 2007.

[7] Y. M. Gao and E. Mehrdad, "Design and control methodology of plug-in hybrid electric vehicles," *IEEE Trans. Ind. Electron.*, vol.57, no.2, pp.633-640, 2010.

[8] S. G. Wirasingha and A. Emadi, "Classification and review of control strategies for plug-in hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol.60, no.1, pp.111-122, 2011.

[9] N. Jalil, N. A. Kheir, and M. Salman, "A rule-based energy management strategy for a series hybrid vehicle," *In: Proc. American Control Conference*., pp.689-93, 1997.

[10] J. K. Peng, H. W. He, and R. Xiong, "Rule based energy management strategy for a series–parallel plug-in hybrid electric bus optimized by dynamic programming," *Applied Energy*, 2016.

[11] L. Li, C. Yang, and Y. H. Zhang, "Correctional DP-based energy management strategy of plug-in hybrid electric bus for city-bus route," *IEEE Trans. Veh. Technol.*, vol.64, no.7, pp.2792-2803, 2015.

[12] V. Larsson, L. Johannesson, and B. Egardt, "Analytic solutions to the dynamic programming subproblem in hybrid vehicle energy management," *IEEE Trans. Veh. Technol.*, vol.64, no.4, pp.1458-1467, Apr. 2015.

[13] L. Serrao, S. Onori, and G. Rizzoni, "A comparative analysis of energy management strategies for hybrid electric vehicles," *J. Dyn. Sys. Meas. Control.*, vol.133, no.3, pp.1-9, 2011.

[14] X. Hu, N.Murgovski, L. M. Johannesson, and B. Egardt, "Comparison of three electrochemical energy buffers applied to a hybrid bus powertrain with simultaneous optimal sizing and energy management," *IEEE Trans. Intell. Transp. Syst.*, vol.15, no.3, pp. 1193-1205, 2014.

[15] X. Hu, Y. Zou, and Y. Yang, "Greener plug-in hybrid electric vehicles incorporating renewable energy and rapid system optimization," *Energy*, vol.111, pp.971-980, 2016.

[16] V. Sezer, M. Gokasan, and S. Bogosyan, "A novel ECMS and combined cost map approach for high-efficiency series hybrid electric vehicles," *IEEE Trans. Veh. Technol.*, vol.60, no.8, pp.3557-3570, 2011.

[17] M. Morari, and J. H. Lee, "Model predictive control: past, present and future," *Computers & Chemical Engineering*, vol.23, no.4, pp.667-682, 1999.

[18] S. Onori, L. Serrao, and G. Rizzoni, "Adaptive equivalent consumption minimization strategy for hybrid electric vehicles," *In: Proc. ASME 2010 Dynamic Systems and Control Conference*., pp.499-505, 2010.

[19] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ECMS: An adaptive algorithm for hybrid electric vehicle energy management," *Eur J Control*., vol.11, no.4, pp.509-24, 2005.

[20] C. Zhang, and A. Vahidi, "Route preview in energy management of plug-in hybrid vehicles," *IEEE Control Syst. Tech.*, vol.20, no.2, pp.546-553, 2012.

[21] C. Manzie, O. Grondin, and A. Sciarretta, "Robustness of ECMS-based optimal control in parallel hybrid vehicles," *IFAC Proceedings Volumes*., vol.46, no.21, pp.127-132, 2013.

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

[23] D. Rotenberg, V. Vahidi, and I. Kolmanovsky, "Ultracapacitor assisted powertrains: Modeling, control, sizing, and the impact on fuel economy," *IEEE Control Syst. Tech.*, vol.19, no.3, pp.576-589, 2011.

[24] 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 Control Syst. Tech.*, vol.20, no.3, pp.593-603, 2012.

[25] S. Vazquez, J. I. Sergio, L. G. Franquelo, J. Rodriguez, H. A. Young, "Model predictive control: A review of its applications in power electronics," *IEEE Trans. Ind. Mag.*, vol.8, no.1, pp.16-31, 2014.

[26] C. Sun, X. S. Hu, S. J. Moura, and F. C. Sun, "Velocity predictors for predictive energy management in hybrid electric vehicles," *IEEE Control Syst. Tech.*, vol.23, no.3, pp.1197-1204, 2015.

[27] I. Arsie, M. Graziosi, C. Pianese, G. Rizzo, and M. Sorrentino, "Optimization of supervisory control strategy for parallel hybrid vehicle with provisional load estimate," *in Proc. AVEC*, 2004, pp.23-27.

[28] 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 Control Syst. Tech.*, vol.19, no.3, pp.545-555, May 2011.

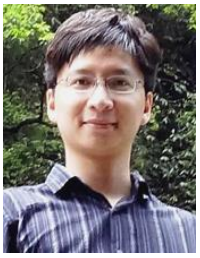
[29] C. Dextreitand I. V. Kolmanovsky, "Game theory controller for hybrid electric vehicles," *IEEE Control Syst. Tech.*, vol.22, no.2, pp.652-63, 2014.

[30] C. Liu, and Y. L. Murphey, "Power management for plug-in hybrid electric vehicles using reinforcement learning with trip information," *In: Proc. Transportation Electrification Conference and Expo (ITEC)*, 2014,

pp.1-6.
[31] X. Lin, Y. Wang, P. Bogdan, N. Chang, and M. Pedram, "Reinforcement learning based power management for hybrid electric vehicles," *In: Proc. IEEE/ACM International Conference*, 2014, pp.33-38.
[32] H. Chen, J. Kessels, and M. C. F. Donkers, "Game-theoretic approach for complete vehicle energy management," *In: Proc. IEEE Vehicle Power and Propulsion Conference (VPPC)*, 2014, pp.1-6.
[33] X. W. Qi, G. Wu, K. Boriboonsomsin, and J. B. Matthew, "A Novel Blended Real-Time Energy Management Strategy for Plug-in Hybrid Electric Vehicle Commute Trips," *In: Proc. IEEE 18th International Conference on Intelligent Transportation Systems*, 2015, pp.1002-1007.
[34] T. Liu, Y. Zou, D. X. Liu, and F. C. Sun, "Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle," *IEEE Trans. Ind. Electron.*, vol.62, no.12, pp.7837-7846, 2015.
[35] Y. Zou, T. Liu, D. X. Liu, and F. C. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Applied Energy*, vol.171, pp.372-382, 2016.
[36] Y. Zou, T. Liu, F. C. Sun, and H. Peng, "Comparative Study of Dynamic Programming and Pontryagin's Minimum Principle on Energy Management for a Parallel Hybrid Electric Vehicle," *Energies*, vol. 6, pp.2305-2318, 2013.
[37] B. Geng, J. K. Mills , and D. Sun, "Energy management control of microturbine-powered plug-in hybrid electric vehicles using the telemetry equivalent consumption minimization strategy," *IEEE Trans. Veh. Technol.*, vol.60, no.9, pp.4238-4248, 2011.
[38] V. H. Johnson, "Battery performance models in ADVISOR," *J.Power Sources*, vol. 110, no. 2, pp.321-329, 2002.
[39] J. M. Liu, and H. Peng, "Modeling and control of a power-split hybrid vehicle," *IEEE Control Syst. Tech.*, vol.16, no.6, pp.1242-1251, 2008.
[40] D. P. Filevand I. Kolmanovsky, "Generalized markov models for real-time modeling of continuous systems," *IEEE Trans. Fuzzy. Syst.*, vol.22, pp.983-998, 2014.
[41] L. Johannesson, M. Asbogard, and B. Egardt, "Assessing the potential of predictive control for hybrid vehicle powertrains using stochastic dynamic programming," *IEEE Trans. Intell. Transp. Syst*.,vol.8, no.1,pp. 71-83, 2007.
[42] D. P Filev and I. Kolmanovsky, "Markov chain modeling approaches for on board applications," *In: Proc.IEEE American Control Conference*, 2010, pp.4139-4145.
[43] G. Grimmet and Stirzaker D, "Probability and Random Processes," London: Oxford University Press, 2004.
[44] R. C. Hsu, C. T. Liu, and D. Y. Chan, "A Reinforcement-learning-based assisted power management with QoR provisioning for human–electric hybrid bicycle," *IEEE Trans.Ind. Electron.*, vol.59, no.7, pp.3350-3359, 2012.
[45] V. Mnih, K. Kavukcuoglu, and K. Silver, "Human-level control through deep reinforcement learning," *Nature.*, vol.518, no.7540, pp.529-533, 2015.
[46] K. Etessami, and M. Yannakakis, "Recursive markov decision processes and recursive stochastic games," *Journal of the ACM (JACM)*, vol.62, no.2, pp.11, 2015.

Teng Liu received the B.S. degree in mathematics from Beijing Institute of Technology, Beijing, China, 2011. He received his Ph.D. degree in the vehicle engineering from Beijing Institute of Technology (BIT), Beijing, in 2017. His Ph.D. dissertation, under the supervision of Dr. Fengchun Sun, was entitled "Reinforcement learning based energy management for hybrid electric vehicles." He is currently a post doctorate in the center for automotive engineering, Cranfield university, UK.

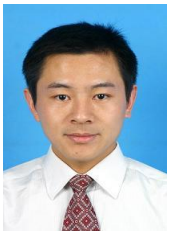
Dr. Liu has more than 6 years' research and working experience in new energy vehicle and control, where he has contributed over 15 papers. His research interests include modeling, control, and optimization of sustainable transport systems, as well as decision making for automated vehicle.



Xiaosong Hu (SM'16) received the Ph.D. degree in Automotive Engineering from Beijing Institute of Technology, China, in 2012.

He did scientific research and completed the Ph.D. dissertation in Automotive Research Center at the University of Michigan, Ann Arbor, USA, between 2010 and 2012. He is currently a professor at the State Key Laboratory of Mechanical Transmissions and at the Department of Automotive Engineering, Chongqing University, Chongqing, China. He was a postdoctoral researcher at the Department of Civil and Environmental Engineering, University of California, Berkeley, USA, between 2014 and 2015,as well as at the Swedish Hybrid Vehicle Center and the Department of Signals and Systems at Chalmers University of Technology, Gothenburg, Sweden, between 2012 and 2014. He was also a visiting postdoctoral researcher in the Institute for Dynamic systems and Control at Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, in 2014. His research interests include modeling and control of sustainable energy systems, including energy storage, electrified vehicles, and automated vehicles.

Dr. Hu has been a recipient of several prestigious awards/honors, including Emerging Sustainability Leaders Award in 2016, EU Marie Currie Fellowship in 2015, ASME DSCD Energy Systems Best Paper Award in 2015, and Beijing Best Ph.D. Dissertation Award in 2013.



Shengbo Eben Li received the M.S. and Ph.D. degrees from Tsinghua University in 2006 and 2009.

He worked as a visiting scholar at Stanford University in 2007, a postdoctoral research fellow in University of Michigan from 2009 to 2011, and a visiting professor in University of California, Berkeley, in 2015. His active research activities include autonomous vehicle control, driver behaviors and modeling, control topics of battery, optimal control and multi-agent control, etc. He is the author of more than 80 journal/conference papers, and the co-inventor of more than 20 patents.

Dr. Li was the recipient of the Distinguished Doctoral Dissertation of Tsinghua University (2009), Award for Science and Technology of China ITS Association (2012), Award for Technological Invention in Ministry of Education (2012), National Award for Technological Invention in China (2013), Honored Funding for Beijing Excellent Youth Researcher (2013), NSK Sino-Japan Outstanding Paper Prize in Mechanical Engineering (2014/2015), Best Student Paper Award in 2014 IEEE Intelligent Transportation System Symposium (as student advisor), Top 10 Distinguished Project Award of NSF China (2014), Best Paper Award in 14th ITS Asia Pacific Forum, 2015. He also served as the Associate editor of IEEE Intelligent Vehicle Symposium (2012/2013), Chairman of organization committee of China ADAS forum (2013).

Dongpu Cao(M'08) received the Ph.D. degree from Concordia University, Canada, in 2008.

He is currently a Lecturer at Advanced Vehicle Engineering Centre, Cranfield University, UK. His research focuses on vehicle control and intelligence, where he has contributed more than 100 publications and 1 US patent. He received the ASME AVTT'2010 Best Paper Award and 2012 SAE Arch T. Colwell Merit Award.

Dr. Cao serves as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL ELECTRONICS, and ASME Journal of Dynamic Systems, Measurement, and Control. He has been a Guest Editor for IEEE/ASME TRANSACTIONS ON MECHATRONICS, VEHICLE SYSTEM DYNAMICS, and IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS. He serves on the SAE International Vehicle Dynamics Standards Committee and a few ASME, SAE, IEEE technical committees.