IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 27, 2021, accepted May 9, 2021, date of publication May 13, 2021, date of current version May 21, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3079903

A Real-Time Intelligent Energy Management Strategy for Hybrid Electric Vehicles Using Reinforcement Learning

WOONG LEE[1], HAESEONG JEOUNG[1], DOHYUN PARK[1], (Graduate Student Member, IEEE), TACKSU KIM[1], HEEYUN LEE[2], AND NAMWOOK KIM[1]

[1]Department of Mechanical Engineering, Hanyang University, Ansan 15588, Republic of Korea
[2]Department of Mechanical Engineering, Seoul National University, Seoul 08826, Republic of Korea

Corresponding author: Namwook Kim (nwkim21@gmail.com)

This work was supported by the Technology Innovation Program (Development of application technologies for heavy duty fuel cell electric trucks using multi-input motor based 400kW class multi speed electrified powertrain system) through the Ministry of Trade, Industry and Energy (MOTIE), South Korea, under Grant 20011834.

ABSTRACT Equivalent Consumption Management Strategy (ECMS), a representative energy management strategy for hybrid electric vehicles (HEVs) derived from Pontryagin's minimum principle, is known to produce a near-optimal solution if the costate or equivalent factor of electric use is appropriately determined according to the driving conditions. One problem when applying the control concept to real-world scenarios is that it is difficult to precisely evaluate the performance of the control parameter before driving is complete, so the costate cannot be determined properly. To address this issue, this study proposes a practical method for estimating an appropriate costate based on Deep Q-Networks (DQNs), which is a reinforcement learning algorithm that uses a Deep Neural Network to evaluate the performances and determine the best control parameter or costate. The control concept benefits vehicle energy management by selecting the control parameter most related to stochastic conditions or future driving information based on artificial intelligence (AI), while optimal control is deterministically conducted by ECMS if the control parameter is given. Simply, only the implicit part of the optimal controller is solved via artificial intelligence. In the simulation results, not only does the proposed control concept outperform an existing ECMS that uses an adaptive technique for determining the costate, but the concept is also very feasible, in that it does not need a model for evaluating the performances.

INDEX TERMS Energy management strategy, adaptive ECMS, machine learning, reinforcement learning, hybrid electric vehicles, deep Q-learning, optimal control.

I. INTRODUCTION

Extensive research and development to reduce energy consumption based on alternative vehicle technologies have been conducted over the last few decades. In particular, Fuel Cell Electric Vehicles (FCEVs) and Battery Electric Vehicles (BEVs) that can realize zero emissions have become popular. However, it will take time to replace all conventional vehicles with zero-emission vehicles because infrastructure such as hydrogen and electric charging stations is necessary; additionally, petroleum is still a cost-competitive energy solution. Therefore, by taking advantage of partial electrification, Hybrid Electric Vehicles (HEVs) may be promising solutions not only for saving fuel usage but also for penetrating automotive markets. It is known that HEVs could improve fuel economy by up to 50% or more by using electric motors, but this requires sophisticated control for maximizing the use of electric components.

Many control concepts for balancing energy and managing powertrain components have been proposed, and the Equivalent Consumption Minimum Strategy (ECMS) shows outstanding fuel-saving performances [1], [2]. The ECMS defines the sum of fuel and electricity consumption as an equivalent value, and then it finds the control inputs that

The associate editor coordinating the review of this manuscript and approving it for publication was Chao Yang.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

minimize this consumption at every moment. It turns out that this energy management strategy can be interpreted in terms of the optimal control, as realized through Pontrygin's Minimum Principle (PMP), so it is possible that near-optimal solutions can be obtained based on ECMS [3], [4]. In this control concept, an equivalent factor called the costate is used to evaluate the energy consumption, which determines the relative value of the electric energy. For optimal energy management, an appropriate costate should be determined according to the driving conditions, which is related to future driving information. Assuming that future driving information is fully known, a method to find the ideal costate using a numerical approach (e.g., Newton Raphson, Shooting Method) was studied [5]. However, such methods are difficult to apply to real applications because it finds an ideal costate through iterative simulations. Another approach for a practical control concept is selecting an appropriate costate based on the current vehicle status [6],[7]. For instance, an adaptive ECMS that updates the costate based on current SOC levels showed the best performance for HEVs in the IEEE VTS Motor Vehicles Challenge 2018 [8]. On the other hand, given that sensors and communication technologies have advanced in recent years, vehicles can predict and utilize future driving information for better control. Therefore, there is need for an intelligent method that can determine costates in real-time by utilizing future driving information.

Reinforcement learning is a specialized field in machine-learning control optimization, and the control policy is reinforced to maximize the total rewards of the system while the environment and the agent interact with each other [9]. Reinforcement learning enables model-free control, which makes it possible to deal with control problems even when no information is provided regarding the environment.

The traditional method, model-based control, can be used when the correlations between the states and the output of the system are known, and an absolute solution can be obtained if the model is deterministic because reinforcement learning is based on Bellman's optimality [10]. Model-free control relies on the experience of the model to learn, and the evaluation of the reward can be conducted through Monte Carlo (MC) learning or Temporal Difference (TD) learning algorithms [11], [12]. The MC learning algorithm grants rewards to the agent collectively after the episode is over. Therefore, its convergence is excellent, but it is not suitable for online learning. On the other hand, the TD learning algorithm assigns rewards to the agent every moment, so the convergence may be lower than that of the MC algorithm, but this makes the online learning process possible.

As a representative algorithm, Q-learning, which is based on TD learning methods, is used in this study to estimate the costate of ECMS. The policy in Q-learning is determined based on a Q-table, where optimal rewards are updated according to states and actions [13]. If the Q-table is saturated, it is possible to utilize the table to select the best control option. However, the dimensions of the state and action are often too large to be effectively handled by the controller. Therefore, deep Q-learning, which approximates the Q-table using deep neural networks, has been used in this study [14]–[17].

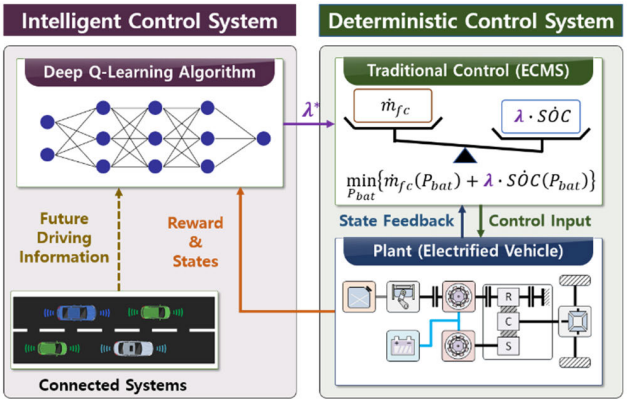


FIGURE 1. The real-time intelligent energy management strategy based on ECMS.

Whereas Q-learning is already becoming popular for HEV control problems [18], [19], we propose an efficient control concept that combines the optimal control strategy with the reinforcement learning algorithm, as shown in Fig. 1. The organization of the proposed control is divided into two parts, a stochastic part and a deterministic part. The stochastic, or intelligent, control part estimates the optimal costate based on Q-learning, where future driving information is the input status of the learning algorithm. The future driving information concerns what can be predicted from the connected systems, such as average vehicle power demand, average vehicle speed and average vehicle acceleration. Such control cannot be deterministic because there is no available perfect future prediction. In the deterministic control part, the powertrain components are managed by the ECMS for saving equivalent consumption. This control concept is very efficient because it can be applied to a real-time controller, and optimality of the EMCS is still guaranteed in the deterministic part, which means that the stochastic information related to future driving is handled in the intelligent part only.

This paper is organized as follows: Section II introduces multi-mode hybrid systems and ECMS. Section III proposes two methodologies to estimate the optimal costate in the controller. Section IV evaluates the performances of the two control concepts. Finally, in Section V, conclusions derived from this study are presented.

II. ENERGY MANAGEMENT STRATEGY BASED ON OPTIMAL CONTROL

This section introduces a multi-mode hybrid system used in this study and an energy management strategy for the vehicle.

A. MULTI-MODE HYBRID SYSTEM

There are several hybrid systems that have been introduced by global manufacturers. Toyota has introduced power-split hybrid systems (e.g., the Prius series) using planetary gear

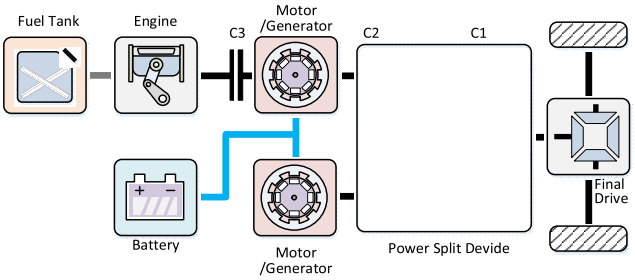


FIGURE 2. The powertrain system configuration of volt 1st Gen.

TABLE 1. Operating modes of volt 1st Gen.

Operating Modes	C1	C2	C3
EV1	O		
EV2		O	
Series	O		O
Output Split		O	O

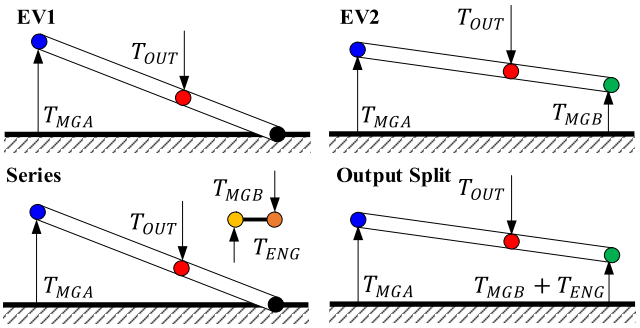


FIGURE 3. Operating modes of volt 1st Gen characterized by a lever system.

sets [20]–[22], and Hyundai Motors has been developing Transmission-Mounted Electric Device (TMED) systems (e.g., Ioniq, Sonata) [23], [24]. General Motors has introduced multi-mode hybrid systems (e.g., Voltec, Malibu, Cadillac CT6) that can implement several operating modes with clutches and planetary gears [25]–[27]. This system can selectively use an optimal operating mode according to the driving conditions. Fig. 2 shows the powertrain system of GM Volt 1st Gen and it is an Extended Range Electric Vehicles (EREV) with plug-in. This system can realize two EV modes, a series mode and an output split mode, with one planetary gear and three clutches as shown in Table 1 and Fig. 3. The EV1 mode is operated at low speed and the series mode is activated when SOC charging is required. The vehicle drives in either EV2 or output split mode at high speeds. This hybrid system is selected for comparative study because this vehicle model was present in the IEEE VTS Motor Vehicles Challenge 2018 [8], and the results of this competition can serve as a reference for evaluating the control performance using the reinforcement learning algorithm.

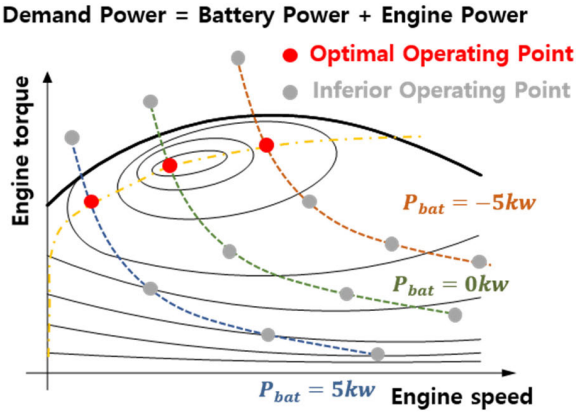


FIGURE 4. Engine power depending on the battery power.

B. THE CONCEPT OF ECMS

Drivers control the acceleration and brake pedals according to the driving situation ahead, including traffic information, signal lights, or speed limits, and the demand power for driving is determined from these pedal signals. The engine, which is the only power source in a conventional vehicle, must provide the demand power requested by the driver.

HEVs, however, have multiple power sources, and thus there are many control options to distribute the demand power between the motors and engine. Fig. 4 and Fig. 5 show the process of determining optimal candidates for engine operating points in HEVs when the vehicle speed and demand power are given. First, if battery power is selected, the engine can choose any of the points on the blue, green, or brown dashed lines seen in Fig. 4, where particular lines indicate the output power is equal to the demand power. However, the best control option is given by the yellow dot-dash Optimal Operating Line (OOL) because it contains the most efficient engine operating points out of the available options. Fig. 5 shows all available engine operating points and the best operating points with respect to battery power, which is obtained by projecting all operating points in Fig. 4 into a different space in Fig. 5, with axes defined by the battery power and the fuel consumption rate. In terms of fuel economy, if the battery power is determined, the controller should use the minimum point to save on fuel usage, as described by the Pareto frontier in Fig. 5 [3]. On the other hand, the ECMS concept is derived from the idea of minimizing both battery and engine power at the same time.

This concept can be interpreted in terms of optimal control theory and can be expressed as the following equations [3].

$$minJ = \int_{t_0}^{t_f} g\left(P_{bat}, t\right) dt, \tag{1}$$

$$\dot{SOC} = f\left(SOC, P_{bat}\right), \tag{2}$$

$$SOC_{init} = SOC_{final} \tag{3}$$

Here, g is the fuel consumption rate, P_{bat} is the battery power and SOC is the state of charge of the battery. The cost

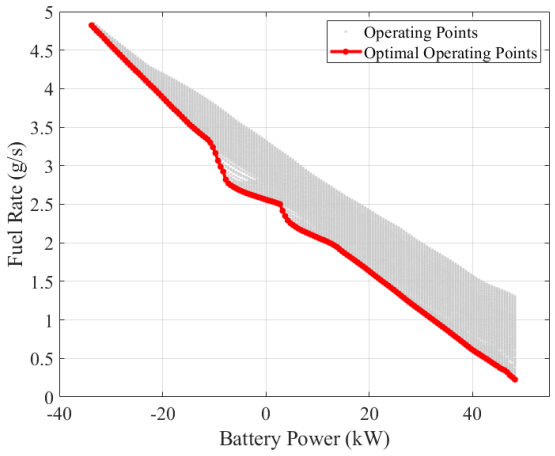


FIGURE 5. Pareto Frontiers for GM volt 1st when the transmission output torque and speed are 200Nm & 200rad/s.

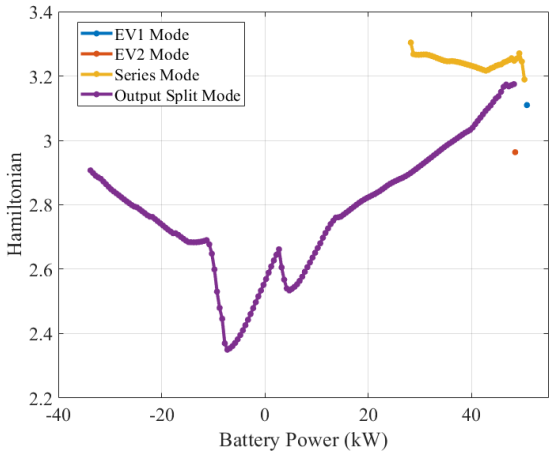


FIGURE 6. Hamiltonian (Output Torque: 200Nm, Wheel Speed: 200rad/s, Costate: -2400g/s). The point that minimizes the Hamiltonian in output split mode is −7kW.

function and state equation can be combined with λ, called the costate, based on constrained optimization, as follows [3]:

$$H = g\left(P_{bat}, t\right) + \lambda \cdot S\dot{O}C \tag{4}$$

where H is the Hamiltonian, which can be interpreted as equivalent consumption. This is obtained based on the Pareto frontier with a specific costate value, and an example of the Hamiltonian is shown in Fig 6. There can be up to four Hamiltonian lines available for a specific output torque and wheel speed because GM VOLT 1st has four different operating modes. For instance, the output split mode with the battery power of -7kW is the optimal point, which is superior to any other operating option. As described above, the main concept of ECMS is to calculate the Hamiltonian with a specific costate value at every moment, so the costate becomes a significantly important parameter. The problem is, however, that the costate is not given prior to driving because it is related to the electric energy, which itself depends on the driving conditions. Therefore, the controller should have methodologies to appropriately estimate the costate.

III. COSTATE ESTIMATION

This section explains the meaning of the costate and proposes a method for estimating it. In this study, an adaptive control concept based on current driving information and an intelligent control technique based on reinforcement learning using short-term future driving information are introduced, where the former was introduced in the 2018 IEEE VTS Challenge [8]. The performances of the two control strategies will be compared via simulation results.

A. IMPACT OF COSTATE

The impact of the costate on energy management has been an interesting topic in HEV control problems [28], [29]. As mentioned in Section II, the costate plays a role in determining the relative value of electric use. If the absolute value of the costate is high, the SOC will be charged as electricity

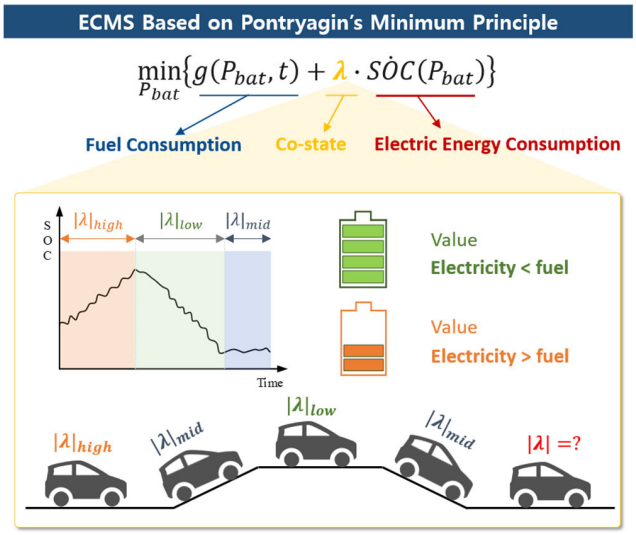


FIGURE 7. Costate meaning according to the SOC and future driving information.

becomes more valuable, and the SOC will be discharged otherwise. Additionally, when the absolute value of the costate is appropriately determined, the energy use is properly distributed between the fuel and the electricity, so the SOC is well-balanced. This balancing task and the costate are related to future driving information. For example, recognizing an uphill load ahead, the controller can predict that the vehicle will need high power for climbing the hill during the next few minutes. Therefore, a possible control strategy is to charge the SOC in advance by setting the absolute value of the costate to high, so the motors can assist the engine on the uphill road. Conversely, if the vehicle knows that there is a downhill road ahead, it consumes the SOC in advance, so the vehicle can fully recuperate its braking energy when it arrives at the downhill road. As shown in Fig. 7, the costate can be adjusted to maximize the efficiency by observing future driving information

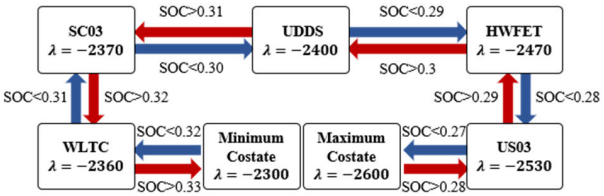


FIGURE 8. A concept of costate estimation of the adaptive energy management strategy.

B. ADAPTIVE ENERGY MANAGEMENT STRATEGY

The IEEE Vehicular Technology Society (VTS) held the IEEE VTS Challenge 2018 with the aim to minimize fuel and electric consumption of GM Volt 1st under the assumption that future driving information is unknown. In this competition, an adaptive energy management strategy based on optimal control using only the current driving information was developed, showing the best performance [8]. Although the costate can be appropriately estimated by considering future driving information and the current SOC, the proposed controller in [8] utilized only the current SOC because future information was not available. The adaptive energy management strategy updates the costate by selecting one of five optimal costate values that appropriately manage the SOC during representative driving cycles, such as UDDS, HWFET, SC03, US06, and WLTC. In case the controller fails to manage the SOC with the 5 costate values, an emergency state is activated to restore the SOC to a normal operating range, as shown in Fig. 8. Based on this concept, adaptive ECMS is expected to improve the fuel economy of the vehicle and properly manage the SOC level of the battery. This was verified when this adaptive ECMS showed the best performance out of other controllers in the competition [8].

C. INTELLIGENT ENERGY MANAGEMENT STRATEGY

Among machine learning techniques, reinforcement learning is a specialized concept in control optimization that reinforces the control policy to maximize the rewards obtained from the environment while the agent and the environment continually interact. There are various reinforcement learning algorithms, which can be classified into the two categories: model-based control and model-free control.

As mentioned in the introduction section, there are two types of model-free control: MC learning method and TD learning method. In this study, the online controller using TD learning algorithm is selected because i) a practical control algorithm should be able to estimate and update the costate in real-time and ii) a high-fidelity vehicle model is available to evaluate the energy consumption in the online controller. In this study, the intelligent controller in Fig. 1 using the TD learning algorithm implements the online learning process by observing the reward without information about the vehicle model.

Reinforcement learning calculates rewards while observing states and actions, and selects controls that can maximize

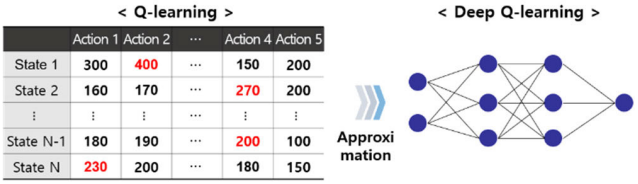


FIGURE 9. A concept of Q-learning and deep Q-learning.

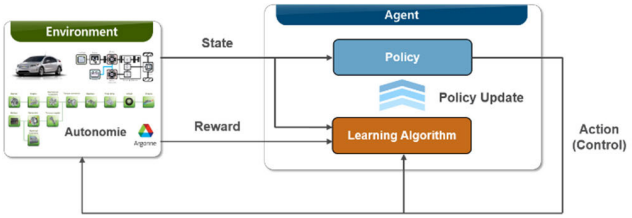


FIGURE 10. Reinforcement learning framework with Autonomie.

long-term rewards, Q, as follows:

$$Q(s,a) = E_{\pi}[r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots | s_t = s, a_t = a] \tag{5}$$

where r is a reward, γ is a discount factor, s is the state, a is the action. The Q-learning algorithm calculates this long-term reward every moment as follows:

$$Q(s,a) = r + \gamma \max_{a} Q(s', a) \tag{6}$$

where s′ is the next state. The Q-learning algorithm updates the policy to improve the performance based on the Q-table. Here, a neural network model is used for the Q-table because the states inhabit a large space and the actions are too numerous to be implemented in a table, which is why the Deep Q-Network (DQN) is used, as shown in Fig. 9. Deep Q-learning, simply, updates the policy by minimizing the loss function, which is defined as follows:

$$L_i(\theta_i) = \mathbb{E}_{(s,a)}\left[\left(r + \gamma \max_{a} Q(s', a'; \theta_{i-1}) - Q(s, a; \theta_i)\right)^2\right] \tag{7}$$

where θ is the weighting factor of deep neural networks. The loss function represents a measure of the deep neural network performance, and the deep Q-learning algorithm updates the weighting factor to minimize this loss function, which is used for estimating the costate in this study [31], [32]. If the loss is 0, the equation (7) becomes the same as equation (6), which means that ideally deep Q-learning can achieve the same performance as Q-learning despite having an approximated Q-table. Fig. 10 shows the reinforcement learning framework applied in this study, and a high-fidelity model built in Autonomie, a performance analysis tool, developed by Argonne National Laboratory (ANL), which was used for the environment [33]. The states include current driving information and short-term future driving information as follows:

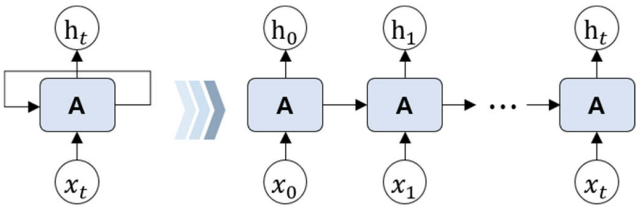


FIGURE 11. The recurrent neural network.

- Current battery SOC
- Current wheel torque demand
- Current costate value
- Predicted average power demand (10s)
- Predicted average vehicle speed (10s)
- Predicted average acceleration (10s)

The predicted average power demand is the sum of the average acceleration power demand and driving loss power. This study was conducted under the assumption that short-term future driving information (10s) can be accurately predicted, but the average values were used to make it easier to predict the future when using the prediction model [34], [35]. In the ECMS, based on optimal control, when the costate is determined, the equivalent consumption of electric use is estimated and optimization is performed to minimize the total consumption. Optimality can be guaranteed if the SOC is appropriately managed by the selected costate [4]. Therefore, the reward from Q-learning is defined to help manage the SOC in the desired range, which is expressed as:

$$r = -\left(SOC_{current} - SOC_{desire}\right)^2 / D \tag{8}$$

$$SOC = \frac{1}{C_b}\int_{t_0}^{t} i\, dt \tag{9}$$

where D is the total travel distance of the vehicle, which is used to normalize the performances in different driving cycles, and C_b is the battery capacity. The agent uses a deep neural network to approximate the Q-network. Alternatively, a solution to the Q-table that selects the action with the maximum expected reward while observing the states can be used, where a Recurrent Neural Network (RNN) is used for the network model. The RNN includes a hidden state with memory functions, so it handles sequential problems such as time-series simulations well, as shown in Fig. 11 [31]. The hidden states and output can be expressed as follows

$$h_t = f_{w_{x,h}}\left(h_{t-1}, x_t\right) \quad \text{and } y_t = f_{w_y}(h_t) \tag{10}$$

where h is a hidden state, x is the input vector and y is the output vector. The deep neural networks based on the RNN are designed by using the parameters, as shown in Table 2 and Fig. 12. A sequence input layers, two fully connected layers, two relu layers, a long short-term memory model (LSTM) layer and an output layer are used. The sequence input layer inputs sequence data to the network. The fully connected layers and output layer multiplies the input by a weight matrix and then adds a bias vector. The LSTM layer, one of the RNN, learns the long-term dependencies between the time steps of the time series and the sequence data. The relu layer, an activation function, performs a threshold operation on each element of the input, and values less than zero are set to zero. The activation function helps the model learn complex data by normalizing the output of each neuron. The action, control, is defined as the discretized costate values.

TABLE 2. Parameters of the deep neural networks.

Layers	Neurons
Sequence Input	16
Fully Connected	50
LSTM	20
Fully Connected	20
Output	1

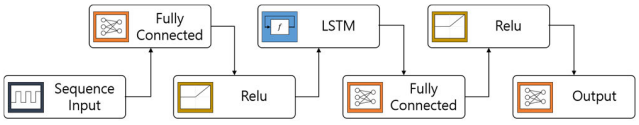


FIGURE 12. The deep Q-Network based on the RNN.

$$a = [-2800, -2790, \ldots, -2110, 2100] \tag{11}$$

Since the model-free control depends on its experience, it is important to explore the various paths the state can reach. If the optimal action is selected only by the deep Q-network, the state may only go to the path previously searched. To address this issue, the epsilon-greedy exploration method is used for this study. It randomly chooses the action with a specific probability, called epsilon, without depending on the deep Q-network, where epsilon ϵ can be expressed as follows:

$$\epsilon = \epsilon \cdot (1 - k) \quad \text{and } \epsilon \geq 0.01 \tag{12}$$

where k is the epsilon decay, and the minimum value for epsilon is set to 0.01 in this study. The simulation results for the controller using the deep Q-network are obtained based on these training conditions.

IV. SIMULATION

This section shows the process and results for the controller using the reinforcement learning framework introduced in the previous section. The control policy is, then, validated by testing untrained episodes, and its performance is compared to that of the controller using the adaptive costate concept.

A. PERFORMANCES OF INTELLIGENT CONTROLLER

To train the networks and obtain the control policy, the learning algorithm uses 1500 episodes of a UDDS cycle, where the parameters listed in Table 3 are used for the training.

TABLE 3. Parameters for reinforcement learning.

Parameters	Values
Discount Factor	0.99
Learning Rate	0.0005
Epsilon / Epsilon Decay	1 / 0.01
Reward (SOC_{desire})	0.3

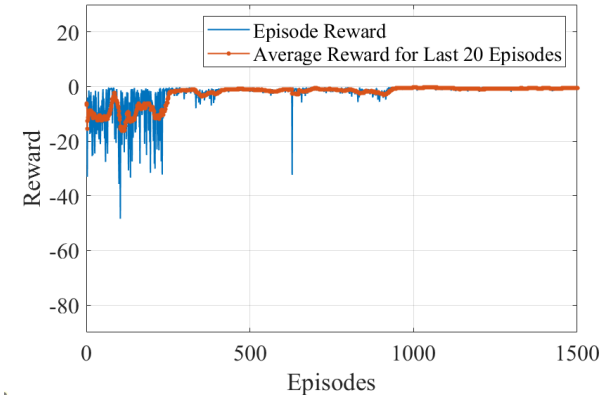
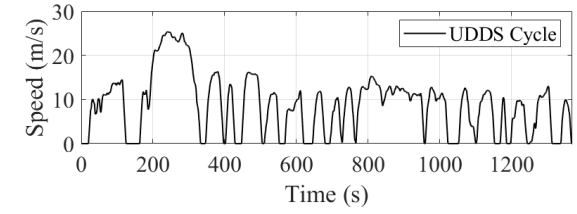
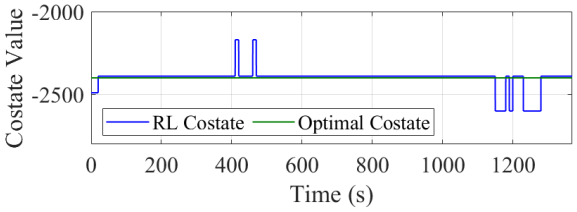


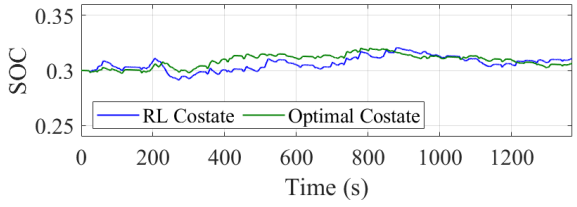
FIGURE 13. Training reward for one cycle (UDDS).



(a) Speed Profile: UDDS Cycle



(b) Costate Trajectories: Reinforcement Learning and Optimal Case



(c) SOC Trajectories: Reinforcement Learning and Optimal Case

FIGURE 14. Training results for the UDDS Cycle: (a) Speed Profile, (b) Costate Trajectories, (c) SOC Trajectories.

The change of reward is altered according to the progress of the training process, as shown in Fig. 13. The blue line represents the total reward per episode, and the orange line represents the average reward for the latest 20 episodes. The total reward is well-converged following the training process, and the intelligent controller using the deep Q-networks can be obtained via the learning algorithm, although there is still room for further improving the training performance by optimizing the learning parameters.

The performance of the intelligent controller is evaluated by comparing its simulation results with the results obtained by a controller using an optimal costate, where the optimal costate is determined by an iterative process to satisfy the boundary condition of the SOC [5]. Since most plug-in hybrids balance SOC around 0.3, the SOC boundary is defined as 0.3. Fig. 14 shows the training results for the UDDS cycles. In Fig. 14 (b), it was found that the intelligent controller, which implements reinforcement learning, estimates the costate very well, with an estimate similar to the optimal value for SOC balancing. Therefore, the SOC is appropriately managed by the intelligent controller, in which the controller does not only exceed the limits of the SOC range, but also produces a trajectory of the SOC that is similar to the optimal SOC, as shown in Fig. 14 (c).

B. MULTIPLE DRIVING SCENARIOS AND VALIDATION

The feasibility and performance of the intelligent controller produced from reinforcement learning are evaluated over

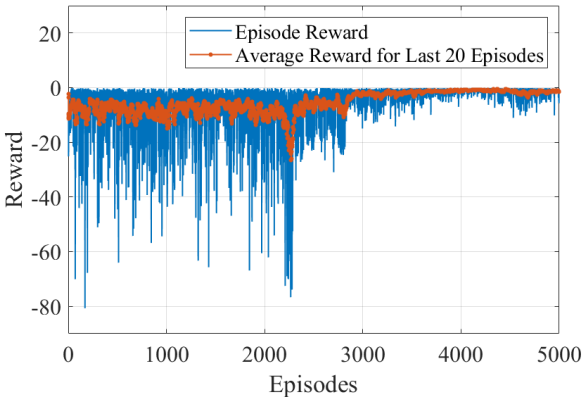


FIGURE 15. Training reward for multiple cycles (UDDS, HWFET, SC03, US06, WLTC).

multiple driving cycles by comparing the simulation results with the results obtained by the controller using the adaptive concept, introduced in Section III. B. The controller based on adaptive concepts is a good reference for comparison with the intelligent controller since such an adaptive controller showed the best performance and has been tested in unknown driving cycles during competition [8].

For this comparative study, the intelligent controller results are obtained by training the networks in 5 representative cycles, particularly UDDS, HWFET, SC03, US06, and

TABLE 4. Parameters for reinforcement learning.

Parameters	Values
Discount Factor	0.99
Learning Rate	0.0001
Epsilon / Epsilon Decay	1 / 0.01
Reward (SOC_{desire})	0.3

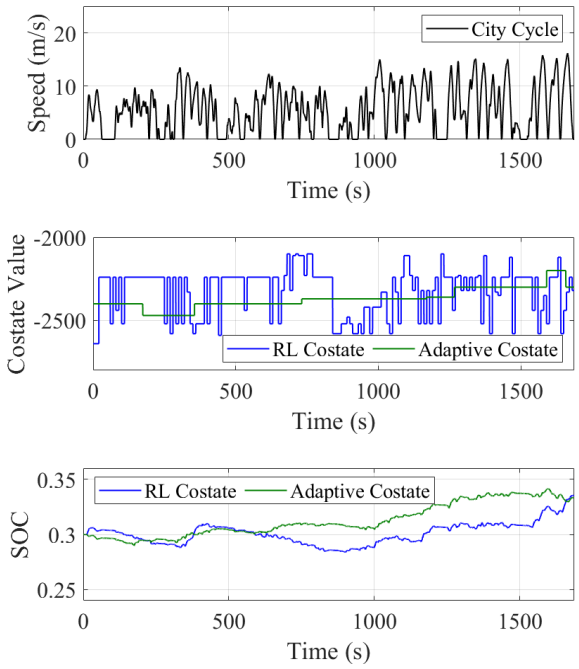


FIGURE 16. Validation results with untrained episodes (City cycle).

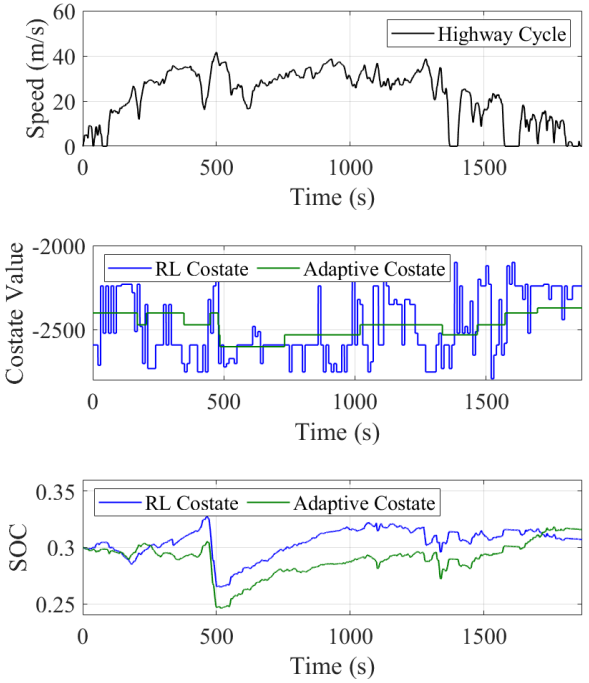


FIGURE 17. Validation results with untrained episodes (Highway cycle).

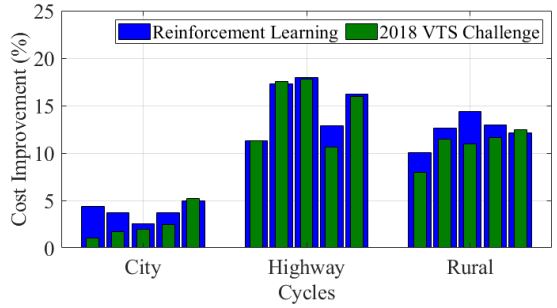


FIGURE 18. Cost improvement of the adaptive and intelligent energy management strategy compared to the rule-based control strategy from Autonomie.

WLTC, using the training parameters in Table 4. The learning rate represents how much the reward from the current episode is significant for estimating the final reward [15].

The smaller it is, the slower the learning speed is expected, however the stability of convergence increases. Training on multiple driving cycles that have different characteristics requires additional training episodes to ensure convergence of the reward, so a lower learning rate is empirically selected to obtain an appropriate control policy rather than using the learning rate for a single cycle in Table 3. In addition to the reward being well-converged in a single cycle in Fig. 13, the reward also appropriately converges over multiple cycles, as shown in Fig. 15, although it requires additional episodes for the training process. In the next step, driving cycles that are not used for the training process are selected for comparing the performances between the intelligent controller and the adaptive controller. First, the intelligent controller appropriately updates the costate and manages the SOC in the desired ranges, even in the untrained driving cycles, as shown in Fig. 16 and Fig. 17.

TABLE 5. Average cost improvement.

Control Strategies	City	Highway	Rural
Intelligent Energy Management Strategy (Reinforcement Learning)	3.86%	15.12%	12.43%
Adaptive Energy Management Strategy (2018 VTS Challenge)	2.48%	14.64%	10.89%

Simply, the intelligent controller is tested in 15 driving conditions by including 5 city, 5 rural, and 5 highway cycles, and the controller did not once fail to manage the SOC. Further, the performances of the intelligent controller using reinforcement learning are compared to those of the adaptive ECMS used in IEEE VTS Challenge 2018, as shown in Table 5 and

Fig. 18. The improvements of the two control concepts are estimated by using results produced by a baseline controller provided in Autonomie, where the baseline controller uses a rule-based concept for managing the operating modes and the component controls. The combined cost of gas and electricity was derived from the U.S. Energy Information Administration (IEA) in 2016 and the average electricity cost in 2016, which is the same as the method used in the competition [8].

$$Cost = gas_{cost} + electricity_{cost} \tag{13}$$

$$\begin{cases} gas_{cost} = 0.795\ US\ \$/kg. \\ electricity_{cost} = 0.137\ US\ \$/kWh. \end{cases} \tag{14}$$

In the comparative results, the intelligent controller shows better performances on average. Given that the controller using reinforcement learning is created based on a stochastic model, it is not possible that the intelligent controller is always superior to the other control concept—the other one is selected as the best controller in the competition. It is, however, expected that the intelligent controller would provide better performance than others if a sufficient number of tests are completed. Further, a control obtained from stochastic dynamic programming does not always guarantee superiority, but is considered as an absolute one if the stochastic mode is fixed [36], [37]. Based on the simulations, we have shown that intelligent, AI-based control could be a promising solution that can be implemented in real-world and real-time applications and which can produce outstanding performances.

V. CONCLUSION

Although the ECMS derived from an optimal control has shown excellent performances, it is not a very feasible solution because it is difficult to determine the key control parameter or costate, which is related to upcoming driving conditions. In this study, an intelligent control concept using deep Q-learning, which is a representative reinforcement learning algorithm, is proposed, so that the costate is estimated as soon as future driving information is available. The important feature of the control concept is that the stochastic part of the controller that determines the costate utilizes the reinforcement learning, whereas the deterministic part of the controller still relies on the optimal control concept. Based on the proposed concept, the fuel efficiency can be improved compared to the adaptive ECMS. These study results can be used to bring the benefits of reinforcement learning technology to the control problems of HEVs. A feasible control organization and development process were introduced in this study, and the performance of the control concept was evaluated via a comparative study. Here, the intelligent control concept showed better improvements in fuel economy, from 0.5 to 1.5% depending on the cycle, than the improvements of the adaptive concept.

REFERENCES

[1] A. Sciarretta, M. Back, and L. Guzzella, "Optimal control of parallel hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 12, no. 3, pp. 352–363, May 2004.

[2] M. Pourabdollah, V. Larsson, and B. Egardt, "PHEV energy management: A comparison of two levels of trip information," SAE Tech. Paper 2012-01-0745, Apr. 2012, doi: 10.4271/2012-01-0745.

[3] N. Kim, S. Cha, and H. Peng, "Optimal control of hybrid electric vehicles based on Pontryagin's minimum principle," *IEEE Trans. Control Syst. Technol.*, vol. 19, no. 5, pp. 1279–1287, Sep. 2011.

[4] N. Kim and A. Rousseau, "Sufficient conditions of optimal control based on Pontryagin's minimum principle for use in hybrid electric vehicles," *Proc. Inst. Mech. Eng., D, J. Automobile Eng.*, vol. 226, no. 9, pp. 1160–1170, Sep. 2012.

[5] N. W. Kim, D. H. Lee, C. Zheng, C. Shin, H. Seo, and S. W. Cha, "Realization of PMP-based control for hybrid electric vehicles in a backward-looking simulation," *Int. J. Automot. Technol.*, vol. 15, no. 4, pp. 625–635, Jun. 2014.

[6] S. Onori and L. Tribioli, "Adaptive Pontryagin's minimum principle supervisory controller design for the plug-in hybrid GM chevrolet volt," *Appl. Energy*, vol. 147, pp. 224–234, Jun. 2015.

[7] C. Sun, F. Sun, and H. He, "Investigating adaptive-ECMS with velocity forecast ability for hybrid electric vehicles," *Appl. Energy*, vol. 185, pp. 1644–1653, Jan. 2017.

[8] W. Lee, H. Jeoung, D. Park, and N. Kim, "An adaptive concept of PMP-based control for saving operating costs of extended-range electric vehicles," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11505–11512, Dec. 2019.

[9] F. L. Lewis and D. Vrabie, "Reinforcement learning and adaptive dynamic programming for feedback control," *IEEE Circuits Syst. Mag.*, vol. 9, no. 3, pp. 32–50, Aug. 2009.

[10] H. Lee, C. Song, N. Kim, and S. W. Cha, "Comparative analysis of energy management strategies for HEV: Dynamic programming and reinforcement learning," *IEEE Access*, vol. 8, pp. 67112–67123, 2020.

[11] M.-B. Radac, R.-E. Precup, and R.-C. Roman, "Model-free control performance improvement using virtual reference feedback tuning and reinforcement Q-learning," *Int. J. Syst. Sci.*, vol. 48, no. 5, pp. 1071–1083, Apr. 2017.

[12] D. Quillen, E. Jang, O. Nachum, C. Finn, J. Ibarz, and S. Levine, "Deep reinforcement learning for vision-based robotic grasping: A simulated comparative evaluation of off-policy methods," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 6284–6291.

[13] T. Liu, X. Hu, S. E. Li, and D. Cao, "Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle," *IEEE/ASME Trans. Mechatronics*, vol. 22, no. 4, pp. 1497–1507, Jun. 2017.

[14] Y. Hu, W. Li, K. Xu, T. Zahid, F. Qin, and C. Li, "Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning," *Appl. Sci.*, vol. 8, no. 2, p. 187, Jan. 2018.

[15] X. Liu, Y. Liu, Y. Chen, and L. Hanzo, "Enhancing the fuel-economy of V2I-assisted autonomous driving: A reinforcement learning approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 8, pp. 8329–8342, Aug. 2020.

[16] C. Yang, M. Zha, W. Wang, K. Liu, and C. Xiang, "Efficient energy management strategy for hybrid electric vehicles/plug-in hybrid electric vehicles: Review and recent advances under intelligent transportation system," *IET Intell. Transp. Syst.*, vol. 14, no. 7, pp. 702–711, Jul. 2020.

[17] E. Walraven, M. T. J. Spaan, and B. Bakker, "Traffic flow optimization: A reinforcement learning approach," *Eng. Appl. Artif. Intell.*, vol. 52, pp. 203–212, Jun. 2016.

[18] X. Qi, G. Wu, K. Boriboonsomsin, M. J. Barth, and J. Gonder, "Data-driven reinforcement learning-based real-time energy management system for plug-in hybrid electric vehicles," *Transp. Res. Rec.*, vol. 2572, no. 1, pp. 1–8, 2016.

[19] G. Du, Y. Zou, X. Zhang, Z. Kong, J. Wu, and D. He, "Intelligent energy management for hybrid electric tracked vehicles using online reinforcement learning," *Appl. Energy*, vol. 251, Oct. 2019, Art. no. 113388.

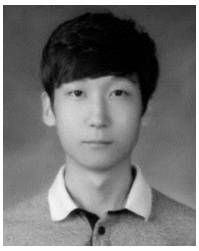
[20] M. Taniguchi, T. Yashiro, K. Takizawa, S. Baba, M. Tsuchida, T. Mizutani, H. Endo, and H. Kimura, "Development of new hybrid transaxle for compact-class vehicles," SAE Tech. Paper 2016-01-1163, Jun. 2016, doi: 10.4271/2016-01-1163.

[21] T. Furukawa, R. Ibaraki, H. Kimura, K. Kondo, M. Watanabe, T. Mizutani, H. Hattori, and A. Takasaki, "Development of new hybrid transaxle for sub-compact-class vehicles," SAE Tech. Paper 2012-01-0623, Dec. 2012, doi: 10.4271/2012-01-0623.

[22] S. Fushiki, "The new generation front wheel drive hybrid system," *SAE Int. J. Alternative Powertrains*, vol. 5, no. 1, pp. 109–114, Apr. 2016.

[23] W. Lee, T. Kim, J. Jeong, J. Chung, D. Kim, B. Lee, and N. Kim, "Control analysis of a real-world P2 hybrid electric vehicle based on test data," *Energies*, vol. 13, no. 16, p. 4092, Aug. 2020.
[24] H. Jeoung, K. Lee, and N. Kim, "Methodology for finding maximum performance and improvement possibility of rule-based control for parallel type-2 hybrid electric vehicles," *Energies*, vol. 12, no. 10, p. 1924, May 2019.
[25] M. A. Miller, A. G. Holmes, B. M. Conlon, and P. J. Savagian, "The GM 'Voltec' 4ET50 multi-mode electric transaxle," *SAE Int. J. Engines*, vol. 4, no. 1, pp. 1102–1114, 2011.
[26] X. Zhang, S. E. Li, H. Peng, and J. Sun, "Design of multimode power-split hybrid vehicles—A case study on the voltec powertrain system," *IEEE Trans. Veh. Technol.*, vol. 65, no. 6, pp. 4790–4801, Jun. 2016.
[27] W. Lee, J. Park, and N. Kim, "Analysis of transmission efficiency of a plug-in hybrid vehicle based on operating modes," *Int. J. Precis. Eng. Manuf.-Green Technol.*, vol. 7, pp. 1–11, Aug. 2019.
[28] F. Lacandia, L. Tribioli, S. Onori, and G. Rizzoni, "Adaptive energy management strategy calibration in PHEVs based on a sensitivity study," *SAE Int. J. Alternative Powertrains*, vol. 2, no. 3, pp. 443–455, Sep. 2013.
[29] J. Zhang, C. Zheng, S. W. Cha, and S. Duan, "Co-state variable determination in Pontryagin's minimum principle for energy management of hybrid vehicles," *Int. J. Precis. Eng. Manuf.*, vol. 17, no. 9, pp. 1215–1222, Sep. 2016.
[30] D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, "Mastering the game of go with deep neural networks and tree search," *Nature*, vol. 529, no. 7587, pp. 484–489, Jan. 2016.
[31] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. [Online]. Available: http://www.deeplearningbook.org
[32] X. Qi, Y. Luo, G. Wu, K. Boriboonsomsin, and M. Barth, "Deep reinforcement learning enabled self-learning control for energy efficient driving," *Transp. Res. C, Emerg. Technol.*, vol. 99, pp. 67–81, Feb. 2019.
[33] (2017). *Argonne National Laboratory, Autonomie*. [Online]. Available: https://www.autonomie.net
[34] B. Yao, C. Chen, Q. Cao, L. Jin, M. Zhang, H. Zhu, and B. Yu, "Short-term traffic speed prediction for an urban corridor," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 2, pp. 154–169, Feb. 2017.
[35] J. Lemieux and Y. Ma, "Vehicle speed prediction using deep learning," in *Proc. IEEE Vehicle Power Propuls. Conf. (VPPC)*, Oct. 2015, pp. 1–5.
[36] C.-C. Lin, H. Peng, and J. W. Grizzle, "A stochastic control strategy for hybrid electric vehicles," in *Proc. Amer. Control Conf.*, 2004, pp. 4710–4715.
[37] C. Yang, S. You, W. Wang, L. Li, and C. Xiang, "A stochastic predictive energy management strategy for plug-in hybrid electric vehicles based on fast rolling optimization," *IEEE Trans. Ind. Electron.*, vol. 67, no. 11, pp. 9659–9670, Nov. 2020.

WOONG LEE received the B.S. degree in mechanical engineering from Hanyang University ERICA Campus, Republic of Korea, in 2016, where he is currently pursuing the Ph.D. degree in mechanical engineering. He is currently working on vehicle system modeling and control optimization for electrified vehicles.



HAESEONG JEOUNG received the B.S. degree in mechanical engineering from Hanyang University ERICA Campus, Ansan, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in mechanical engineering. His research interests include system optimization and predictive control using machine learning for electrified vehicles.

DOHYUN PARK (Graduate Student Member, IEEE) received the B.S. degree in mechanical engineering from Hanyang University ERICA Campus, Ansan, South Korea, in 2017, where he is currently pursuing the Ph.D. degree in mechanical engineering. He is specialized for studies of vehicle electrification and control problems.

TACKSU KIM received the B.S. degree in mechanical engineering from Hanyang University ERICA Campus, Republic of Korea, in 2019, where he is currently pursuing the master's degree in mechanical engineering. He is currently studying deterministic and intelligent optimal control for vehicles.

HEEYUN LEE received the B.S. degree in mechanical engineering from Sungkyunkwan University, South Korea, in 2013, and the Ph.D. degree in mechanical engineering from Seoul National University, South Korea, in 2018. His research interests include optimal control, reinforcement learning, modeling, and simulation of electrified vehicles. He is currently with the Research and Development Division, Hyundai Motor Company, South Korea.

NAMWOOK KIM received the B.S. and Ph.D. degrees from Seoul National University, in 2003 and 2009, respectively. In 2009, he held a postdoctoral position with the Transportation Research Center, Argonne National Laboratory, and worked as a Research Engineer, from 2012 to 2015. He is currently working as an Associate Professor with Hanyang University, Republic of Korea.