Title
Reinforcement Learning for Hybrid and Plug-In Hybrid Electric Vehicle Energy Management: Recent Advances and Prospects

Journal
IEEE Industrial Electronics Magazine, 13(3)

ISSN
1932-4529 1941-0115

Authors
Hu, Xiaosong
Liu, Teng
Qi, Xuewei
et al.

Publication Date
2019-09-01

DOI
10.1109/MIE.2019.2913015

Peer reviewed

©ISTOCKPHOTO.COM/MRPLUMO

XIAOSONG HU, TENG LIU, XUEWEI QI, and MATTHEW BARTH

Reinforcement Learning for Hybrid and Plug-In Hybrid Electric Vehicle Energy Management

Recent Advances and Prospects

Digital Object Identifier 10.1109/MIE.2019.2913015
Date of publication: 24 September 2019

Energy management is a critical technology in plug-in hybrid-electric vehicles (PHEVs) for maximizing efficiency, fuel economy, and range, as well as reducing pollutant emissions. At the same time, deep reinforcement learning (DRL) has become an effective and important methodology to formulate model-free and real-time energy-management strategies for HEVs and PHEVs. In this article, we describe the energy-management issues of HEVs/PHEVs and summarize a variety of potential DRL applications for onboard energy management. We also discuss the prospects for various DRL approaches in the energy-management field.

 1932-4529/19©2019IEEE

Overview

Powertrain electrification, fuel decarbonization, and energy diversification are techniques that are spreading across the world, leading to cleaner and more efficient vehicles. HEVs/PHEVs have significantly increased fuel economy while maintaining extremely low emissions. To achieve these gains, it is critical that HEVs/PHEVs have sophisticated energy-management systems. PHEVs can operate in different modes, such as full electric and power split modes. These modes are controlled by the energy-management system under diverse driving conditions [1].

In general, an energy-management system adjusts and regulates the output power from multiple sources to fulfill the power request and minimize a predefined objective cost [2]. Battery packs, internal combustion engines (ICEs), and supercapacitors are often used as power sources. In practical applications, the optimization of energy consumption and performance is the primary concern and objective of a hybrid powertrain. Constructing appropriate and efficient energy-management strategies in a PHEV or an HEV is a challenging optimization problem that many researchers have investigated during the past decade.

Several published works have summarized the research progress in HEV/PHEV energy management. For example, Serrao et al. constructed a comparative analysis of three known global optimal algorithms [3], including dynamic programming, Pontryagin's minimum principle, and the equivalent consumption minimization strategy. In addition, [4] and [5] discussed the real-time and global optimization methods for energy management, such as game theory, the genetic algorithm, and model predictive control (MPC). Furthermore, Ganji et al. focused on studying predictive energy-management strategies, such as velocity and power demand prediction [6]. However, the research advance represented by learning methods in the HEV/PHEV energy-management field has not been significantly covered in the literature.

Recently, academic and industrial researchers have shown an increasing interest in learning-based energy management approaches that are founded in artificial intelligence. Machine learning is a prevalent and useful technique to address a variety of problems in many research fields [7]. The greatest challenges in HEV/PHEV energy management are shortening the counting process and improving adaptability, and they can be overcome by learning methods. Therefore, RL has become increasingly popular in the HEV/PHEV energy-management field. The technique can derive a model-free and instantaneous energy-management strategy for online application purposes.

In this article, a comprehensive survey of the recent progress in learning-based energy-management strategies is presented. In general, applications of RL methods to energy management can be classified into two categories. The first is "simplex algorithms," meaning that only a single algorithm is used to derive the energy-management policy, such as the Q-learning, Dyna, and Sarsa algorithms [8]. The second is "hybrid algorithms," indicating the commixture of other information or algorithms with RL, such as predictive algorithms, trip information, deep learning, and MPC [9]. These additional details are integrated into the RL framework to deduce more efficient and real-time controls. We summarize the application of RL approaches aimed at HEV/PHEV energy-management systems.

The remaining content of this article is organized as follows. First, the energy-management problem is sketched, with the vital optimization objective and constraints. Next, the different executions of RL methods, including the simplex and hybrid algorithms, are reviewed in their applications in HEVs/PHEVs. This article covers a variety of vehicle types and different techniques and compares their key performance measures. Finally, the future research prospects for RL-based energy-management systems are discussed.

The Energy-Management Problem

The core function of an energy-management system is balancing the power distribution among multiple onboard energy/power sources, with the goal of optimizing some cost functions, such as fuel consumption, battery life, pollution emissions, and driving mobility. This issue is usually formulated as an optimal control problem that has desired control objectives and particular physical constraints [10]. The control objectives may contain one or several options ranging from exhaust temperature, nitrogen oxide and sulfur oxides emissions, fuel consumption, shift frequency, the battery's state of charge (SOC) and state of health (SOH), and the cost of electricity. Figure 1 illustrates the energy-management problem for PHEVs.

The optimal control problem is often subject to three kinds of physical constraints: the powertrain dynamics, initial and final values of the state variables, and limitations on the control actions and state variables. Once the inputs (for example, power demand, vehicle velocity, the current SOC, and the steering angular speed) of this problem are provided in advance, the desired power from each energy source and the fuel cost can be calculated based on the powertrain dynamics. The battery SOC, position of the gearbox, and motor/generator speed are usually chosen as the state variables. The engine's output torque or the throttle position, the gear shifting, and the status of the clutches (in multimode HEVs, including the Toyota Prius and Chevrolet Volt) are often selected as the control actions. To solve this optimal control problem, limits for these parameters are necessary.

In addition to the control objective and constraints, an elaborate model of the powertrain components is necessary as part of the solution. For example, the modeling of an engine involves the calculation of the fuel consumption, an estimate of efficiency, and the derivation of the torque and angular speed. Also, a computation of efficiency and an expression of the power balance are required for motor modeling. The transfer process of the speed and power from the motor/generator to the final drive is part of the transmission modeling. The battery pack modeling incorporates the evolution of the SOC and the relationship of the open-circuit voltage and internal resistance with respect to the SOC.

Reinforcement Learning

In RL, an agent learns how to establish a mapping, from input states to optimal control actions, to maximize a cumulative reward as illustrated in Figure 2. The "learner" (for example, the vehicle controller) needs to discover which actions contribute to the largest reward, typically through a trial-and-error search process. Each action may influence the current and delayed rewards simultaneously. Sensing the states from the environment, taking particular actions, and achieving goal-directed rewards are pivotal steps in RL applications [11].

Three representative features belong to RL. The first is the coordination between exploitation and exploration. The agent utilizes exploration to acquire knowledge about the environment and applies exploitation to achieve a control action based on current knowledge. The ε-greedy algorithm is frequently utilized to balance exploitation and exploration. The second feature is that the environment in RL is sometimes uncertain. Hence, adaptive and model-free control actions are necessary. Owing to the interactions with the environment, the agent can recognize the states and choose suitable actions to affect the environment, which constitutes a grand picture of the states and actions. The third is the Markovian property of the environment, which means that the conditional probability distribution of the future states of the environment depends only upon the present state, not the sequence of events that preceded it.

Figure 2 depicts the essential RL framework and its representation in the energy-management problem. In the original form, an agent receives state and reward signals from the environment and decides the control action. In the PHEV energy-management problem, the environment model can be regarded as the driving conditions, powertrain dynamics, and modeling. The agent is a particular power-split controller with different algorithms. The objective of this controller is to search for a sequence of actions according to the received state and reward information. Hence, different RL algorithms indicate that the techniques used for obtaining control actions are different, such as Q-learning, Dyna, temporal difference (TD), and the deep Q-network.

To train the RL algorithm, a value function is established in the agent. It is a function of state, action, and reward, and it is often represented as $Q(s, a)$ (s is state, and a is action). In the PHEV energy-management problem, the state, action, and reward information can be collected in real driving situations. Then, Markov decision processes (MDPs) are usually exploited to mimic the variables, meaning that the next state and reward are determined only by current information and are independent of historical data. Finally, the value function can be computed and applied to decide the best control action. Varying RL algorithms mean that the updating criteria of the value function are different. Moreover, RL algorithms can be classified into model-based and model-free versions, and they are distinguished by the dependency of the environment model. The model-free RL algorithms can easily handle the

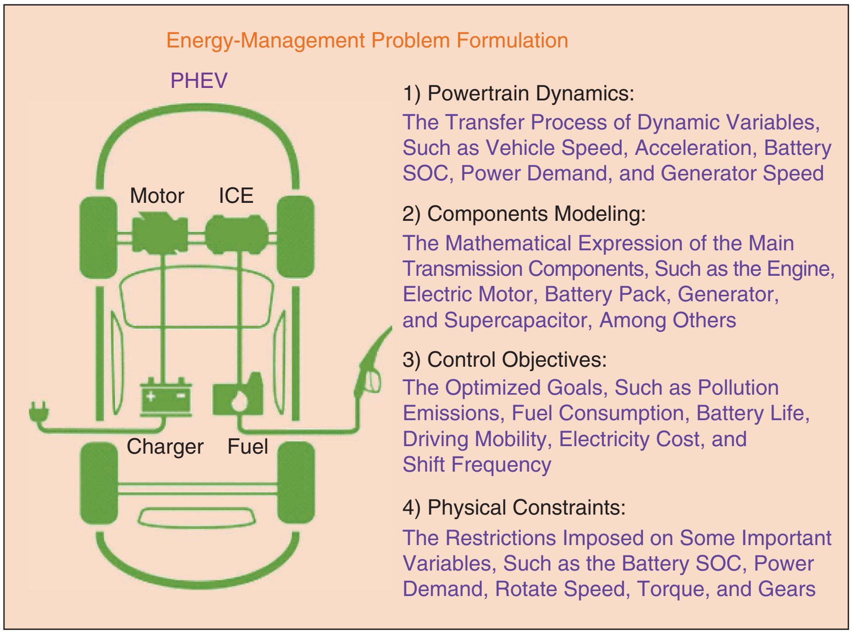


FIGURE 1 – The formulation of the energy-management problem for the PHEV.

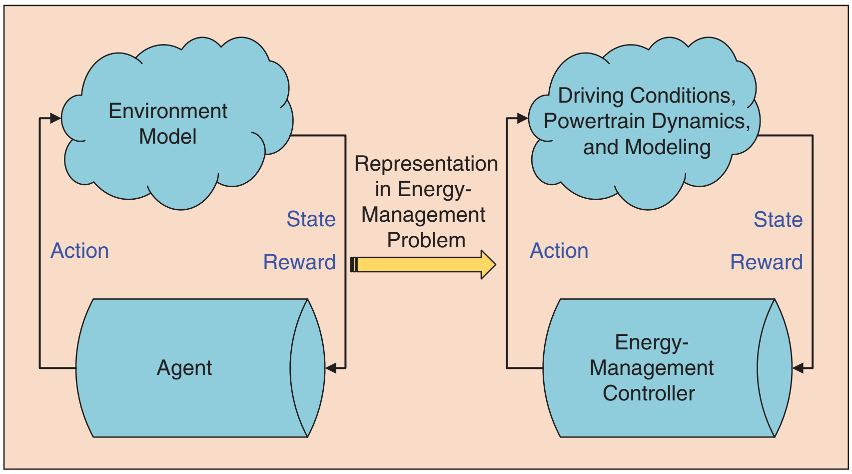


FIGURE 2 – The RL framework and its representation in the energy-management problem.

battery aging, engine abrasion, and driver behavior changes.

The Application of RL in PHEV Energy Management

In this section, state-of-the-art RL approaches in the HEV/PHEV energy-management field are summarized. First, the primary attempts of the simplex algorithms are discussed. Then, the recent progress and development of multiple-algorithm commixture, the hybrid algorithms, are introduced. Table 1 provides an overview of the different types of algorithms in the HEV/PHEV energy management.

Simplex Algorithms

In 2012, Hsu et al. assisted with the power management of a "pedelec," or hybrid bicycle, using the Q-learning algorithm [12]. They quantified the pedelec's safety and comfort targets as the quality of riding and improved energy utilization in the battery. Simulation results indicated that the quality-of-riding and energy objectives could be increased by 24 and 50%, respectively. Subsequently, many researchers changed the methods of HEV/PHEV energy management from optimization control theory to RL algorithms. For example, Yue et al. proposed a TD(λ)-learning-based, model-free, and online strategy to manage the energy flows in the supercapacitor and battery of an HEV [13]. In [14], the authors also applied a TD(λ)-learning algorithm to train and learn the optimal Q-function based on collected historical driving data. Implemented on a hybrid-electric bus, the RL-based controls aimed to increase fuel economy and reduce emissions. Using the same algorithm, TD(λ) learning, [15] constructed a power management system in an advanced vehicle simulator (ADVISOR). The convergence and complexity of the method were analyzed, and the deduced strategy was compared with the rule-based policy on different driving cycles.

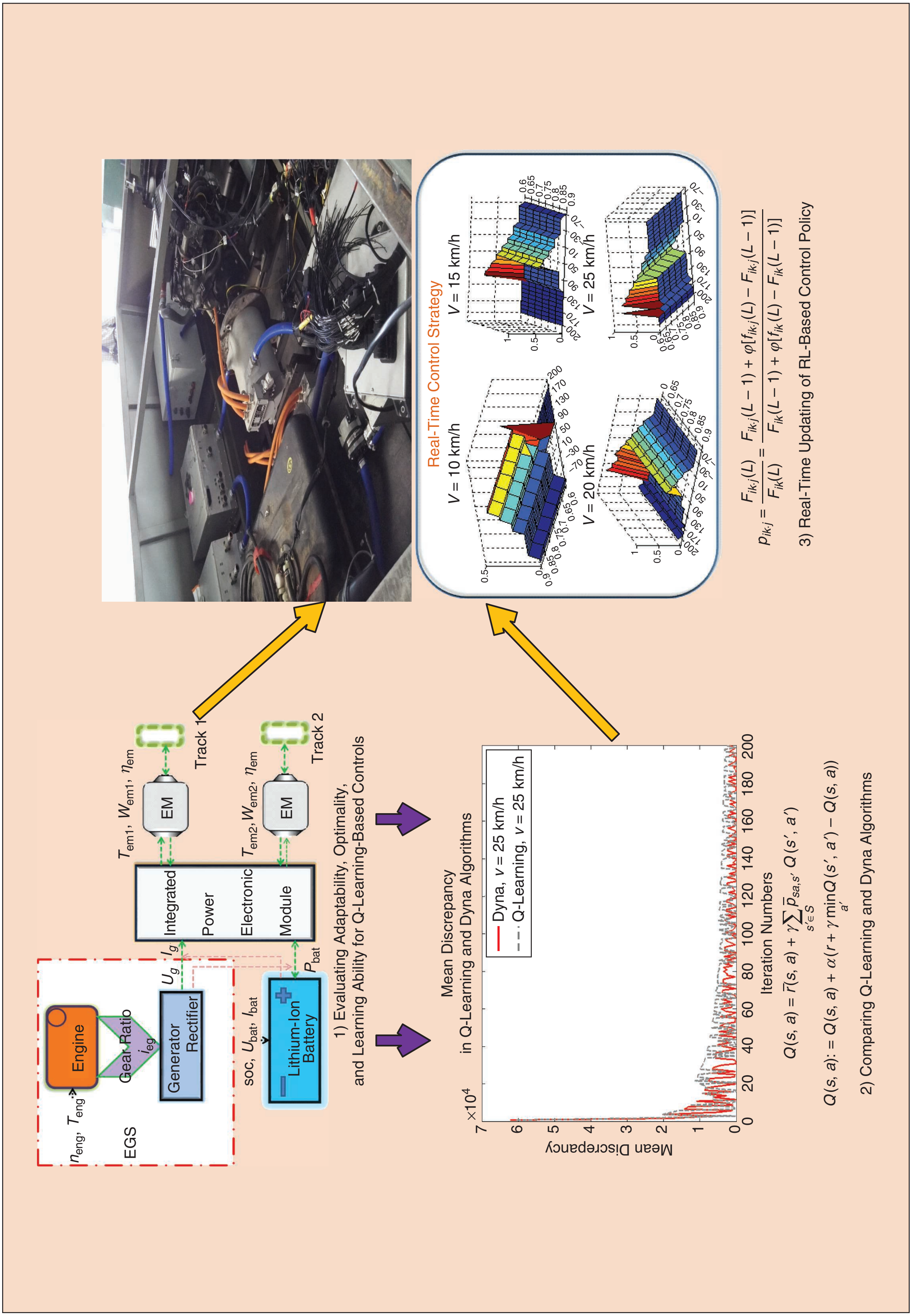
In addition to these efforts, several researchers began to seek breakthroughs in diverse aspects of energy management. For a PHEV, Qi et al. leveraged the Q-learning algorithm to optimize the charge-depletion strategy of the battery SOC [16]. Combined with the charge-sustaining strategy, the proposed method could balance the optimality and real-time performance. In [17], the authors built a new reward signal related to the power demand, SOC, and remaining distance to travel. The latter information was obtained from GPS data, and the TD(0) technique was utilized to train the estimated state-value tables. In [18], the authors leveraged the inverse RL (IRL) method to establish the probabilistic driving route prediction system, wherein driver behavior was predicted, and then the power-split rate between the engine and battery was calculated. Lin et al. presented a nested RL-based framework to address the operating cost of an HEV [19], in which the inner loop focused on the fuel cost minimization, and the outer loop aimed to optimize the battery replacement cost. Furthermore, in Johri's doctoral dissertation, the author built a self-learning system upon NDP and RL [20]. That system could minimize fuel consumption and predict real-time engine-out transient particulate and nitrogen oxide emissions.

During the past several years, Liu et al. also focused on researching RL-based power-split controls for hybrid powertrains. They first evaluated the adaptability, optimality, and learning ability of a Q-learning-based energy-management strategy for a hybrid-electric tracked vehicle [21]. Next, they compared the control performance of the Dyna and Q-learning algorithms, including fuel cost and calculation speed [22]. To achieve real-time controls for

TABLE 1 – A COMPARISON OF THE RL ALGORITHMS IN THE PHEV.

ALGORITHM	POWERTRAIN STRUCTURE	ADVANTAGES	DISADVANTAGES	REFERENCES
Q-learning	Bicycle	Adaptive to the riding environment	Strong model simplifications	[12]
TD(λ)	HEV	High accuracy	Dependency on driving data	[13]–[15]
Q-learning	Plug-in HEV	Possible online implementation	Local optimization sometimes	[16]
TD(0)	Plug-in HEV	Self-improvement capability	Dependency on GPS	[17]
Inverse RL	HEV	Finding-reward function	Training data quality	[18]
Q-learning	HEV	Multiple control objectives	Computational burden	[19]
NDP and RL	Hydraulic HEV	Robust against parameter changes	Design complexity	[20]
Dyna	Hybrid tracked vehicle	Real-time controls	Complex mathematics	[22], [23]
Deep RL	Plug-in HEV	Data-driven model	Special training cases	[24]–[27]
Online RL	All-climate EV	Fast computation	Driving cycle sensitivity	[28], [29]
Continuous RL	Plug-in HEV	Lower calculation effort	Tuned neural network	[30], [31]
MDP and RL	HEV and V2G	Improved battery life	Data requirement	[32], [50]
Q(λ)-learning	Plug-in HEV	Multiagent framework	Nonunique solution	[33]
Predictive RL	HEV	Robust against uncertainties	Online performance	[34]–[37]
Dyna-H	Multimode HEV	Convergence rate	Different driving scenarios	[42]

NDP: neuro-dynamic programming; V2G: vehicle-to-grid.



$$Q(s,a) = \bar{r}(s,a) + \gamma \sum_{s' \in S} \bar{p}_{sa,s'} Q(s',a')$$

$$Q(s,a) := Q(s,a) + \alpha\left(r + \gamma \min_{a'} Q(s',a') - Q(s,a)\right)$$

$$p_{ik \cdot j} = \frac{F_{ikj}(L)}{F_{ik}(L)} = \frac{F_{ikj}(L-1) + \varphi[f_{ikj}(L) - F_{ikj}(L-1)]}{F_{ik}(L-1) + \varphi[f_{ik}(L) - F_{ik}(L-1)]}$$

FIGURE 3 – Examples of the simplex algorithms in the PHEV's energy-management field. EGS: engine–generator system.

a hybrid powertrain, they integrated an online recursive algorithm into the Q-learning framework to achieve real-time updating of the control strategy [23]. Their research studies of simplex algorithms are summarized in Figure 3. Dyna and Q-learning are respectively applied in the energy management of a hybrid tracked vehicle, and their performance is compared in simulation results. Furthermore, the formulated control strategies are validated in a real vehicle, which proves the online realization of simplex algorithms. However, these algorithms may not handle mutable driving situations, which means that driving behaviors, motion areas, and road environments are different.

Hybrid Algorithms

In recent years, with the rapid development of deep learning and artificial intelligence, HEV/PHEV energy-management strategies have become increasingly intelligent. Two or more algorithms, or varying amounts of information, are usually integrated into the RL framework to constitute more efficient and real-time controls, such as velocity and power demand prediction, shared information from vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, and interaction with smart grids and smart cities.

DRL has been found to be an efficient tool to design the adaptive energy-management strategy based on historical driving data. For example, Qi et al. generated DRL-based power-split controls for a PHEV that did not depend on predefined principles or prediction [24]. Hu et al. evaluated their DRL-based energy management-strategy in an ADVISOR and proved its online learning character via a comparison with a rule-based strategy [25] as shown in Figure 4. Different energy-management policies are, first, computed and stored as historical experiences via DRL. In current driving situations, the best matching strategies are chosen by value function error and applied to manage the vehicle in real time. Thus, more data are required in hybrid-algorithm cases. In [26], the authors applied a deep neural network (DNN) to train the offline value functions and used the Q-learning algorithm to compute the online controls, which can be adaptive to different powertrain modeling and driving situations. The authors of [27] constructed DRL-enabled energy-management strategies that considered different drivers' behaviors, which could improve fuel efficiency.

Xiong et al. made a number of attempts at real-time and continuous RL-based energy-management strategies. They presented a real-time control strategy via combining the Q-learning algorithm and an online updating algorithm of value function [28], which means that the control actions can be refreshed in real time. Next, they validated this strategy by considering the battery and ultracapacitor in the loop, where a hardware-in-loop platform was established to execute the simulations [29]. To exploit the trained value function provided by RL methods, He et al. leveraged an actor-critic approach to handle the continuous action space and used stochastic gradient descent to train the state-value function [30], [31].

Two additional studies expanded the horizon of DRL-based learning construction in the PHEV energy-management field. In [32], Hoang et al. considered PHEV discharging and charging strategies for conditions when information from a vehicle to the grid is available. The MDP was utilized

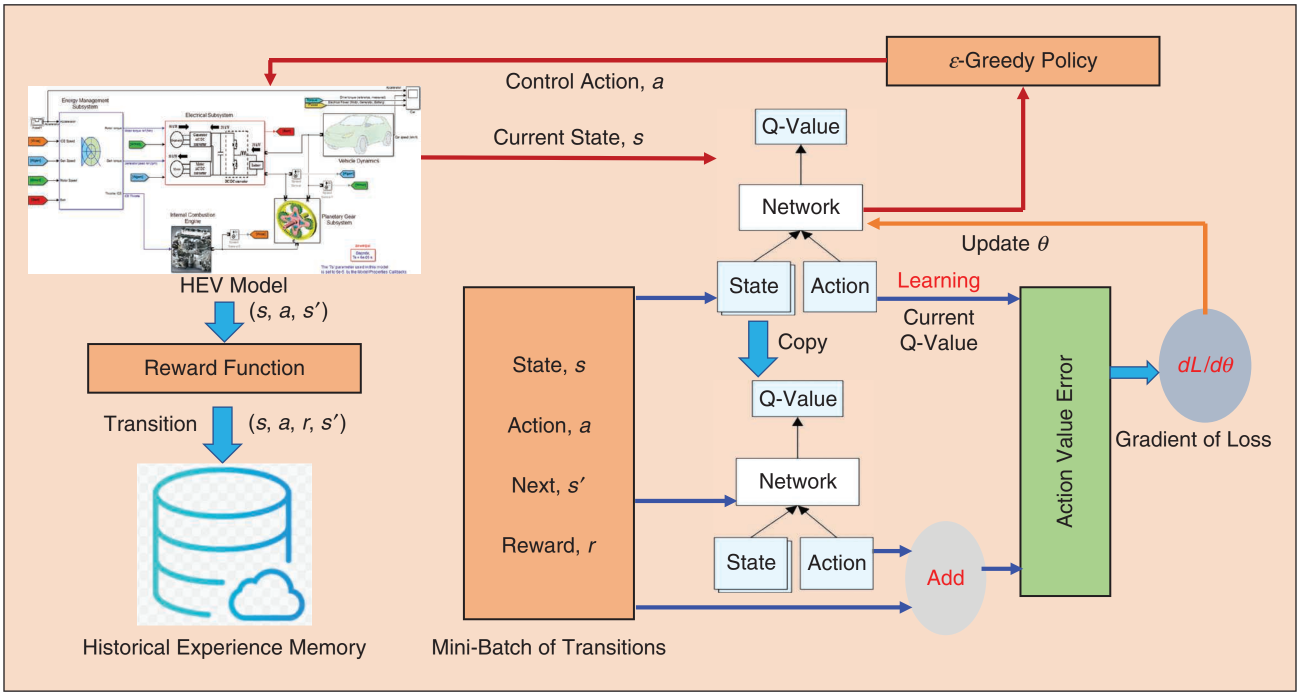


FIGURE 4 – The deep RL-enabled energy-management strategies for different driving situations [25]. Nm: Newton meter. (Source: Multidisciplinary Digital Publishing Institute: adapted and used with permission.)

to formulate the energy cost problem, and an RL-based learning structure was used to decide the online charging and cyberinsurance strategies. The relevant framework is described in Figure 5. The energy-management strategies should improve each vehicle's energy efficiency and consider price information and the use of charging stations. The energy-management problem grows from a single vehicle to multiple vehicles in a connected environment. Furthermore, to build an accurate charging-load model for multiple plug-in electric taxis, the authors in [33] applied multiple agents and a multiple-step Q(λ)-learning algorithm to search for a precise and detailed charging strategy for those vehicles. The reward performance and convergence rate were verified to be much better.

Liu et al. implemented many attempts at advanced and novel RL-based power-split controls with hybrid algorithms. To predict the vehicle speed or power demand accurately, they developed a fuzzy encoding predictor to forecast those variables [34]. Based on future information, they used the Q-learning algorithm to derive the guided control strategy for a hybrid tracked vehicle [35]. The computational time and energy efficiency achieved remarkable improvements. As depicted in Figure 6, they proposed decision-making and energy-efficient controls for a group of automated HEVs. The upper level used MPC to determine the velocity trajectories to avoid red-light idling, and the lower level applied RL to optimize the energy-efficient controls for each HEV [36]. This example indicates that anticipative energy-management strategies should be formulated to consider information communicated from other vehicles and infrastructures. They also used a speedy Q-learning algorithm to reserve a series of control policies offline. Further, an induced matrix norm was leveraged to choose the appropriate policy related to the current driving conditions [37].

Most of the HEVs and PHEVs in the market adopt rule-based energy-management strategies that are built from humans' engineering experience, which indicates that the techniques are simple enough and impose a low computational burden. Although DRL-based energy-management strategies are superior to rule-based ones, two conditions restrict their real-time applications. The first is the calculative capability of the onboard control unit; an extra computer needs to be installed in the vehicle to process the data. The second is data collection and storage; DRL needs an enormous amount of data to enable the derived strategies that are adaptive to different driving situations. With the development of network communication and intelligent transportation systems (ITSs), DRL-based energy-management strategies may be easily applied in real-time, in the future.

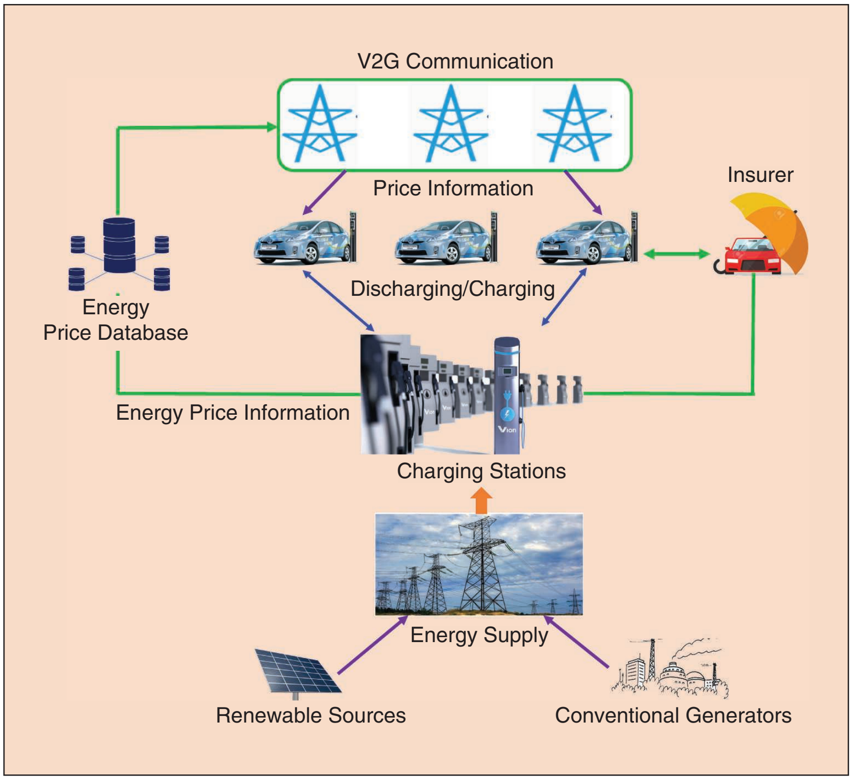


FIGURE 5 – The PHEV's energy-management strategy, considering connected information and grid interactions. Adapted from [32].

The Future Prospects for RL Applications

This section examines the future prospects and trends for RL-based HEV/PHEV energy management. The prospects consist of four circumstances: novel and efficient RL algorithms are going to be applied in this field, energy management will integrate with ITSs to construct a smart city or smart grid, the optimization control objectives will become more comprehensive and complicated, and there will be distributed or multiagent DRL systems for cooperative learning between vehicles in a connected environment.

Novel RL Algorithms

Fast development of the calculative capacity enables neoteric algorithms to be used in the HEV/PHEV energy-management field. Different kinds of deep-learning algorithms can be exploited to classify, train, and learn the massive scale of the data. For example, deep belief networks, stacked autoencoders, and recurrent neural networks [38] are promising approaches to learn the special model or table from the generous data. Then, double Q-learning, speedy Q-learning, and deep deterministic policy gradient algorithms [39] are able to formulate the optimal policy based on the trained model or table. With the help of cloudy control and management, these methods are practical and useful for the real-time operations of HEVs/PHEVs.

Furthermore, other RL algorithms have proved to be superior to the traditional Q-learning and Dyna, such as the k-nearest-neighbors TD [kNN-TD(λ)] and Dyna-H. In the former, a kNN method could represent probabilistic characteristics of the state variables, and TD back-propagation could be used to learn the control actions [40]. Dyna-H is a model-free online algorithm that adds a heuristic planning strategy to a Dyna agent to choose the optimal controls more efficiently [41], [42]. Finally, IRL is also suitable for energy-management problems, especially when the control objectives are unknown. This method can search for the proper reward signal through trial and error and learn from the existing experience [43].

Energy Management in Intelligent Transportation Systems

Information from emerging ITS technologies [for example, vehicle-to-everything (V2X) communication] provides great assistance to the process of improving energy management, such as real-time trip information, specific traffic situations, cloudy prediction, and weather conditions. For example, future trip information can be learned and forecast from historical driving data. Based on this information, the energy-management strategy can be more adaptive and robust to dynamic driving conditions, especially for HEVs/PHEVs with stationary routes [44]. The current onboard devices have the ability to get real-time traffic conditions via wireless communication, GPS, and geographical information systems. The information can regulate the online control strategy through advanced computation methods [45].

Furthermore, future vehicle velocities and power demands are potent information for influencing the power-split controls of a PHEV. The development of reliable algorithms may be utilized to acquire these data through a cloud platform or V2X communication. The data processing can be executed online, and the achieved controls are feasible for a group of vehicles [46]. Finally, weather conditions are essential factors in driver behavior and in fuel and electricity costs. The wind direction and temperature may affect aerodynamics and rolling resistance, and the weather may influence the driving styles of different operators. How to adjust the energy-management strategy according to weather information is an open question [47].

The Combination of Multiple Objectives

A transition from one common objective (fuel cost) to multiple goals is another research interest in the future energy-management field. These objectives include greenhouse gas emissions [48], the battery SOH [49], [50], safety, comfort, user convenience [51], and powertrain mobility [52]. Eco-driving is a promising method to reduce the use of ICEs to lower emissions. Battery health is a critical parameter that limits the driving range of a vehicle with electric power. Drivability for safety and comfort is significant for today's vehicles that have human drivers. Researchers need to strive to reach these representative objectives in solving future energy-management problems. More importantly, the question of how to handle the enormous

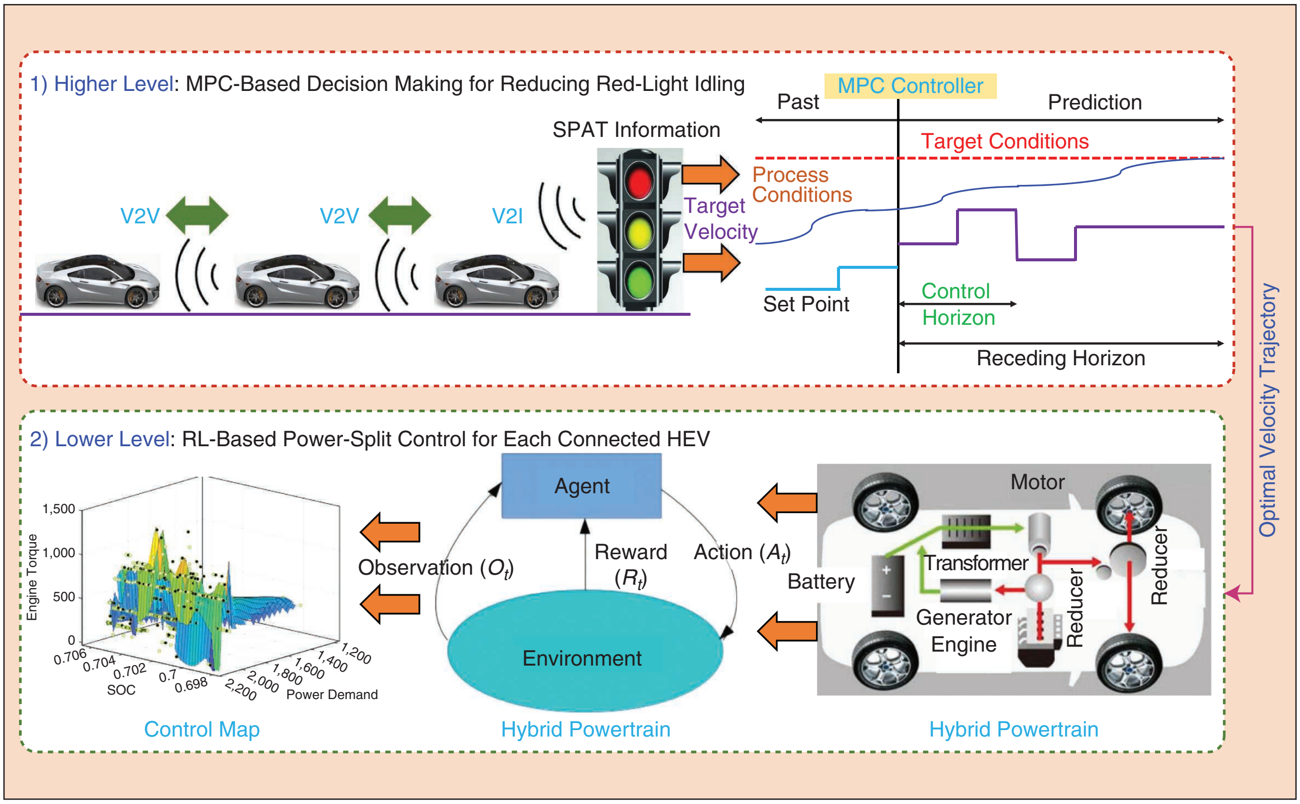


FIGURE 6 – The combination of MPC and RL for decision-making and energy-efficient controls for automated HEVs. SPAT: Signal Phasing and Timing.

computational overhead demanded by the multiple control objectives must be resolved.

Cooperative Learning in a Connected Environment

Vehicle automation is a research hotspot in the automobile industry. In the future, numerous HEVs/PHEVs could communicate with each other, and their driving behaviors might affect one another. In this connected and networked environment, a central controller should consider more objectives than improving one vehicle's energy efficiency. The energy-management controller should realize each vehicle's control objective while taking other vehicles' influences into account, in a harmonized and coordinated way. This goal should be achieved by advanced learning methods. For example, the asynchronous variants of standard RL algorithms are proposed in [53]. This concept trains a DNN through asynchronous gradient descent, and it can shorten computational time and realize parallel calculation. By doing this, different energy-management problems can be solved in a connected environment where HEVs/PHEVs consider each other's driving behaviors.

Conclusion

This article summarized DRL-based energy-management strategies for PHEVs. We began with an introduction of the energy-management problem and RL framework. Then, extensive applications with multiple control objectives were discussed. Finally, the prospects of an RL-based energy-management system were described.

One of the future research directions lies in applying more efficient artificial intelligence approaches in the energy-management field. The theoretical feasibility could be validated by simulation, and the practical implementation should be conducted through real-vehicle evaluations. Another major future work could access and improve energy-management strategies in the intelligent transportation environments. Since traffic information can be acquired, a procedure to tune strategies according to vehicles' and infrastructures' behaviors should be further addressed.

Acknowledgment

This work was supported, in part, by the National Natural Science Foundation of China (grant 51875054).

Biographies

Xiaosong Hu (xiaosonghu@ieee.org) earned his B.S. and Ph.D. degrees in automotive engineering at the Beijing Institute of Technology, China, in 2006 and 2012, respectively. He researched and earned his Ph.D. degree at the Automotive Research Center at the University of Michigan, Ann Arbor, from 2010 to 2012. He is a professor at the State Key Laboratory of Mechanical Transmissions and in the Department of Automotive Engineering at Chongqing University, China. He has received several awards, including the Emerging Sustainability Leaders Award in 2016 and European Union Marie Currie Fellowship in 2015. He is a Senior Member of the IEEE.

Teng Liu (tengliu17@gmail.com) earned his B.S. degree in applied mathematics from the Beijing Institute of Technology, China, in 2011 and his Ph.D. degree in automotive engineering at the Beijing Institute of Technology in 2017. He worked as a research fellow at Vehicle Intelligence Pioneers, Beijing, for one year and is currently a postdoctoral fellow in the Department of Mechanical and Mechatronics Engineering at the University of Waterloo, Ontario, Canada. His research focuses on reinforcement learning (RL)-based energy management in hybrid electric vehicles and RL decision making for autonomous vehicles. He has published more 40 Science Citation Index papers and 15 conference papers in those areas.

Xuewei Qi (qixuewei@gmail.com) earned his B.S. degree in automation at the China Agriculture University, Beijing, in 2008, his M.S. degree in engineering at the University of Georgia, Athens, in 2013, and his Ph.D. degree in electrical and computer engineering at the University of California, Riverside, in 2016. He is an artificial intelligence scientist in autonomous vehicle technology at General Motors, Detroit. He is a committee member of the Alternative Transportation Fuels and Technologies Standing Committee and the Artificial Intelligence Standing Committee and Advanced Computing Standing Committee of the Transportation Research Board at the National Academy of Sciences, Engineering, and Medicine, Washington, D.C.

Matthew Barth (barth@ece.ucr.edu) earned his B.S. degree in electrical engineering/computer science at the University of Colorado, Boulder, in 1984 and his M.S. (1985) and Ph.D. (1990) degrees in electrical and computer engineering at the University of California, Santa Barbara. His research focuses on applying engineering system concepts and automation technology to transportation systems. He is active on the Transportation Research Board at the National Academy of Sciences, Engineering, and Medicine, Washington, D.C., serving on the Transportation and Air Quality Committee and the Intelligent Transportation Systems Committee. He is a Fellow of the IEEE.

References

[1] C. Depature et al., "Energy management in fuel-cell/battery vehicles: Key issues identified in the IEEE Vehicular Technology Society Motor Vehicle Challenge 2017," *IEEE Veh. Technol. Mag.*, vol. 13, no. 3, pp. 144–151, June 2018. doi: 10.1109/MVT.2018.2837154.

[2] J. Solano, D. Hissel, and M. Pera, "Fail-safe power for hybrid electric vehicles: Implementing a self-sustained global energy management system," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 34-39, June 2018. doi: 10.1109/MVT.2017.2776670.

[3] L. Serrao, S. Onori, and G. Rizzoni, "A comparative analysis of energy management strategies for hybrid electric vehicles," *J. Dyn. Syst. Meas. Control*, vol. 13, no. 3, p. 031012, May 2011. doi: 10.1115/1.4003267.

[4] T. Liu, H. Yu, H. Guo, Y. Qin, and Y. Zou, "Online energy management for multimode plug-in hybrid electric vehicles," *IEEE Trans. Ind. Informat.*, vol. 15, no. 7, pp. 4352–4361, July 2019. doi: 10.1109/TII.2018.2880897.

[5] A. Malikopoulos, "Supervisory power management control algorithms for hybrid electric vehicles: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1869–1885, Oct. 2014. doi: 10.1109/TITS.2014.2309674.

[6] B. Ganji and A. Kouzani, "A study on look-ahead control and energy management strategies in hybrid electric vehicles," in *Proc. IEEE Int. Conf. Control Automation (ICCA 2010)*, pp. 388–392. doi: 10.1109/ICCA.2010.5524178.

[7] E. Foruzan, L. Soh, and S. Asgarpoor, "Reinforcement learning approach for optimal distributed energy management in a microgrid," *IEEE Trans. Power Syst.*, vol. 33, no. 5, pp. 5749–5758, Sept. 2018. doi: 10.1109/TPWRS.2018.2823641.

[8] R. S. Sutton, "Dyna, an integrated architecture for learning, planning, and reacting," *ACM SIGART Bulletin*, vol. 2, no. 4, pp. 160–163, Aug. 1991. doi: 10.1145/122344.122377.

[9] G. Williams et al., "Information theoretic MPC for model-based reinforcement learning," in *Proc. 2017 IEEE Int. Conf. Robotics Automation (ICRA)*, pp. 1714–1721. doi: 10.1109/ICRA.2017.7989202.

[10] F. Soares, D. Rua, C. Gouveia, B. Tavares, A. Coelho, and J. Lopes, "Electric vehicles charging: Management and control strategies," *IEEE Veh. Technol. Mag.*, vol. 13, no. 1, pp. 130–139, Mar. 2018. doi: 10.1109/MVT.2017.2781538.

[11] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*, 2nd. ed. Cambridge, MA: MIT Press, 2018.

[12] R. Hsu, C. Liu, and D. Chan, "A reinforcement-learning-based assisted power management with QoR provisioning for human-electric hybrid bicycle," *IEEE Trans. Ind. Electron.*, vol. 59, no. 8, pp. 3350–3359, Aug. 2012. doi: 10.1109/TIE.2011.2141092.

[13] S. Yue, Y. Wang, Q. Xie, D. Zhu, M. Pedram, and N. Chang, "Model-free learning-based online management of hybrid electrical energy storage systems in electric vehicles," in *Proc. 40th Annu. Conf. IEEE Industrial Electronics Society (IECON 2014),* pp. 3142–3148. doi: 10.1109/IECON.2014.7048959.

[14] Y. Fang, C. Song, B. Xia, and Q. song, "An energy management strategy for hybrid electric bus based on reinforcement learning," in *Proc. 27th Chinese Control Decision Conf. (2015 CCDC)*, pp. 4973–4977. doi: 10.1109/CCDC.2015.7162814.

[15] X. Lin, Y. Wang, P. Bogdan, N. Chang, and M. Pedram, "Reinforcement learning based power management for hybrid electric vehicles," in *Proc. 2014 IEEE/Association Computing Machinery Int. Conf. Computer-Aided Design (ICCAD)*, pp. 32–38. doi: 10.1109/ICCAD.2014.7001326.

[16] X. Qi, G. Wu, K. Boriboonsomsin, M. Barth, and J. Gonder, "Data-driven reinforcement learning-based real-time energy management system for plug-in hybrid electric vehicles," *J. Transportation Res. Board*, vol. 2572, no. 1, pp. 1–8, Jan. 2016. doi: 10.3141/2572-01.

[17] C. Liu and Y. Murphey, "Power management for plug-in hybrid electric vehicles using reinforcement learning with trip information," in *Proc. 2014 IEEE Transportation Electrification Conf. Expo (ITEC)*, pp. 1–6. doi: 10.1109/ITEC.2014.6861862.

[18] A. Vogel, D. Ramachandran, R. Gupta, and A. Raux, "Improving hybrid vehicle fuel efficiency using inverse reinforcement learning," in *Proc. 26th Association Advancement Artificial Intelligence Conf. Artificial Intelligence*, 2012, pp. 384–390.

[19] X. Lin, P. Bogdan, N. Chang, and M. Pedram, "Machine learning-based energy management in a hybrid electric vehicle to minimize total operating cost," in *Proc. 2015 IEEE/Association Computing Machinery Int. Conf. Computer-Aided Design (ICCAD)*, pp. 627–634. doi: 10.1109/ICCAD.2015.7372628.

[20] R. Johri, "Neuro-dynamic programming and reinforcement learning for optimal energy management of a series hydraulic hybrid vehicle considering engine transient emissions," Ph.D. dissertation, Dept. Mech. Eng., Univ. Michigan, Ann Arbor, MI, 2011.

[21] T. Liu, Y. Zou, D. 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, Dec. 2015. doi: 10.1109/TIE.2015.2475419.

[22] T. Liu, Y. Zou, D. Liu, and F. Sun, "Reinforcement learning-based energy management strategy for a hybrid electric tracked vehicle," *Energies*, vol. 8, no. 7, pp. 7243–7260, July 2015. doi: 10.3390/en8077243.

[23] Y. Zou, T. Liu, D. X. Liu, and F. C. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," *Appl. Energy*, vol. 171, pp. 372–382, June 2016. doi: 10.1016/j.apenergy.2016.03.082.

[24] X. Qi, Y. Luo, G. Wu, K. Boriboonsomsin, and M. Barth, "Deep reinforcement learning-based vehicle energy efficiency autonomous learning system," in *Proc. 2017 IEEE Intelligent Vehicles Symp. (IV)*, pp. 1228–1233. doi: 10.1109/IVS.2017.7995880.

[25] 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. doi: 10.3390/app8020187.

[26] P. Zhao, Y. Wang, N. Chang, Q. Zhu, and X. Lin, "A deep reinforcement learning framework for optimizing fuel economy of hybrid electric vehicles," in *Proc 2018. 23rd Asia South Pacific Design Automation Conf. (ASP-DAC)*, pp. 196–202. doi: 10.1109/ASP-DAC.2018.8297305.

[27] R. Liessner, C. Schroer, A. Dietermann, and B. Baker, "Deep reinforcement learning for advanced energy management of hybrid electric vehicles," in *Proc. 10th Int. Conf. Agents and Artificial Intelligence (ICAART 2018)*, pp. 61–72. doi: 10.5220/0006573000610072.

[28] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," *Appl. Energy*, vol. 211, pp. 538–548, Nov. 2017. doi: 10.1016/j.apenergy.2017.11.072.

[29] R. Xiong, Y. Duan, J. Cao, and Q. Yu, "Battery and ultracapacitor in-the-loop approach to validate a real-time power management method for an all-climate electric vehicle," *Appl. Energy*, vol. 217, pp. 153–165, May 2018. doi: 10.1016/j.apenergy.2018.02.128.

[30] Y. Li, H. He, J. Peng, and H. Zhang, "Power management for a plug-in hybrid electric vehicle based on reinforcement learning with continuous state and action spaces," *Energy Procedia*, vol. 142, pp. 2270–2275, Dec. 2017. doi: 10.1016/j.egypro.2017.12.629.

[31] J. Wu, H. He, J. Peng, Y. Li, and Z. Li, "Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus," *Appl. Energy*, vol. 222, pp. 799–811, July 2018. doi: 10.1016/j.apenergy.2018.03.104.

[32] D. Hoang, P. Wang, D. Niyato, and A. Hossain, "Charging and discharging of plug-in electric vehicles (PHEV) in vehicle-to-grid (V2G) systems: A cyber insurance-based model," *IEEE Access*, vol. 5, pp. 732–754, Jan. 2017. doi: 10.1109/ACCESS.2017.2649042.

[33] C. Jiang, Z. Jing, X. Cui, T. Ji, and Q. Wu, "Multiple agents and reinforcement learning for modelling charging loads of electric taxis," *Appl. Energy*, vol. 222, pp. 158–168, July 2018. doi: 10.1016/j.apenergy.2018.03.164.

[34] T. Liu, X. Hu, S. 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, Aug. 2017. doi: 10.1109/TMECH.2017.2707338.

[35] T. Liu and X. Hu, "A bi-level control for energy efficiency improvement of a hybrid tracked vehicle," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1616–1625, Apr. 2018. doi: 10.1109/TII.2018.2797322.

[36] T. Zhang, G. Kahn, S. Levine, and P. Abbeel, "Learning deep control policies for autonomous aerial vehicles with mpc-guided policy search," in *Proc. 2016 IEEE Int. Conf. Robotics and Automation (ICRA),* pp. 528–535, May, 2016.

[37] T. Liu, B. Wang, and C. Yang, "Online Markov chain-based energy management for a hybrid tracked vehicle with speedy Q-learning," *Energy*, vol. 160, pp. 544–555, Oct. 2018. doi: 10.1016/j.energy.2018.07.022.

[38] S. Levine, P. Pastor, A. Krizhevsky, J. Ibarz, and D. Quillen, "Learning hand-eye coordination for robotic grasping with deep learning and large-scale data collection," *Int. J. Robotics Res.*, vol. 37, no. 4–5, pp. 421–436, Apr. 2018. doi: 10.1177/0278364917710318.

[39] T. Lillicrap et al., "Continuous control with deep reinforcement learning," presented at the Int. Conf. Learning Representations (ICLR), San Juan, Puerto Rico, 2016.

[40] J.A. Martin, J. Lope, and D. Maravall, "Robust high-performance reinforcement learning through weighted k-nearest neighbors," *Neurocomputing*, vol. 74, no. 8, pp. 1251–1259, Mar. 2011. doi: 10.1016/j.neucom.2010.07.027.

[41] M. Santos, J.A. Martin, V. Lopez, and G. Botella, "Dyna-H: A heuristic planning reinforcement learning algorithm applied to role-playing game strategy decision systems," *Knowledge-Based Syst.*, vol. 32, pp. 28–36, Aug. 2012. doi: 10.1016/j.knosys.2011.09.008.

[42] T. Liu, X. Hu, W. Hu, and Y. Zou, "A heuristic planning reinforcement learning-based energy management for power-split plug-in hybrid electric vehicles," *IEEE Trans. Ind. Informat.*, Mar. 2019, to be published. doi: 10.1109/TII.2019.2903098.

[43] A. Ng and S. Russell, "Algorithms for inverse reinforcement learning," in *Proc. 17th Int. Conf. Machine Learning (ICML '00)*, 2000, pp. 663–670.

[44] Y. Hay, M. Kuang, and R. McGee, "Trip-oriented energy management control strategy for plug-in hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1323–1336, July 2014. doi: 10.1109/TCST.2013.2278684.

[45] C. Sun, S. Moura, X. Hu, J. Hedrick, and F. Sun, "Dynamic traffic feedback data enabled energy management in plug-in hybrid electric vehicles," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 3, pp. 1075–1086, May 2015. doi: 10.1109/TCST.2014.2361294.

[46] E. Ozatay et al., "Cloud-based velocity profile optimization for everyday driving: A dynamic-programming-based solution," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2491–2505, Dec. 2014. doi: 10.1109/TITS.2014.2319812.

[47] F. Yan, J. Wang, and K. Huang, "Hybrid electric vehicle model predictive control torque-split strategy incorporating engine transient characteristics," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2458–2467, July 2012. doi: 10.1109/TVT.2012.2197767.

[48] P. You et al., "Scheduling of EV battery swapping, I: Centralized solution," *IEEE Trans. Control Netw. Syst.*, vol. 5, no. 4, pp. 1887–1897, Dec. 2018. doi: 10.1109/TCNS.2017.2773025.

[49] W. Tushar, C. Yuen, S. Huang, D. Smith, and H. Poor, "Cost minimization of charging stations with photovoltaics: An approach with EV classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 1, pp. 156–169, Jan. 2016. doi: 10.1109/TITS.2015.2462824.

[50] X. Wang, C. Yuen, N. Hassan, N. An, and W. Wu, "Electric vehicle charging station placement for urban public bus systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 1, pp. 128–139, Jan. 2017. doi: 10.1109/TITS.2016.2563166.

[51] H. Chung, W. Li, C. Yuen, C. Wen, and N. Crespi, "Electric vehicle charge scheduling mechanism to maximize cost efficiency and user convenience," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 3020–3030, May 2019. doi: 10.1109/TSG.2018.2817067.

[52] R. Yu, W. Zhong, S. Xie, C. Yuen, S. Gjessing, and Y. Zhang, "Balancing power demand through EV mobility in vehicle-to-grid mobile energy networks," *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 79–90, Feb. 2016. doi: 10.1109/TII.2015.2494884.

[53] V. Mnih et al., "Asynchronous methods for deep reinforcement learning," in *Proc. Int. Conf. Machine Learning (ICML 2016)*, pp. 1928–1937.