


Review

# Towards a Smarter Energy Management System for Hybrid Vehicles: A Comprehensive Review of Control Strategies

Nan Xu <sup>1,\*</sup> , Yan Kong <sup>1</sup>, Liang Chu <sup>1</sup>, Hao Ju <sup>1</sup>, Zhihua Yang <sup>1</sup>, Zhe Xu <sup>2</sup> and Zhuoqi Xu <sup>1</sup>

<sup>1</sup> State Key Laboratory of Automotive Simulation and Control, Jilin University, Changchun 130022, China; kongyan17@mails.jlu.edu.cn (Y.K.); chuliang@jlu.edu.cn (L.C.); juhao18@mails.jlu.edu.cn (H.J.); yzh@jlu.edu.cn (Z.Y.); xuzhuoqi.jlu@gmail.com (Z.X.)

<sup>2</sup> General R&D Institute of China FAW, Changchun 130011, China; xuzhe1@faw.com.cn

\* Correspondence: nanxu@jlu.edu.cn; Tel.: +86-431-8509-5165

Received: 30 March 2019; Accepted: 12 May 2019; Published: 16 May 2019



**Abstract:** This paper presents a comprehensive review of energy management control strategies utilized in hybrid electric vehicles (HEVs). These can be categorized as rule-based strategies and optimization-based strategies. Rule-based strategies, as the most basic strategy, are widely used due to their simplicity and practical application. The focus of rule-based strategies is to determine and optimize the optimal threshold for mode switching; however, they fall into a local optimal solutions. To have better performance in energy management, optimization-based strategies were developed. The categories of the existing optimization-based strategies are identified from the latest literature, and a brief study of each strategy is discussed, which consists of the main research ideas, the research focus, advantages, disadvantages and improvements to ameliorate optimality and real-time performance. Deterministic dynamic programming strategy is regarded as a benchmark. Based on neural network and the large data processing technology, data-driven strategies are put forward due to their approximate optimality and high computational efficiency. Finally, the comprehensive performance of each control strategy is analyzed with respect to five aspects. This paper not only provides a comprehensive analysis of energy management control strategies for HEVs, but also presents the emphasis in the future.

**Keywords:** energy management strategy; rule-based strategies; optimization-based strategies; data-driven strategies; hybrid electric vehicles

## 1. Introduction

With the increasingly serious issues of energy shortage and environmental pollution, gas-electric hybrid vehicles, as a new energy-fueled automobile, have received extensive attention from governments and automobile companies, and include the plug-in hybrid electric vehicle (PHEV) and traditional hybrid electric vehicles [1–3]. Hybrid electric vehicles (HEVs) can take full advantage of traditional fuel vehicles and pure electric vehicles into account, and have become a transitional stage from the traditional vehicle with internal combustion engines in electric vehicles (EV).

The energy management strategy (EMS) organically coordinates the power units to achieve optimal energy distribution by directly controlling fuel consumption rate and battery state of charge (SOC), which can achieve lower energy consumption and lower pollution [4–7]. Aiming towards common driving cycles, there are three methods for obtaining better performance and higher energy efficiency. The first method is to increase the efficiency of each powertrain component (such as engine, motor, power battery); however, this may lead to a small increase in overall efficiency, while increasing

its cost greatly. Forcing the engine and motors to operate in the high-efficiency zone by incorporating shifting strategies can also achieve higher energy efficiency; however, there may occur an abnormal phenomenon whereby efficiency and fuel consumption increase simultaneously. Moreover, based on the efficiency characteristics of the motor/engine, making the position and distribution of the high efficiency zone match the driving cycle is another method for obtaining better performance. In summary, in terms of the short-term and long-term range, the key point for reducing fuel consumption is to reasonably distribute the energy between the engine and motors so that the engine can operate in the high efficiency zone for as long as possible [8].

HEVs are equipped with two or more power sources, and vehicle powertrains can provide regenerative braking during deceleration and allow efficient auxiliary electricity and recharging operations. Therefore, HEVs can be operated alternately in pure electric, hybrid modes, regenerative braking mode, and so forth, which can have a better fuel consumption performance and effectively solve the mileage problem of EV [9]. Due to the issue of multiple power sources, the energy distribution between engine and motors is the core problem of energy management for HEVs. Vehicles need to meet the required driving force, and the energy consumption of the battery is completely dependent on the motor, that is, the torque distribution between engine and motors is the focus to EMS for HEVs.

Based on road conditions, power demand and battery SOC, to the means by which the power distribution between engine and motors to reduce fuel consumption as much as possible can be effectively determined has become a crucial issue in energy management control for HEVs. One important purpose of energy management system is guaranteeing that battery SOC is kept within a reasonable range while improving the fuel economy over the whole driving cycle [10]. Only in this way can we ensure a global balance of SOC and the lowest fuel consumption in a true sense.

Furthermore, implementation of a control strategy to update the driving condition in order to optimize the energy management system in real time is another key point for HEVs. Meanwhile, the higher the computational efficiency, the more likely it is to achieve real-time control. In other words, it is also necessary to consider whether real-time optimization can be achieved with low computational burden and high computational speed.

Based on the above problems, current control strategies can be categorized as rule-based (RB) strategies and optimization-based strategies [10,11] in terms of energy optimization for HEVs. To efficiently distribute the energy between oil and electricity, rule-based (RB) strategies select operating modes based on predefined rules, which are subdivided into deterministic rule-based strategies and fuzzy logic-based strategies [10,12]. Although rule-based strategies are the most basic strategies, and are widely applied due to their simplicity and practicality, they usually lead to local optimization such that they cannot obtain the globally optimal solution. On the other hand, it does not take the actual changes in driving conditions into account and its robustness cannot be guaranteed.

To achieve a better performance in terms of fuel economy in HEVs, numerous efforts been made in the last decade in the area of optimization-based energy management control strategies, mainly with respect to instantaneous optimization and global information driven optimization [13–15].

Instantaneous optimization strategies, such as Equivalent consumption minimization strategies (ECMS), and global information-driven optimization, such as Pontryagin's Minimum Principle (PMP) [16–19] and dynamic programming (DP) [20–23], are able to obtain near-optimal or globally optimal control strategies, however, they should acquire entire driving cycle information in advance. The objective function can be optimized in a global sense to obtain the global optimal fuel economy for global information-driven optimization strategies. Theoretically, deterministic dynamic programming strategies can obtain the global optimal fuel economy for HEVs, providing a benchmark for assessing the optimality of other energy management strategies. However, they suffer from difficulties in practical application because they are time consuming, and tremendous memory is required for the calculation process, while they also need to acquire the driving information in advance.

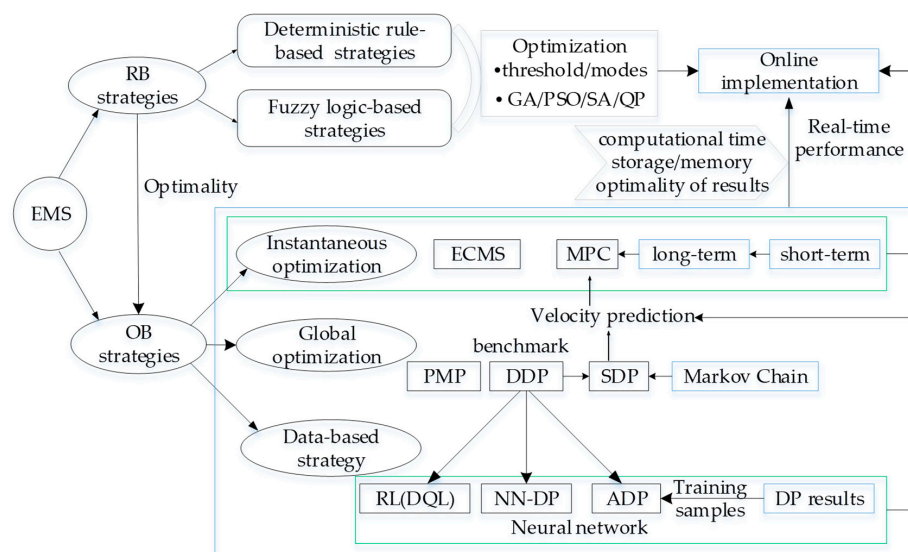
To obtain an applicable strategy, some researchers have attempted to design the EMS via model predictive control (MPC) [24,25] and stochastic dynamic programming (SDP) [26], which can be directly

applied online. The MPC strategy can apply various optimization methods to design the control strategy based on a period of the predicted velocity, which can be predicted via Markov Chain or artificial neural network. The optimization methods could be PMP algorithms, DDP algorithms, or some intelligent algorithms, including particle swarm optimization (PSO), genetic algorithm (GA), simulated annealing (SA), and quadratic programming (QA). On the basis of DDP-based strategy and for the purpose of practical application, the SDP-based strategy constructs a stochastic transition probability matrix of the driver demand by Markov chain and captures the statistical characteristics in traffic information with a stochastic model. Therefore, the SDP-based strategy is a good choice for a bus driving cycle with a fixed route, due to its high regularity. However, whether it is SDP-based or MPC-based strategy, speed prediction is required, which influences the computational accuracy for EMS.

With the rapid development of artificial intelligence algorithms and the large data processing technology, optimal data-driven control strategies show superior performance because of their on-line implementation and approximation to DP results. Data-driven strategies use neural networks to approximate the control sequence and cost function or the prediction of future driving patterns. At present, data-driven strategies are roughly composed of neural network-dynamic programming (NN-DP) [27], reinforcement learning (RL) [28,29], and adaptive dynamic programming (ADP) [28–31]. Among these, the ADP-based strategy can not only significantly reduce computational time and memory storage, but can also obtain similar results to DP results in achieving real-time optimization based on updated driving information.

In addition, with the development of new technologies such as hydrogen fuel cells, powertrain systems in a HEV can be the combination of engines, motors, batteries, super-capacitors, and hydrogen fuel cells. Since the objective of energy management control strategies is to reasonably distribute the energy of each power source, the various control strategies mentioned above are also applicable to a hydrogen fuel cell hybrid vehicle or a hybrid vehicle with a super-capacitor.

The purpose of this paper is to review the existing energy management control strategies and the advanced control strategies (such as ADP/deep learning algorithms) for gas-electric HEVs, which are shown in Figure 1. The general research ideas, the advantages and disadvantages of the current energy management control strategies related to HEVs will be emphatically discussed. It should be noted that control strategies described in this paper are still applicable to hydrogen fuel cell hybrid vehicles. The difference lies in allocating the energy between hydrogen consumption and electricity consumption, rather than fuel consumption and electricity consumption.



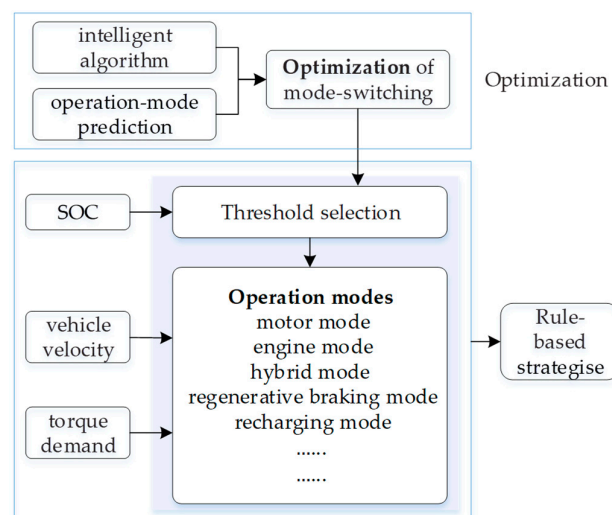
**Figure 1.** Depiction of the schematic diagram of this paper.

The structure of this paper is as follows: the first part reviews the development and the classification of energy management control strategies for HEVs. The second section introduces the main study idea and the optimization of rule-based strategies. The third section presents optimization-based strategies in terms of instantaneous optimization, global information driven optimization and data-driven strategies. This chapter is mainly about the overview of conventional optimization control strategies, like ECMS/MPC/PMP/DDP/SDP, and the novel data-driven strategies, like NN-DP/RL/ADP, which can achieve approximation to the global optimal results of DDP and can be implemented online based on the updated driving conditions. Based on the comprehensive performance analysis of each strategy in the fourth section, the conclusions and suggestions (outlook and future trends) for energy management control strategies for HEVs are compiled in the fifth section.

## 2. Rule-Based Strategies

To reasonably manage the multi-power source energy coupling system, an energy management control strategy is employed to distribute the power or torque of the multiple power sources. Meanwhile, the braking energy recovery is coordinated to improve fuel economy and system efficiency under the premise of ensuring the power requirements, safety and comfort of the vehicle. According to the steady-state efficiency map of the powertrain components, the rule-based control strategy, as the most practical control strategy, determines the torque/power distribution between engine and motors. It selects the operating mode based on predefined rules to make the vehicle operate in the high-efficiency zone to improve the fuel economy of the vehicle. Due the fact that the EMS of HEVs can be easily extended to PHEVs via a charge depleting–charge sustaining (CD–CS) mode, the charge depleting–charge sustaining (CD–CS) strategy [12,32,33] and the blended control strategy [4] were developed to achieve energy saving and emissions reduction.

In further research, rule-based control strategies combining CS, CD, EV and hybrid modes was used to improve fuel economy by making the engine and motors operate in the high efficiency area. According to the characteristics of predefined rules, rule-based control strategies are mainly subdivided into deterministic rule-based strategies and fuzzy logic-based strategies. The study of rule-based strategies is mainly directed towards three aspects: threshold selection, optimization of mode-switching thresholds, and determination of operation modes, as shown in Figure 2.



**Figure 2.** Depiction of the main research areas for rule-based strategies.

### 2.1. CD–CS Strategy and Blended Strategy

The limited charge/discharge capability and low battery capacity/power of conventional HEVs leads to a small operating area of battery SOC; therefore, the battery SOC of the conventional HEV has

a certain convergence. In other words, the final value of SOC should be same as its initial value over the whole driving cycle, which is referred to as charge sustaining (CS) mode. In CS mode, vehicles are powered by an electric machine, engine, or both to maintain the battery SOC. Plug-in hybrid electric vehicles (PHEVs) can be regarded as a combination of an electric vehicle (EV) and a conventional HEV, which have the all-electric capability of an EV and the extended range capability of an HEV. Due to the ability to charge via an external grid, the EMS of PHEVs increases charge depleting (CD) mode. In CD mode, the electric machine is primarily used to power the vehicle with a net decrease in battery SOC, while the engine turns on when the electric machine cannot provide the required power or SOC drops too low.

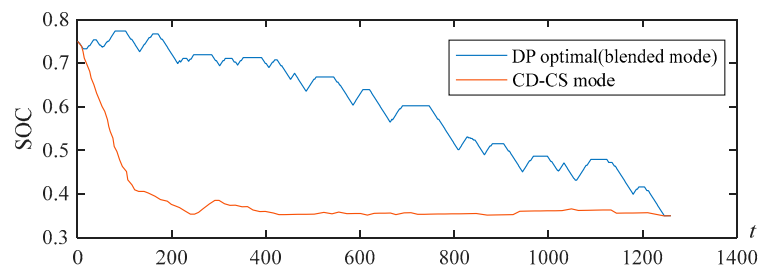
Based on the above characteristics, the charge depleting–charge sustaining (CD–CS) strategy, as the simplest method, is first proposed to generate an optimal SOC trajectory. This strategy operates in an all-electric mode in its initial stage, following by CS mode when the battery SOC decreases to a predefined low threshold along with an increase in driving distance. It is generally desirable that the battery reach its depleting limit by the end of the given driving distance. Furthermore, simulation results indicate that the CD–CS strategy, along with electric assistance, is more effective in PHEV than in HEV that have the same battery energy capacity as PHEV [12]. A CD control strategy for a generic parallel PHEV (SUV) was investigated in [34], which uses electric power to drive the vehicle until the power demand reaches a preset threshold, and then the engine turns on to meet the desired power requirements with the assistance of the motor. Meanwhile, the mechanical power of the motor remains constant from when the engine turns on until the end of the drive cycle. With the development of intelligent transportation systems (ITSs), geographical information systems (GISs) and global positioning systems (GPSs), the trip information can be obtained in advance for a certain route. Therefore, if the basic trip information can be acquired in advance, the CD strategy would slightly increase the fuel economy of a PHEV compared with the AER strategy [35].

Although the CD–CS strategy is characterized by its simplicity, ease of implementation, and allowing HEVs to operate as an “electric vehicle”, it is far from obtaining the optimal fuel economy of energy management for HEVs. To obtain relatively satisfying fuel economy, the CD–CS strategy should equip a relatively large battery to meet power performance requirements in CD mode over the whole driving cycle, which will increase the vehicle cost. Generally speaking, sufficient CD operations may lead to more electric loss and a sharp decrease in engine efficiency under high power demand, while inadequate CD operations may not well achieve an improvement of fuel economy and make full use of EMS. To further reduce fuel consumption, the blended strategy with gradual battery depletion in a blended mode (BM) is implemented for an optimal EMS of HEVs. Engine and motors would operate in coordination, causing the electricity to not be completely consumed until the end.

A blended-mode HEV can achieve cruising with low electric drive. Simultaneously, it can moderately accelerate at low to moderate vehicle speeds in electric mode. Under moderate to high vehicle speeds, the engine will turn on to meet power/torque requirements, and the motor will serve as auxiliary power to provide more power/torque when the output power of engine cannot satisfy power demand. In [36], the blended-mode energy management of PHEVs was introduced to achieve the minimum total fuel consumption for a given driving cycle, while maintaining a constant battery energy. Compared with foremost strategies, the proposed strategy improves fuel savings by 8.7%, on average.

In contrast to the CD–CS strategy, the blended strategy with basic trip information can have a better performance in terms of fuel saving for HEVs. However, the performance of the blended-based strategy depends on the trip length. Longer trips may lead to battery exhaustion in advance, while shorter trips may result in power remaining. In addition, it is worth mentioning that the blended strategy without trip information may have a worse performance than a fine-tuned CD–CS strategy. When providing the information of the driving cycle, a comparison between CD–CS strategy and the blended strategy in terms of the battery SOC evolution is shown in Figure 3.





**Figure 3.** Depiction of the comparison between the CD-CS strategy and blended strategy.

## 2.2. Deterministic Rule-Based Strategy

Deterministic rule-based strategies are mainly based on several predetermined threshold parameters to make engines and motors operate in their high-efficiency area, which also takes battery charging and discharging efficiency into consideration in order to properly distribute the demand torque. According to power requirements, battery SOC, and vehicle speed or acceleration, the vehicle can be divided into several operation modes, which mainly include motor-only mode, engine-only mode, power-assist mode, regenerative braking mode, mechanical braking mode, recharging mode, and so forth. The mode switches when the operating state meets threshold switching condition. As detailed in [10], it takes four operation modes into consideration, namely, motor-only mode, engine-only mode, regenerative braking mode and hybrid mode, and the threshold of mode-switching is determined by desired torque and battery SOC.

Generally speaking, threshold selection of the deterministic rule-based strategy is mainly based on engineering experience and the efficiency characteristics of the related power components, and it requires considerable parameter debugging time to acquire satisfactory results. Thus, some optimization methods are required to automatically obtain global optimal threshold parameters to more reasonably distribute the torque between engine and the motors.

The main objective of mode-switching is to obtain minimum fuel consumption under the premise of meeting driver's demand and maintaining battery SOC. To make deterministic rule-based strategies more effective, mode-switching thresholds can be optimized using several intelligent algorithms, such as the genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), quadratic programming (QP), and so forth. For instance, mode-switching thresholds were optimized by simulated annealing–particle swarm optimization (SA–PSO) in [26] to obtain the ideal mode-switching sequence. In [37], a hybrid algorithm combining GA with SA was applied to simultaneously optimize powertrain and control parameters, resulting in a better convergence speed and offering a global searching ability to obtain the best comprehensive performance for a plug-in hybrid electric bus (PHEB). In order to have good real-time performance, the direct algorithm has been used to optimize extracted key parameters globally due to its low computational burden and rapid convergence [38].

In addition, the method of operation-mode prediction can also be introduced to optimize the deterministic rule-based strategy. A torque correction strategy based on operation-mode prediction is proposed in [39]. The current operation mode of HEV is determined by demand power, vehicle speed, and battery SOC, and then the operation mode in the immediate future time-horizon is predicted based on the Markov probability matrix [9,24,25]. In the end, based on ECMS algorithm, the optimized factor of torque correction is selected as the control parameter to correct engine torque and motor torque in real-time.

## 2.3. Fuzzy Logic-Based Strategy

With the in-depth study of fuzzy theory and based on fixed threshold control, the fuzzy logic-based strategy was developed to optimize the predefined control rules by fuzzification, rule base, fuzzy reasoning and defuzzification, offering strong adaptability and robustness.

Simple fuzzy rule-based strategies consider battery SOC as the primary input and combine this with other vehicle parameters, such as torque demand, vehicle speed, vehicle acceleration, motor speed and engine speed. For instance, the fuzzy controller takes vehicle speed, battery SOC and torque demand as inputs and outputs the torque request of internal combustion engine (ICE), resulting in a more than 10% increase in fuel economy compared with the standard system [40]. In addition, the performance of this proposed strategy was discussed with respect to fuel mileage, battery usage and driver performance. Furthermore, a fuzzy logic-based EMS was proposed for a through-the-road HEV, which applied fuzzy logic with a pair of membership functions to determine the appropriate power distribution in real time [41].

Simple fuzzy rule-based strategies have low computation requirements, but require engineering experience and have non-optimal results. In order to have good performance, there are three main methods of optimization. One method is to intelligently select fuzzy rules by looking up the table of off-line computed DP results. In [42], the mode selection between all-electric and hybrid is based on the fuzzy logic controller, which uses a set of rules extracted from DP results. Meanwhile, the controller can be adaptive to different driving conditions by using driving condition information. To further improve the fuel efficiency of EMS for HEVs, [43] proposes an energy management control strategy combining the conventional rule-based strategy and the global optimization strategy. Meanwhile, driving pattern cycle recognition is used to classify current driving conditions into one of the driving patterns, and the DP method is applied to design fuzzy-logic control strategy of each driving pattern.

Similar to threshold optimization for the deterministic rule-based strategy, the fuzzy logic-based strategy can also be optimized by intelligent algorithms, such as GA, PSO, QP, SA, and so on [44,45]. Another method is the adaptive controller based on fuzzy logic, which can identify driver behavior and optimize itself for these situations. This method does not need predetermined rules and can be optimized for various drive cycles. For instance, a fuzzy rule-based EMS can automatically identify the driver's style, intentions and preferences in order to inform driver of the optimal operating mode to minimize fuel consumption [46]. The driver's style can be roughly divided into aggressive, conservative, and multi-variant styles. Based on the collected data of vehicle speed, acceleration, throttle opening, slope, or slip rate, the classification of driving style can be acquired by K-means clustering or neural network. The driving style, as one of the influencing factors, is beneficial to having a better performance in terms of fuel economy, which is mainly reflected in the impact on shift frequency, or acceleration/deceleration.

Simple fuzzy logic controllers can effectively reduce computational burden; however, they are still based on predetermined rules and lack of optimality in fuel economy for HEVs. Although some adaptive fuzzy logic-based controllers can solve the above problems, they cannot be used easily in real-time implementations due to their computational burden. Moreover, fuzzy logic controllers need more than one set of rules to optimize a system if control variables/objectives are multiple, which will increase the strategy's complexity.

To achieve practical application while ensuring a good performance of control results, the rule-based strategy can be improved by combining intelligent transportation systems. As for the prediction of vehicle operation in the future, the trip information can be considered as one of the determining factors, and can be obtained by intelligent transportation systems (ITSs), geographical information systems (GISs), global positioning systems (GPSs), or advanced traffic flow modeling techniques [47,48]. When trip information is predictable to a large extent, EMS becomes a global optimization problem, which can utilize DP to improve the optimality of RB strategy in the process of offline optimization. As described in [49], a rule-based EMS was proposed for PHEVs. Based on historical traffic information, the driving cycle is modeled, and the DP algorithm is applied to reinforce the charge-depletion control to ensure that SOC drops to a predefined value at the end of the driving cycle.

In summary, rule-based control strategies have been widely used in EMS for HEVs due to their convenience in adjusting parameters and practical applications. Computational time and memory

storage of rule-based strategies are acceptable for online control. However, the drawback of such strategies is that they tend to fall into local optimal solutions rather than global optimal solutions, that is, they cannot fully optimize fuel consumption over the whole trip.

### 3. Optimization-Based Strategies

Theoretically, the local optimization is to find the minimum value of the objective function in a limited space, while the global optimization is to find the minimum value of the objective function in the whole space. If the information for the entire driving cycle is acquired in advance, the optimal control consequence at each moment can be obtained by minimizing the objective function of the entire driving cycle, which means that the global optimal fuel economy can be obtained. The corresponding control strategy is implemented based on the known global driving information, which is defined as global information-driven optimization. If only the current state of the vehicle is available in the entire drive cycle, the optimal control sequence at each moment is obtained by minimizing the objective function of the current moment. Relative to the entire drive cycle, the local optimal fuel economy is obtained. The corresponding control strategy is implemented based on the current driving information, which is defined as instantaneous optimization.

Therefore, the optimization-based control strategy contains instantaneous optimization and global information driven optimization. Equivalent consumption minimization strategy (ECMS) and model predictive control (MPC) strategy belong to instantaneous optimization. When the information for the entire driving cycle is acquired in advance, the deterministic dynamic programming (DDP) and Pontryagin's minimum principle (PMP) strategies are designed to achieve theoretical global optimal fuel economy over the entire driving cycle. Equivalent consumption minimization strategies (ECMS) expresses electrical energy as an equivalent fuel quantity by introducing an equivalence factor, which can achieve online control due to requiring less computational time compared with DDP. Meanwhile, due to the prior knowledge of the driving cycle and the uncertainty of the optimal equivalent factor, real-vehicle application of this strategy is limited.

The deterministic dynamic programming strategy can obtain global optimal fuel economy; however, it suffers from the "curse of dimensionality", and can only be implemented offline, owing to its time consumption and requirement of a tremendous amount of memory. Additionally, the entire driving cycle information needs to be acquired in advance when applying the DDP or PMP algorithms, which is extremely difficult in real applications for energy management in HEVs. To overcome these issues, the stochastic dynamic programming (SDP) strategy was designed to optimize power distribution between different energy sources based on the prediction of driving condition.

Since the driving information is updated in real time, the data-driven control strategy is adopted to achieve real-time optimization of energy management for HEVs, which includes the neural network–dynamic programming (NN–DP) strategy, the reinforcement learning (RL) strategy, and the adaptive dynamic programming (ADP) strategy.

#### 3.1. Instantaneous Optimization

Instantaneous optimization strategies mainly consist of the equivalent consumption minimization strategy and the model predictive control strategy, which can achieve real-time optimization control. The focus of ECMS is the determination and adaptive adjustment of optimal equivalent factor, while the key points of the MPC strategy are short-term speed prediction and the short-term optimization of power distribution.

##### 3.1.1. Equivalent Consumption Minimization Strategy (ECMS)

The equivalent consumption minimization strategy (ECMS), as an instantaneous optimization algorithm, can obtain a near-optimal control strategy and can be implemented online due to its smaller requirements in terms of storage memory and computation time compared with DDP-based strategies.



The core idea of ECMS is to convert electricity consumption of electric machines into fuel consumption by using an equivalent factor (EF), and adds this to the actual fuel consumption of the engine to obtain the equivalent fuel consumption at each moment. The equivalent fuel consumption is chosen as the objective function to solve the optimal energy distribution of HEVs at the minimum. The objective function can be expressed as:

$$\min J(t) = \min \int_0^t \dot{m}_{eq}(\tau) d\tau \quad (1)$$

$$\dot{m}_{eq}(t) = \dot{m}_{eng}(t) + \dot{m}_{bat\_eq}(t) \quad (2)$$

$$\dot{m}_{bat\_eq}(t) = \gamma \dot{m}_{batt}(t) = \begin{cases} \gamma \cdot \frac{P_{batt}(t)}{Q_{hw}} \eta_{chg}(t) P_{batt} < 0 \\ \gamma \cdot \frac{P_{batt}(t)}{Q_{hw}} \frac{1}{\eta_{dis}(t)} P_{batt} \geq 0 \end{cases} \quad (3)$$

where,  $\dot{m}_{eq}$  is the total equivalent fuel consumption rate of the HEV,  $\dot{m}_{eng}$  is the actual fuel consumption rate of the engine,  $\dot{m}_{bat\_eq}$  is the equivalent fuel consumption rate of the battery,  $P_{batt}$  is battery power,  $Q_{hw}$  is fuel calorific value,  $\eta_{chg}$  is battery charging efficiency,  $\eta_{dis}$  is battery discharging efficiency, and  $\gamma$  is the equivalent factor, whose value is positive during the battery discharging process, while being negative during the battery charging process.

It is worth noting that the results of ECMS are very sensitive to the equivalent factor, which is influenced by driving conditions, battery SOC, driving style, road gradient, and so forth. If the equivalent factor is too large, the strategy will prefer to use fuel, which results in increasing fuel consumption and higher battery power. Conversely, the strategy will tend to use electricity, which leads to excessive power consumption and lower battery power. Generally, the equivalent factor can be calculated by the efficiency of oil-electric conversion  $\eta_{fuel\_batt}$ , which is defined as the efficiency of converting engine's chemical energy into electrical energy to store in the battery through generator. The efficiency  $\eta_{fuel\_batt}$  can be formulated as:

$$\eta_{fuel\_batt} = \eta_e \eta_{gc} \eta_{chg} \quad (4)$$

where,  $\eta_e$  is engine efficiency,  $\eta_{gc}$  is generator efficiency.

Considering the average effect of a large number of different operation points during the driving cycle, the efficiency  $\eta_{fuel\_batt}$ ,  $\eta_e$ ,  $\eta_{gc}$ ,  $\eta_{chg}$  can be calculated by  $\bar{\eta}_{fuel\_batt}$ ,  $\bar{\eta}_e$ ,  $\bar{\eta}_{gc}$ ,  $\bar{\eta}_{chg}$ , respectively. Where  $\bar{\eta}_e$  is the average efficiency of the engine,  $\bar{\eta}_{gc}$  is the average efficiency of the generator,  $\bar{\eta}_{chg}$  is the average charging efficiency of the battery, and  $\bar{\eta}_{fuel\_batt}$  is the average conversion efficiency, which satisfies  $\bar{\eta}_{fuel\_batt} = \bar{\eta}_e \bar{\eta}_{gc} \bar{\eta}_{chg}$ . In the end, the equivalent factor can be calculated by the following formula:

$$\gamma = \begin{cases} \frac{\bar{\eta}_m \bar{\eta}_{dis}}{\bar{\eta}_e} P_{batt} < 0 \\ \frac{1}{\bar{\eta}_{fuel\_batt}} P_{batt} \geq 0 \end{cases} \quad (5)$$

where  $\bar{\eta}_m$  is the average efficiency of the motor,  $\bar{\eta}_{dis}$  is the average discharging efficiency of the battery.

In addition, the instantaneous fuel consumption rate of the engine  $\dot{m}_{eng}$  can be calculated by the interpolation of engine torque  $T_e$  and engine speed  $w_e$ , and the battery power  $P_{batt}$  can be calculated by the change of SOC and the battery's terminal voltage. Therefore, battery SOC and vehicle speed  $v$  can be chosen as the state variables, and  $T_e$ ,  $w_e$  or  $T_m$ ,  $w_m$  can be chosen as the control variables. Based on the above state variables and control variables, the main research areas of ECMS are shown in Figure 4, including the determination and optimization of the equivalent factor, and the solution of the ECMS algorithm with constraints.

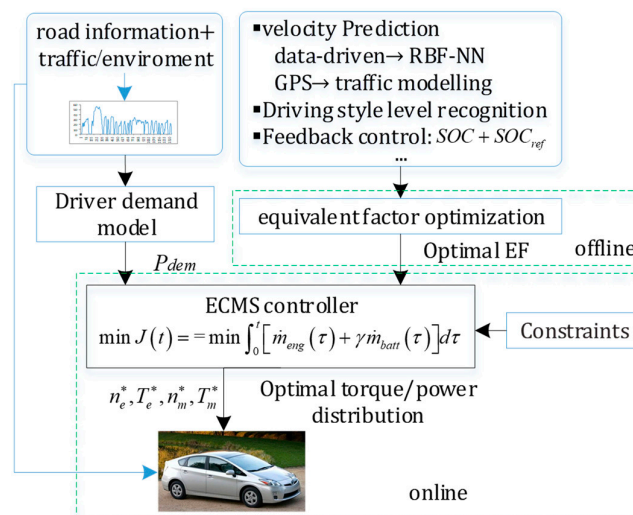


Figure 4. Depiction of the main research areas of ECMS.

For a given driving cycle, the optimal equivalent factor can be obtained by repeatedly adjusting the equivalent factor until the terminal SOC is equal to the initial SOC in the offline calculation. Meanwhile, the equivalent factor should be adaptively adjusted due to uncertain future driving conditions. Therefore, the determination and adaptive adjustment of equivalent factor are the keys of ECMS. In [50], the optimal equivalent factor  $\lambda^*$  was determined by the full trajectory of the driver's demanded power, and the time-varying equivalent factor  $\lambda$  was set as an estimate of  $\lambda^*$  to achieve the adaptive ECMS, which was not needed in order to predict vehicle speed or perform horizon optimization.

However, the optimal equivalent factor  $\lambda^*$  of ECMS can be calculated only if the whole driving cycle is known in advance. Instead of direct estimation of the optimal equivalent factor in ECMS, the optimal equivalent factor  $\lambda^*$  can be estimated through the estimation of the upper and lower bounds of the optimal equivalent factor, which are functions of the HEV's configuration and independent of the driving cycle. The idea of the upper and lower bounds of EF can be employed in designing adaptive ECMS (A-ECMS) [51], which can obtain a charge-sustaining solution and minimize the total fuel consumption. The adaptive ECMS (A-ECMS) strategy as an online control strategy is introduced to estimate  $\lambda^*$  at each moment, which can be categorized into instantaneous A-ECMS and predictive A-ECMS.

Based on the theoretical analysis of Pontryagin's Minimum Principle and introducing soft constraints inside the range, an ECMS–CESO (catch energy-saving opportunity) strategy for series HEVs as the instantaneous A-ECMS is introduced in [52], which can achieve near-optimal fuel economy without the need for predicting future driver demand. The cost function of ECMS–CESO is penalized when SOC exceeds the soft bounds and uses battery power, and its penalty factor is equivalent to the equivalent factor for the latter. A driving-style-oriented adaptive ECMS is developed in [53], which classifies drivers into six groups, from moderate to aggressive, using kernel density estimation and entropy theory. According to driving style, the EF is tuned based on the relationship between SOC and power demand to improve the fuel economy and charge sustainability of HEVs.

An adaptive ECMS with velocity prediction as predictive A-ECMS was proposed to adaptively adjust the equivalence factor in real time [54]. The velocity predictor was constructed by neural network to forecast short-term future driving behaviors based on historical data, and the bisection method was utilized to guarantee the convergence of EF. By deriving the theoretical relationship between the optimal EF and future driving statistics, a strategy synthesized with predictive ECMS was proposed to develop the causal adaptation law to adjust equivalence factor [55]. In this work, the non-causal behavior of EF in various driving scenarios was analyzed using the DP-based extraction method, which obtains the equivalence factor corresponding to optimal SOC trajectory and the optimal control policy by using the DP algorithm.

In addition, RB-ECMS, combining rule-based strategy and ECMS, was developed to further improve fuel economy while ensuring drivability and battery charge-sustaining for HEVs. The equivalent factor in ECMS can vary with power demand and battery SOC, which can be obtained online from a 2-dimensional mapping established offline, which can be optimized by intelligent algorithm.

ECMS, as an instantaneous optimization algorithm, can compromise approximate global optimality and computational burden; however, it is implemented over a driving cycle provided in advance. Even though previous knowledge of the future driving cycle can be predicted based on trip distance, future traffic and terrain conditions, the prediction accuracy and the value of the optimal equivalent factor will have a significant influence on the optimality of the EMS, which limits the further application of this strategy.

### 3.1.2. Model Predictive Control (MPC)

Common approaches for online energy management reduce the computational time by implementing optimization for a short-time finite horizon of the entire trip and repeating it at every time step. The MPC [56–58] strategy, as one of the instantaneous optimization control strategies, and the rolling time domain control, have been extensively studied theoretically and applied in different fields, and is able to achieve an approximate fuel economy to DP and can be implemented online with limited computation and memory resources. The MPC strategy depends on a short-term prediction of driver-demanded power in the future at each moment, which is different from DP, which requires the whole driving cycle in advance.

Model predictive control (MPC) is based on rolling optimization, which converts the optimization process into a limited prediction horizon to reduce calculations and has the potential for real-time control. The short-term prediction can be studied from the perspective of the time domain or the distance domain. In the time domain, each location has the same time of  $\Delta t$ , and control inputs are updated with a fixed time step. Similarly, in the distance domain, each location has the same distance of  $\Delta l$ , and control inputs are updated with a fixed distance step. The majority of studies have implemented the prediction from the perspective of the time domain. With respect to the distance domain, a distance-based ecological driving scheme with long-term speed optimization and short-term adaptation is proposed in [59]. The optimal speed profile for the entire route is optimized by road conditions, and the QP method is used for optimization to save computation time. In the short term, the speed at the next location is adapted by the spacing to the preceding vehicle.

The main procedures of the MPC-based strategy are as follows, and the main research areas are summarized in Figure 5:

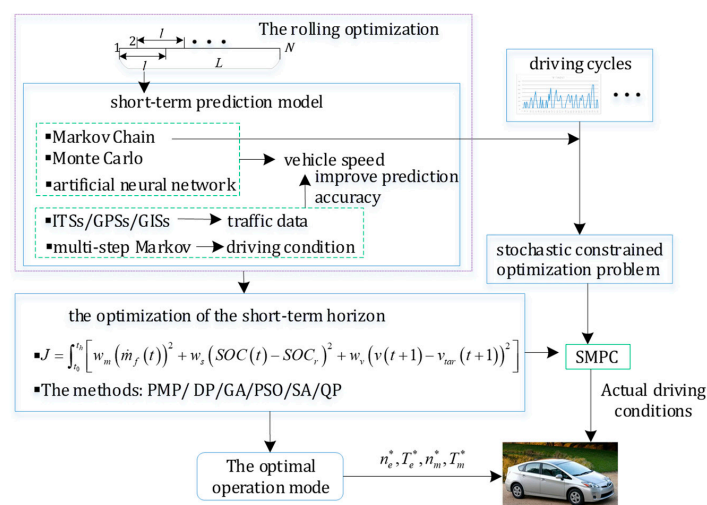


Figure 5. Depiction of the main research areas of the MPC-based strategy.

Step1: Predict vehicle velocity in a prediction horizon, and obtain optimal control trajectory in this horizon based on minimizing objective function with multivariable constraints. If taking fuel consumption, the performance of driving, and battery SOC reference trajectory into consideration, the cost function can be written as:

$$J = \int_{t_0}^{t_h} \left[ w_m (\dot{m}_f(t))^2 + w_s (SOC(t) - SOC_r)^2 + w_v (v(t+1) - v_{tar}(t+1))^2 \right] dt \quad (6)$$

where  $t_0$  is the current time,  $t_h$  is the terminal time during a prediction time horizon,  $w_m, w_s, w_v$  are penalty weights,  $SOC_r$  is the value of reference SOC,  $v$  is vehicle speed,  $v_{tar}$  is the target speed.

- Step2: Implement the first element of the optimal control sequence in the corresponding vehicle model and send the feedback adjustment of the estimated SOC to the optimization section after the vehicle has responded;
- Step3: Move the entire prediction horizon one step forward, that is, refresh the optimization problem with the latest measured value at each moment;
- Step4: Repeat the steps 1 to 3.

In general, the Markov Chain method, Monte Carlo Method [60] or artificial neural network algorithm can be utilized to predict vehicle speed. Markov Chain Monte Carlo Method (MCMC) can be adopted to predict vehicle velocity with post-processing algorithms including average filtering and quadratic fitting to moderate fluctuations of the prediction results, which can effectively improve the predictive accuracy and enhance the control performance of MPC-based strategy [24]. In [61], a velocity predictor based on current driving environment and vehicle information is developed to predict future driving conditions, which are established by the radial basis function neural network (RBF-NN). Meanwhile, the nonlinear model predictive control with forward dynamic programming is utilized to construct the master controller, which obtains optimal control variables online.

To improve prediction accuracy, intelligent transportation systems (ITSs) with some vehicular telemetry technologies including onboard GPSs, geographical information systems (GISs), and advanced traffic flow modeling techniques [49,62] can be used to access information of the traffic or road conditions. Based on the traffic data from telematics, the battery SOC profile can be scheduled systematically by considering the effect of road conditions, battery and real-time implementation ability, which can efficiently improve fuel economy based on the MPC strategy [63]. In addition, a multi-step Markov prediction method can be used to predict the driving conditions, and the DP method can be used to solve the optimization problem within the prediction horizon [64].

As for the optimization of the short-term horizon, the methods used could be the PMP algorithm, the DP algorithm, or an intelligent algorithm, such as the genetic algorithm (GA), particle swarm optimization (PSO), simulated annealing (SA), quadratic programming (QP), and so forth. As described in [65], a predictive energy management strategy is proposed with an online correction algorithm in the optimal energy management strategy of PHEVs, which is optimized with the dynamic neighborhood particle swarm optimization (PSO) algorithm. In [66], several quadratic equations were employed to determine the engine fuel rate with respect to battery power, using the QP and SA methods together to find the optimal battery power commands and engine-on power.

The Markov chain is a series of transition probabilities from one of the limited states at instant  $k$  to another state of all possible states at instant  $(k+1)$ , which is mainly used to model and solve dynamic decision-making problems. The key of the Markov chain is to obtain the transition probability matrix  $T$ , whose element can be defined as:

$$T(i, j) = Pr[x(k+1) = x_j | x(k) = x_i] (i, j \in 1, 2, \dots, N) \quad (7)$$

where the matrix  $T$  represents the transition probability from the state  $x_j$  at the moment  $k$  to the state  $x_i$  at the moment  $k + 1$ .

In the MPC-based strategy, the future torque/power demand of powertrain can be modeled by Markov chain; therefore, this problem can be regarded as a kind of stochastic constrained optimization problem for nonlinear systems. Based on the above theory, a novel stochastic model predictive control (SMPC) [67,68] strategy is developed to contribute the practical application of the MPC-based strategy. To improve real-time performance, an SMPC-based energy management strategy for PHEB uses state reconstitution method to guarantee the continuity of practical application, and employs time-varying predictive steps by an online accuracy estimation method and a corresponding threshold to maintain the prediction accuracy [24]. In [25], the SMPC is modified with the equivalent consumption minimization strategy for PHEB, which considers the reference SOC trajectory in the finite predictive horizons to eliminate undesirable working points. Based on SMPC with learning (SMPC-L), a driver-aware vehicle control was developed in [69], using Markov chain to adapt to changes in driver behavior, and the QP method was used for optimization to handle larger state dimension models. The simulation results showed that the performance of the proposed approach was close to that of MPC with full knowledge of future driver power requests in standard and real-world driving cycles.

Due to the fact that the road conditions significantly influence the battery charging and discharging processes of HEVs, it is worth considering that the previewed road grade can be used to predict the future power demand to improve the performance of energy management without the route being determined in advance. Ref. [70] proposed an SMPC-based energy management strategy to maintain battery SOC within its boundaries and achieve good energy consumption performance. Based on vehicle location, traveling direction, and terrain information for HEVs running in hilly regions, the road grade for stochastic routes was modeled by a finite-state Markov chain model, and the vehicle speed profile was modeled using a similar method.

Factors including prediction accuracy, design parameters, and optimization solution will influence the control performance of the MPC strategy, which means that it cannot guarantee robustness and accuracy when the actual driving cycle is quite different from the training driving cycle. Specifically, prediction uncertainty can heavily affect the performance of the MPC-based strategy for unknown routes with unknown driver behavior. Similarly, as for the SMPC strategy, the transition probabilities in the Markov chain are based on collected driving cycles [71], that is, it cannot guarantee prediction accuracy or the optimality of energy management when the real driving conditions differ from the collected data.

### 3.2. Global Information-Driven Optimization

Global information-driven optimization control strategies look for the optimal control sequence based on global information to obtain the minimum fuel consumption, and include the Pontryagin's minimum principle (PMP) strategy, the deterministic dynamic programming (DDP) strategy, and the stochastic dynamic programming (SDP) strategy. These strategies can obtain the global optimal energy management over the whole driving cycle, and the results of the DDP-based strategy can be regarded as a benchmark for assessing the optimality of other energy management strategies.

#### 3.2.1. Pontryagin's Minimum Principle (PMP)

The minimum principle can be expressed as: the optimal trajectory determined by the optimal control law must be the minimum value within the whole control domain when the control variables are limited to a certain range, which can be applied in the energy management for HEVs. The PMP [72,73] method solves the optimal energy management by finding the instantaneous minimum value of Hamiltonian function at each moment, which is generally defined as the derivative of the stage cost function versus time represented by the adjustable variable as independent variable. The optimal control variables  $U$  will be obtained when the derivative of Hamiltonian function  $H$  versus control variables equals to 0, that is,  $\partial H / \partial U = 0$ . Therefore, Pontryagin's minimum principle provides a



necessary condition, but not a sufficient condition for optimality. When the obtained local optimal trajectory is a unique trajectory that satisfies necessary boundary conditions and constraints and the state function can be used to determine whether the results are the optimal solution, the optimal trajectory of PMP-based strategy is regarded as the global optimum solution.

The essence of the PMP-based strategy for HEVs is to minimize the sum of fuel consumption at each moment under determined terminal constraints, the objective function, and the admissible control range. Within the boundary condition of control variables, the value of the Hamiltonian function, along with the optimal trajectory, is a constant; thus, the optimal control sequence can be obtained by finding the minimum value of performance index function. Thus, the performance index function can be expressed as:

$$J(u) = \int_{t_0}^{t_f} L(x(t), u(t), t) dt = \int_{t_0}^{t_f} \dot{m}_f(u(t), t) dt \quad (8)$$

where  $x$  is the state variable,  $u$  is the control variable,  $\dot{m}_f$  is the fuel mass flow rate (g/s),  $L(\cdot)$  is the objective function,  $t$  is the time,  $t_0$  is the initial time of driving cycle, and  $t_f$  is the terminal time of driving cycle.

Taking the battery SOC (or fuel rate) as the state variable and the power battery, engine speed, engine torque (or torque split ratios) as the control variable, the state equation and boundary condition of the system are formulated as follows:

$$\dot{SOC}(t) = -\frac{U_{OC} - \sqrt{U_{OC}^2 - 4R_{int} \cdot P_{batt}}}{2R_{int} \cdot C} \quad (9)$$

$$\begin{cases} SOC(t_0) = c \\ SOC(t_f) = SOC(t_0) \\ SOC_{min} \leq SOC(t) \leq SOC_{max} \\ P_{batt\_min} \leq P_{batt}(t) \leq P_{batt\_max} \end{cases} \quad (10)$$

where  $U_{OC}$  is the open-circuit voltage,  $R_{int}$  is the internal resistance of the battery,  $P_{batt}$  is the electric power,  $C$  is the nominal capacity of the battery with a fixed value,  $c$  is a constant,  $SOC(\cdot)$  is the boundary values of state variable,  $P_{batt}(\cdot)$  is the boundary values of control variable,  $SOC_{min}$  is the minimum SOC value,  $SOC_{max}$  is the maximum SOC value,  $P_{batt\_min}$  is the minimum value of power battery, and  $P_{batt\_max}$  is the maximum value of the power battery.

When solving the optimal problem based on the minimum principle, it is necessary to introduce the Hamiltonian function and the covariant variable. The Hamiltonian function can be defined as:

$$H(x(t), u(t), \lambda(t), t) = \dot{m}_f(t, u) + \lambda(t)f(x(t), u(t), t) \quad (11)$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x} \quad (12)$$

where  $f(\cdot)$  is the state equation of the system,  $\lambda$  is the covariant variable, which is related to the state transfer function.

In the solving process, it is worth considering the complexity of the solutions to the boundary value problem. On the one hand, the common method for solving the above problem is the shooting method, which converts the boundary value problem to a problem of solving several initial values. The terminal SOC value is calculated by the state equation and the costate equation. The error between the terminal SOC value and the initial SOC value can be calculated, and then the error can be brought within the target range ( $\delta$ ) by readjusting the value of the covariant variable, that is,  $|SOC(t_f) - SOC(t_0)| \leq \delta$  [74]. The method of linear interpolation or the Newton method can be used to correct the value. On the other hand, the value of the state variable must satisfy the boundary condition  $SOC(\cdot)$ , which can be defined as the hard boundary. To avoid missing the optimal value on the boundary, [50] proposed the

concept of soft bounds inside the hard bounds, defined as  $SOC_{min}^{soft} \leq SOC(t) \leq SOC_{max}^{soft}$ . The value of SOC can exceed the soft bounds by an allowable amount; however, exceeding those bounds will be penalized by increasing or decreasing  $\lambda$ .

Compared with the rule-based strategy, the PMP-based strategy has a significant improvement in fuel economy and the battery SOC at the final moment is unchanged [74]. The PMP-based strategy combining with the intelligent algorithms can be developed to improve the optimal performance of control strategy. In [75], the PMP-based strategy combining with SA algorithm is applied to determine the battery current command and engine-on power using a series of quadratic equations to approximate fuel-rate when the engine is operating. Moreover, the proposed algorithm is validated with by considering the battery's state of health (SOH) to extend the application.

In addition, the selection of covariant variable  $\lambda$  can directly affect the optimal power distribution between the power sources, meaning that it will affect the fuel economy and driving performance of HEVs. Therefore, the influence of the co-state variable on the optimal solution of the PMP-based strategy and its determination in terms of certain driving cycles and uncertain driving cycles should be discussed, as detailed in [76]. According to the restriction of the state variable, the periodically updated co-state variable is used for uncertain driving cycles, and the initial  $\lambda$  is given in advance.

The Hamiltonian function is a complicated function of the control variables, and the process of seeking an optimal control variable needs to traverse the whole domain of the control variables, which leads to a large computational burden and difficulty in implementing real-time control. To overcome the challenge of optimizing the Hamiltonian function, the quadratic performance index is widely used, which is an explicit solution for Hamiltonian optimization by applying quadratic fitting to both engine and electric powertrains. Generally speaking, the consideration of engine on/off control is not necessary for a HEV because the engine needs to be on due to the limited electrical energy. However, the engine on/off control should be considered in PHEV because the engine can be kept off for a relatively long period of time due to the sufficient electrical energy obtained from the extra grid.

Aiming to solve the problem mentioned above, [77] introduces the energy management of PHEVs based on an approximate PMP (A-PMP) while utilizing a piecewise linear approximation, which was determined by specifying the turning point of the engine fuel rate. In the A-PMP strategy, the engine on command is separated from the engine torque command, and the engine state with the smaller Hamiltonian function is considered to be the engine on command.

Based on the above research, the main research areas of the PMP-based strategy are described in Figure 6.

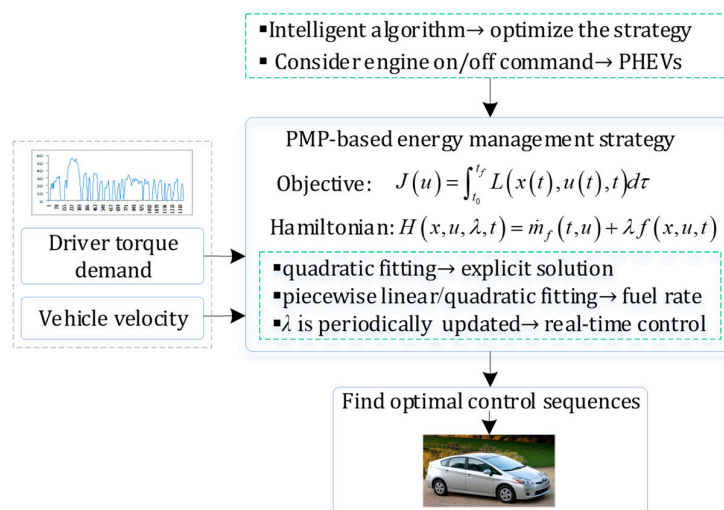


Figure 6. Depiction of the main research areas of the PMP-based strategy.

Although the results of the PMP-based strategy are very close to DP, and it has a lower computational burden compared to the DDP-based strategy, there are still some challenges of PMP. On the one hand, the quadratic fitting to the engine fuel rate is used to simplify the engine fuel map, which changes some important features of the engine. On the other hand, the challenges of solving the Hamiltonian function and acquiring the driving cycle in advance result in difficulty in the real-time application of the PMP-based strategy; therefore, it is necessary to develop approximate PMP to enhance computational efficiency and reduce the computational burden to implement online control.

### 3.2.2. Deterministic Dynamic Programming (DDP)

Applying deterministic dynamic programming (DDP) [78,79] in the EMS of HEVs could greatly improve the fuel economy over the whole driving cycle, which provides a benchmark for assessing the optimality of other energy management strategies despite the off-line optimization.

Assuming that the state equation of a nonlinear system is expressed as:

$$\dot{x}(k+1) = f(x(k), u(k), k) \quad (13)$$

where  $x$  is the state variables,  $u$  is the control variables,  $k = 0, 1, 2, \dots, N-1$ , and  $N$  is the stage number of the whole driving cycle. Under the action of the control variables  $u$ , the cost function can be expressed as:

$$J = \sum_{k=0}^{N-1} U(x_k, u_k) \quad (14)$$

where  $U$  is the instantaneous cost at every moment.

The core of dynamic programming is Bellman's optimal principle [80], which can be solved by dealing with the optimization of a sequence of sub-problems, and is implemented backward from the terminal state to the initial state of the driving cycle by searching for the optimal trajectory among each state. Unlike to Pontryagin's minimum principle, the principle of optimality provides sufficient conditions for optimality, which need to solve the nonlinear Hamilton-Jacobi-Bellman (HJB) equation. According to the principle of optimality, the basic recursive equation of dynamic programming can be obtained:

$$J_k^*(x_k) = \min_{u_k} [U(x_k, u_k) + J_{k+1}^*(x_{k+1})] \quad (15)$$

where,  $J_k^*$  is the optimal cost-to-go function at state  $x_k$  in the  $k$ th ( $0 \leq k \leq N-1$ ) step,  $x_{k+1}$  is the next state, which is obtained under  $u_k$  being applied to the current state  $x_k$  at the  $k$ th step.

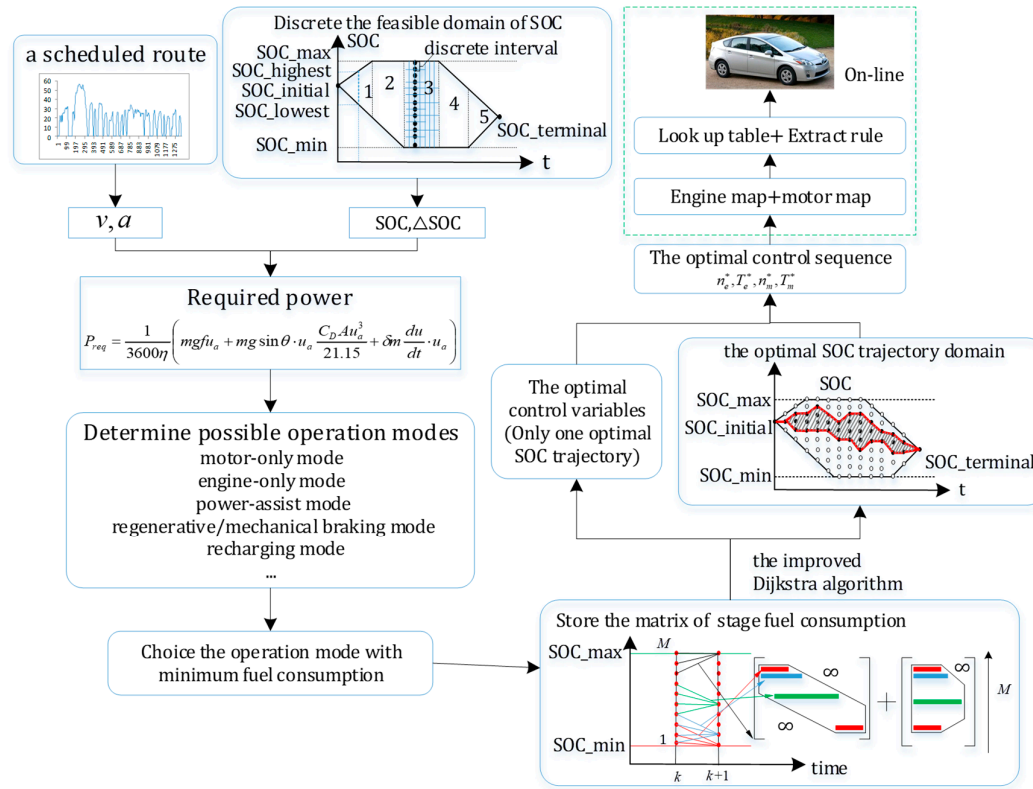
The objectives of DP in HEVs are to find optimal control sequences to obtain the optimal SOC trajectory and minimize fuel consumption over a given driving schedule. That is, the optimal cost function of each step can be regarded as the minimum fuel consumption for that stage. Consequently, how to determine the minimum fuel consumption of each stage and then search for the optimal trajectory to obtain global optimal fuel economy are crucial issues for the DDP-based strategy.

In addition, because that dynamic programming is a numerical algorithm, the state variables, control variables and continuous time should be discretized before formulating the DP, and the boundary issue of their feasible regions should be considered when implementing DP to solve the optimal control problem of HEVs. In addition, the physical constraints on the states and the inputs should be considered to ensure safe operation of the components (such as the engine, motors and battery). Generally, the state variables usually include the vehicle speed  $v$  and SOC, and the control variables could be selected from among engine torque, engine speed, motor torque, engine speed, and power-split ratio. Based on the Rint model, the battery SOC of the battery can be formulated as [20]:

$$SOC_{k+1} = SOC_k - \frac{U_{OC} - \sqrt{U_{OC}^2 - 4R_{int} \cdot P_{batt}}}{2R_{int} \cdot C} \quad (16)$$

where  $SOC_{k+1}$  is the battery SOC at the  $(k+1)$ th step.  $U_{OC}$  is the open-circuit voltage, which is related to the SoC and the battery temperature  $T_{bat}$  and can be obtained from the interpolation function  $U_{OC} = f(SOC, T_{bat})$ .  $R_{int}$  is the internal resistance of the battery, which is related to SOC,  $T_{bat}$ ,  $I$  and can be obtained from the interpolation function  $R_{int} = f(SOC, T_{bat}, I)$ ,  $P_{batt}$  is the electric power, and  $C$  is the nominal capacity of the battery, which is a fixed value.

Therefore, the main steps of DDP-based strategy can be concluded as follows, and the main technical routes are shown in Figure 7:



**Figure 7.** Depiction of the main technical areas of the DDP-based strategy.

- Step 1: Discretize the relevant variables and determine the boundaries of the feasible regions. The shape of the SOC feasible domain is determined by the highest SOC, the lowest SOC, initial SOC, terminal SOC and the maximum charge/discharge current limitation.
- Step 2: Determine the SOC value between each adjacent grid point. According to the start time of SOC rising/falling and the terminal time of SOC rising/falling, the feasible region of SOC would be divided into several districts, and then the highest and lowest SOC at each moment would be determined. Limited by the maximum SOC discrete interval, the number of grid points (i.e., discrete points) at each moment can be calculated in order to obtain the SOC value of each grid points.
- Step 3: Determine the possible operation modes according to the variation of vehicle speed, vehicle acceleration and battery SOC. In general, the operation modes roughly include motor-only mode, engine-only mode, hybrid mode (motor-assist), regenerative braking mode, mechanical braking mode, recharging mode, and stop mode.
- Step 4: Choose the operation mode with the minimum fuel consumption, and the fuel consumption will be stored in the form of a matrix. Moreover, using a three-dimensional matrix to record fuel consumption between each moment contributes to improving the computational efficiency.
- Step 5: Search for all the optimal SOC trajectories among the discretized grid points using the improved Dijkstra algorithm based on the fuel matrix.

Step 6: Obtain the optimal control sequences of the engine and motors and the optimal SOC trajectory domain.

In general, DP is employed to locate the optimal actions for the engine at each stage by minimizing the fuel consumption cost function over a certain driving cycle in HEVs [81]. In [20], the torque  $T_e$  and  $T_{TM}$  were chosen as the independent control variables and the vehicle speed  $v$  and SOC reflecting the state of the PHEB were chosen as state variables. As a result, the results of the DDP-based strategy were an approximately 20% improvement in fuel economy compared with the traditional control strategy.

In implementing the DDP-based strategy, researchers usually find all possible operation modes of the HEV and all possible control solutions in each mode in order to obtain all control solutions at every grid point of the battery SOC, and then the optimal controls of each stage are obtained by backward solution based on the specified initial SOC or interpolation; however, this results in a heavy computational burden. When using the DP algorithm, it is necessary to find a compromise between the optimization accuracy and the computational burden. One of the reasons for the tremendous computational burden in the calculation process is the storage of the fuel matrix. There are three different manners in which to store the fuel matrix. An intuitive one is point to point, which uses the loop inside the loop at each time. The second is an intuitive matrix method based on an  $M \times M$  matrix, which usually leads to a highly sparse matrix. The third is a method using a dense matrix with column by column (rather than line by line), which contributes to having a smaller loop to construct it, thereby reducing the computational time. To effectively solve the offline issue of DP, the different possibilities for reducing the computational time were investigated in [82,83], and were introduced from two perspectives: restraining the area of the feasible domain, and changing the solutions from an intuitive manner to a complex matrix manner. As for the feasible area of SOC, the drawback of the non-regular grid is that the maximum discharge limits have virtually no chance of corresponding to an effective edge, while the regular grid creates errors on the SOC variations, and these are accumulated throughout the whole cycle.

The interaction mechanism between the accuracy of the optimal results and the numerical issues of DP (such as the discretization degree of the state/control variables and the boundary issue) should be investigated in consideration of the computational burden in order to have the potential for practical application. Thus, [84] presents a DP-based EMS for the Toyota Hybrid System (THS-II) powertrain of the Prius. The potential of this DP-controller in reducing fuel consumption on regulatory cycles is better than a rule-based controller, and the fuel reduction enhancements of the DP-controller were achieved in real road tests in a MY06 Prius in Ile-de-France, when compared with the associated consumption measurements.

To realize the adaptive energy management for real-time driving cycles, an adaptive energy management method based on DP for a PHEV was proposed in [85]. This approach classifies typical driving cycles into different driving patterns by considering the average and maximum speed as its classification parameters and identifies the driving types for the driving pattern recognition (DPR) process by a fuzzy logic controller. Meanwhile, the previous duration of historical information is determined to identify a real-time driving pattern in order to improve the real-time and robust performance of the energy management.

Although the DDP-based strategy is powerful and effective in obtaining the global optimal fuel economy of HEVs, it presents some disadvantages for practical application:

- The whole driving information needs to be acquired in advance;
- The method is difficult to implement in practical applications because of the tremendous time and memory consumption of the calculation process;
- The phenomenon of the “curse of dimensionality” occurs with the increase in the dimensions of the state variables  $x$  and the control variables  $u$ , namely, the computation and storage capacity of traditional dynamics programming will increase significantly.



### 3.2.3. Stochastic Dynamic Programming (SDP)

The stochastic dynamic programming algorithm uses the existing standard driving cycles or historical driving data as the sample to establish the statistical model of driver demand, and then uses dynamic programming to solve the energy management problem represented by the statistical model. This optimal strategy is the strategy with the lowest expected cost in the general sense, and it can be applied to energy management of various hybrid vehicles with different drive configurations.

Considering a stochastic dynamic programming problem with a finite state space  $S = \{s^1, s^2, \dots, s^N\}$  and a control space  $A$ , the cost function is  $L : S \times A \rightarrow L$ , and the transition probability matrix is  $T(s, a, s')$ . The cost function means that the cost that the state  $s$  moves to the state  $s'$  under the control action  $a$ , while the  $T(s, a, s')$  represents the probability that the state  $s$  moves to the state  $s'$  under the control action  $a$ . As for a certain state  $s$  in the state space of the system, it has a finite set  $A_s = \{a^1, a^2, \dots, a^M\}$  that the system can take action to move to any other state. That is, the system performs based on the state transition matrix  $T$ .

When applying SDP to solve the energy management for HEVs, the objective is to find the optimal control variables under some constraints to minimize fuel consumption. Therefore, the battery SOC, the vehicle speed and the required power can be chosen as state variables, and the engine speed and torque distribution ratio can be chosen as control variables. In addition, the cost function can be expressed as:

$$J = \lim_{N \rightarrow \infty} E \left\{ \sum_{k=0}^{N-1} \beta^k L(x(k), u(k)) \right\} \quad (17)$$

$$J(s') = \min \sum T(s, s') (L(s, s') + J(s)) \quad (18)$$

$$fuel_{k+1}(s') = \min \sum T(s, u, s') (fuel_{step\ k}(s, u, s') + fuel_k(s)) \quad (19)$$

where  $\beta \in [0, 1]$  is the discount to ensure the convergence of the total cost function at the expected iteration step, and the value reflects the importance of the future costs compared with the current costs. The closer the value is to 1, the greater the proportion of future long-term expectations in the total cost function.  $T$  is the state transition probability matrix,  $L(x(k), u(k))$  is the one step reward function in the  $k$  moment, which consists of the fuel consumption of the engine  $fuel_{step\ k}$ ; sometimes, it also includes the electricity consumption of the motors or the penalty term for the deviation of SOC from the expected final value, which can be defined as  $w_s(SOC(t) - SOC_r)^2$ .

The demand power can be modeled using a Markov chain model, which is established by the transition probability matrix. The state transition probability matrix  $T$  can be obtained by the maximum likelihood estimation method as follows:

$$T(i, j) = \frac{m_{ij}}{m_i} \quad (20)$$

where  $m_{ij}$  is the frequency of the requested power,  $m_i$  is the total frequency.

As for the solution to SDP method, there are three typical algorithms: value iteration, policy iteration and modified policy iteration. The value iteration method calculates the optimal cost function  $J^*$ , and then acquires the optimal strategy based on the optimality principle. Unlike the value iteration, the policy iteration contains policy evaluation and policy improvement, which calculates the value function of all states by giving an initial strategy  $\pi_0$  in advance and calculates a series of new strategies based on the greedy strategy, and then repeats the above steps. The modified policy iteration combining the value iteration and the policy iteration is more effective than the above strategies. The value iteration is implemented in the first few steps to obtain a better estimated value and updates the value function based on the Bellman function, and then implements the policy improvement to acquire the optimal strategy.

Therefore, the key points of the SDP-based strategy mainly include the speed prediction and the optimal solver, which are shown in Figure 8. In addition, the main steps of the SDP-based strategy in HEVs are as follows:

- Step1: Establish the discrete-time dynamic systems and determine state variables, the control variables, the optimal objectives and the constraints of the powertrain components (engine, motors, and battery).
- Step2: Establish the Markov chain model of the driver's torque demand. Implement the statistical analysis of the driver's torque demand based on multiple standard driving cycles or historical driving data, and then use the maximum likelihood estimation method to establish a transition probability matrix.
- Step3: Calculate the matrix of the one step reward function and solve the optimization of SDP based on value iteration, policy iteration or modified policy iteration [86]. If the converged condition is satisfied, the iteration will stop.
- Step4: Obtain the optimal operation mode and the power split between engine and motors.

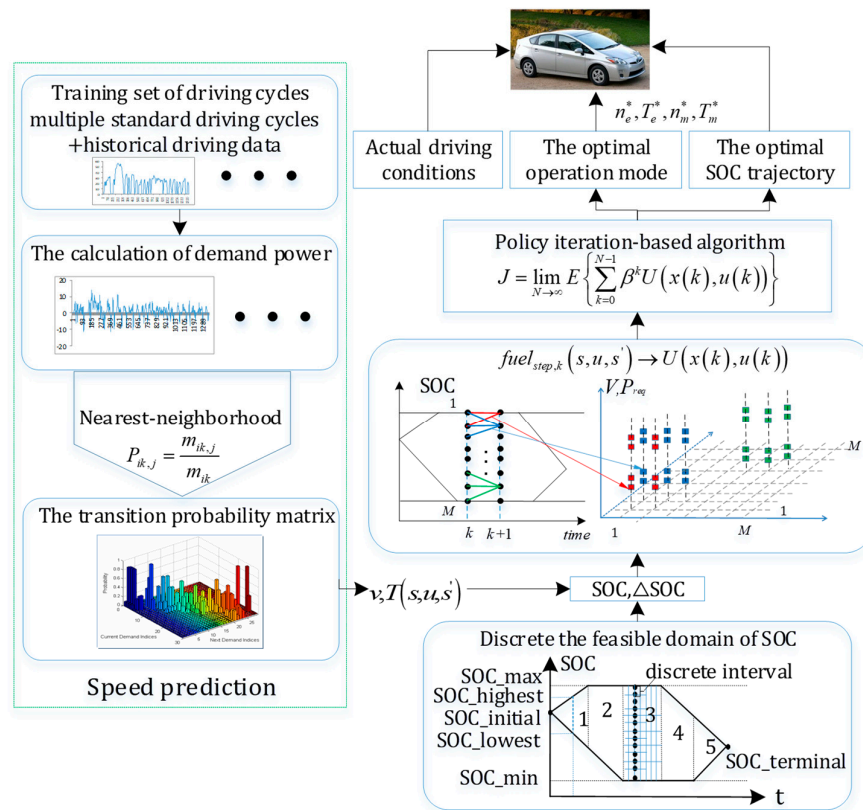


Figure 8. Depiction of the main research areas of the SDP-based strategy.

The future power demand can be formulated as a discrete-time Markov decision process and can be modeled as a stochastic model established by Markov chain or Monte Carlo method, which reflects the probability distribution of the future power demand and the variation of the future driving cycles. In [87], a novel cost function (incorporating the square of battery charge) with a penalty on high-powered systems was used to lessen the affliction of real-world concerns such as battery health and motor temperature. Furthermore, a Markov chain was augmented with the information regarding SOC transitions to complete a full-state transition probability matrix, and the interpolation was used to distribute each state transition between multiple transition probabilities.

Due to the time-invariant characteristics of system dynamics, the optimal EMS can be formulated from the finite horizon and infinite horizon respectively. As for the finite horizon problem, the cost is accumulated in the finite stage, while the infinite horizon problem provides a reasonable approximation of a class of problems, which have a finite range but a very large number [88]. To make the decision-making process independent of driving time, the energy management strategy is solved from the perspective of an infinite horizon in [89]. By considering the statistics of traffic speed profiles, the SDP method based on a modified policy iteration can be adopted to generate a time-invariant state-dependent power split strategy to optimize fuel consumption and charge sustenance of HEVs in a general sense [90]. In [26], a pre-optimization based on the basic operating modes was carried out to obtain the optimal decisions in advance, followed by obtaining the optimal combination of the operating modes and the power split between engine and motor based on SDP. In the SDP algorithm, a homogenous Markov chain in a finite horizon is used to model the driver demand, whose stochastic transition probability matrix is constructed from 15 driving cycles.

In the SDP algorithm, the discount factor can reduce the contribution of the exponentially increasing cost function in the future, and has a significant influence on ensuring the convergence of the cost function. Therefore, it is necessary to consider the impact of different penalty factors on the performance of the optimal control strategy, which is mainly reflected in the SOC domain, the deviation from initial SOC, and the fuel economy. According to the effects of different penalty factors on the optimal results, the most suitable value was chosen as 8 to balance fuel economy and battery performance in [26]. Higher values of  $\lambda$  tend to consider the energy balance costs to make the vehicle operate in a more charge sustaining fashion; however, this requires more iterations to ensure the convergence of the solution and more computational times. Based on the above issues, a compromise between guaranteeing charge sustaining and achieving lower computation time is discussed in [87]. In the end, the value of  $\lambda$  was in the range 0.95–0.995, and the realization of SDP took about 20 iterations, with each policy evaluation step taking about 30 s.

The SDP-based strategy solves the problem that the global driving information of the vehicle needs to be known in advance in the DP-based strategy; however, it needs to predict vehicle speed, and the prediction accuracy would affect the global optimal solution, and the “curse of dimensionality” would occur when the policy-based iteration is used to solve SDP. In other words, the implementation of SDP is not straightforward, and embedded real-time operation may be memory intensive, despite not being computationally intensive. Moreover, the driver demand model was established by using the existing standard driving cycles or historical driving data as the training samples, and it cannot guarantee robustness or accuracy in cases where the actual driving cycle is quite different from the training driving cycle.

### 3.3. Data-Driven Strategies

With the deep learning of neural networks and the rapid development of large data processing technology, data-driven energy management control strategies are increasingly being used for HEVs, which can realize approximate optimal control performance. Data-driven control strategies are introduced from the perspectives of neural network-based dynamic programming (NN-DP), reinforcement learning (RL), and adaptive dynamic programming (ADP) [91].

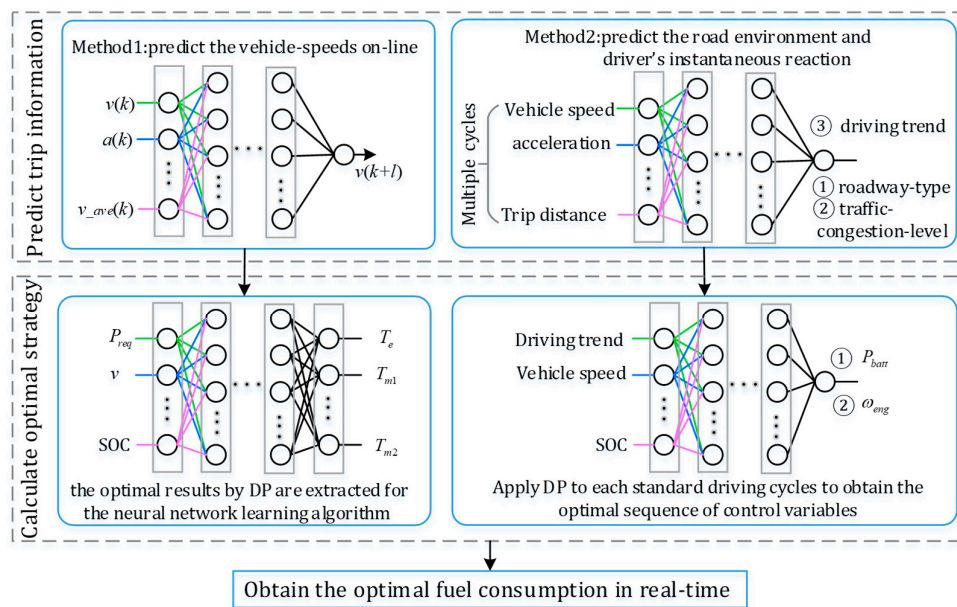
#### 3.3.1. Neural Network-Dynamic Programming (NN-DP)

An artificial neural network is composed of a considerable number of artificial neurons, which are widely interconnected. It can be seen as a directed graph with artificial neuron nodes connected by directed weighted arcs. Artificial neural networks can mainly be divided into feed-forward networks and feedback networks, which can be used to solve optimization problems by the minimum point of energy functions. The back propagation neural network (BP-NN) trains the neural network by using error back propagation; this is the simplest neural network and is widely used in multiple fields at present. Moreover, the radial basis function neural network (RBF-NN) has a better performance for

nonlinear function approximation, generalization ability, and faster learning convergence speed, and is also utilized to construct the model.

Since dynamic programming cannot achieve on-line optimization, neural networks (NN) can be utilized to construct the road environment prediction model and the driving condition prediction model due to their self-learning ability and adaptivity. They can be back propagation neural networks (BP-NN), radial basis function neural networks (RBF-NN) or deep learning neural networks. Based on the simplified neural network module designed to represent trip information, the DP results of standard driving cycle and actual driving cycle can be used to train the neural network module.

There are two ways to effectively combine neural networks with DP. On the one hand, the vehicle speed can be predicted based on current vehicle speed, acceleration, average vehicle speed and so forth, and then the optimal results obtained offline by DP are extracted for the training of neural network, which can select power demand, vehicle speed and SOC as the inputs. On the other hand, it can predict the road environment (including the roadway type and the traffic-congestion level) and the driver's instantaneous reactions according to the vehicle speed, acceleration, trip information, and so forth. The DP method is applied to a set of standard driving cycles to obtain the optimal control sequences and the corresponding driving trend, which are used as inputs of the neural network to obtain battery power and engine torque or motor speed in real time. Based on the above methods, the main research areas for the NN-DP strategy for HEVs are presented in Figure 9.



**Figure 9.** Description of the main research areas for the NN-DP-based strategy.

As for the predetermined bus route, [27,92–94] proposed a length ratio-based neural network energy management strategy for online control of PHEB to reduce the computational time and storage capacity of the micro-controller and to achieve approximate optimal control performance. The length ratio representing the space domain was chosen as the input variable of the neural network module to represent trip information, which consists of four parameters: trip length, trip duration, current driving length and current driving time.

Based on the idea of transforming the global optimization problem into a local optimization problem, the trip condition information at the  $k$ th(s) can be used to predict the future trip condition information according to the prediction transfer relation. The process is from the  $(k+1)$ th(s) to  $(k+l)$ th(s), where  $l$  represents the length of a preview window. Therefore, the vehicle-speed trajectory can be acquired online through the trip condition prediction model, which can be constructed by BP-NN. To improve the prediction accuracy of the prediction model, the initial weights and thresholds

can be optimized by some intelligent optimization algorithm. In [95], the focus was to provide an on-line energy management control strategy based on trip condition prediction to minimize fuel consumption for PHEVs. It predicts the vehicle speeds on-line by establishing the trip condition prediction model based on GA/PSOA-BPNN. Based on the multi-mode trip information prediction module (MTCPPM), the optimal strategy calculation module (OSCM) calculates the optimal control sequence by DP to obtain the optimal fuel consumption in real-time.

To solve the problem that DP cannot be applied in real time, it is possible to develop a machine learning strategy to learn the optimal power split by generalizing the knowledge through neural learning from multiple standard drive cycles. This includes driving environment prediction and optimal energy management by machine learning. A multilayered and multiclass neural network can be developed to predict the road environment including roadway types and traffic congestion levels over a driving trip. Meanwhile, another neural network could be developed to predict the driver's instantaneous reactions to the driving environment at any given moment. Based on the above predicted road information and driving trends, the optimal energy management strategy dictated by DP for the current conditions can be emulated.

On the intelligent optimization of energy management for HEVs, a machine learning framework that combines dynamic programming with machine learning to learn about roadway type and traffic-congestion level was developed in [96]. For each of the 11 drive cycles, two neural networks were trained to emulate the optimal engine speed generated by DP and the optimal battery power. The simulation results indicated that the proposed framework not only minimized the fuel consumption, but also maintained vehicle performance.

Compared with the CD-CS strategy, the engine operating points of the NN-DP strategy are mostly concentrated on the region with the lower fuel consumption, which can greatly decrease the total cost and can be regarded as an approximated global optimal energy management strategy. Due to the fact that the optimal control sequences are based on the DP results obtained from a set of standard driving cycles, which is an approximate nonlinear relationship based on statistical data to some extent, it cannot guarantee the optimality of energy management when the real driving conditions differ from the training samples.

### 3.3.2. Reinforcement Learning

As a new research hotspot in the field of machine learning and artificial intelligence (AI), reinforcement learning (RL) [97] is the system learning from environment to behavior mapping to maximize the value of the enhanced/reward signal. In recent years, the RL-based energy management strategy has been applied to achieve optimal fuel economy of HEVs in real-time. The schematic diagram of the RL-based strategy is shown in Figure 10.

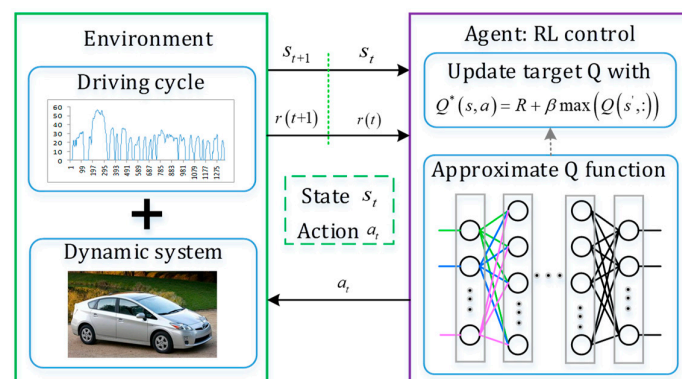


Figure 10. Depiction of the schematic diagram of the RL-based strategy.



Q learning, as a famous RL algorithm, can achieve satisfactory control performance that does not require an environmental model to strengthen learning. The environment can be regarded as a discrete Markov process with a finite state. The agent can select an action  $a$  within the finite action set at each step, and then the environment accepts the action to transfer the state and give the reward  $r_t$ . The task of the agent is to determine an optimal strategy to maximize the expected value of the total discount reward signal. The idea of Q learning is to directly optimize the Q function to find a strategy that maximizes the sum of expected discount rewards. The Q function is defined as a discount cumulative enhancement value when the action  $a$  is performed in the state  $s$  under the optimal action sequence, that is,

$$Q_{t+1}(s_t, a_t) = r_t + \beta \max_{a_t} \{Q(s_{t+1}, a_t)\} \quad (21)$$

where  $\beta$  is the discount factor.

In energy management based on Q learning, vehicle speed, battery SOC, or engine speed can be selected as the state variables, and the torque-split ratio between engine and motor, engine torque or motor torque can be selected as the control variables. The objective is to minimize the sum of fuel consumption at each moment; therefore, the reciprocal of fuel consumption at each time step can be defined as the immediate reward. The neural network can be used to approximate Q function, and outputs the corresponding action of Q learning. The process of Q learning is as follows:

- Initialize the Q network, including selecting the number of neurons in each layer, initializing network connection weights, etc.;
- Obtain the state  $s_t$  at the moment  $t$  and calculate the Q function of each action  $Q(s_t, a_i)$ ;
- Select an action  $a_i$  and obtain the next state  $s_{t+1}$  and the enhanced signal  $r_t$ ;
- Calculate  $Q'_t = r_t + \beta \max_{a_t} \{Q(s_{t+1}, a_t)\}$ , and then adjust the network weights to minimize the error  $\Delta Q_i$  until the error satisfies  $\Delta Q_i \leq \delta$ , where  $\Delta Q_i$  is defined as  $\Delta Q_j = Q' - Q_i$  ( $i = j$ ). The weight vectors of the Q network can be regulated based on the gradient-based adaptation algorithm.

In [97], a Q learning-based EMS was proposed for PHEVs to optimize the power-split control in real time, which would address the trade-off between real-time performance and optimal energy savings. The simulation of a real-world commute trip shows that about a 12% fuel savings can be achieved without taking charging opportunities into account.

However, the Q learning-based strategy would lead to a sharply increasing computational burden and bad convergence ability during the training process due to the “curse of dimensionality”, which makes the Q-learning strategy difficult in practical application for the energy management of HEVs. To solve the above problem, the deep Q learning (DQL) algorithm with deep neural network (DNN) was developed. Unlike Q learning, a deep neural network (DNN) is employed, and is well trained to approximate the action value function (Q function). Common deep learning networks generally include convolutional neural networks and recurrent neural networks. As for the convolutional neural network, it can handle high-dimensional data well, and weight sharing can reduce the complexity of the network. Additionally, it can avoid the gradient loss being too fast during the back propagation in the BP neural network. By maximizing the accumulated reward value obtained by the agent from the environment, DQL-based energy management for both minimizing the fuel consumption and keeping the battery SOC stable over a specific time horizon can be developed. A DQL-based energy management strategy for a power split hybrid electric bus (HEB) was proposed in [98]. An experience pool in the form of a quadruplet was defined to store the data needed by the neural network. Furthermore, the optimality and adaptability of the DQL-based strategy on the different driving cycles were verified by comparison with the DP-based strategy. Compared with the Q learning strategy with the same model, the strategy with DQL had a better performance in terms of training difficulty and the influence of different state variables on the action value function.

Based on the DQL method and considered as a breakthrough in the field of reinforcement learning, a deep reinforcement learning (DRL) framework for optimizing the fuel economy of HEVs was put

forward in [99], which is approximately optimal, model-free, and has no need to have any prior knowledge of the driving cycle. The DRL technique consists of an offline deep neural network and an online deep Q-learning network, and can handle the high dimensional state and action space.

To achieve practical application, the required power can be modeled as a stationary Markov model, and the power transition probability matrix can be calculated by the nearest neighborhood method using the RL-based offline strategy. Ref. [100] proposed a RL-based power management control strategy for PHEV, taking optimal fuel economy, battery life, real-time application, and adaptivity in different conditions into account. This strategy can have a good performance in limiting the maximum discharge current and optimizing the system efficiency. To seek the optimal control, different forgetting factors and Kullback-Leibler (KL) [97,99,101] divergence rates, which decide whether to update the power management strategy, were discussed.

Simultaneously, minimizing the tradeoff between fuel consumption and computational efficiency for HEVs can be achieved with respect to two different aspects. On the one hand, the dimension of action variable can be decreased by setting the engine to operate in a specified area, which contributes to decreasing the training difficulty of the deep neural network. On the other hand, a predictive energy management strategy via a synergy of velocity prediction and RL can be developed. In [102], the nearest neighbor and fuzzy encoding with two transformations (an M-dimensional possibility vector and the proportional possibility-to-probability) was employed to achieve velocity prediction. Furthermore, the transition probability of the vehicle velocity was estimated using the maximum likelihood estimator. The results of an HIL experiment verified that the predictive controller was implementable in real time and had lower fuel consumption than the rule-based strategy.

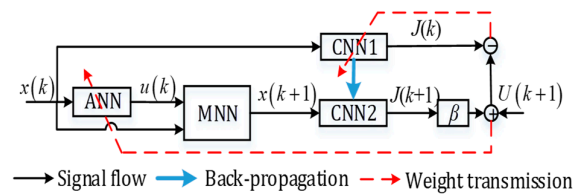
### 3.3.3. Adaptive Dynamic Programming (ADP)

With the increased dimensions of the state variables  $x$  and the control variables  $u$ , the computation and storage capacity of the traditional dynamic programming increase significantly, often leading to the phenomenon of “dimensional disasters”. In addition, the nonlinear optimal control needs to solve the nonlinear Hamilton-Jacobi-Bellman (HJB) equation; however, this involves solving nonlinear partial difference equations, which are difficult to work out. With the development of artificial neural networks characterized by their strong abilities for self-learning and adaptivity, adaptive dynamic programming (ADP) as an approximate solution to DP has been applied in effectively solving the dynamic programming problem of nonlinear discrete-time systems with strong coupling, strong nonlinearity and high complexity. This method has been put into practical application in transportation, power systems, industrial production, and other aspects. The core idea of adaptive dynamic programming is to approximate the performance index function and control strategy in the dynamic programming equation by using the function approximation structure to satisfy the optimality principle and obtain the optimal control sequence. The performance index function can be expressed as:

$$J(x_t) = \sum_{t=0}^{N-1} \beta^t U(x_t, u_t) = U(x_t, u_t) + \beta J(x_{t+1}) \quad (22)$$

where  $U$  is the utility function,  $J$  is the performance index function (cost function), and  $\beta$  is the discount factor.

ADP consists of three parts, each of which can be replaced by a neural network with the simplest neural network-BP, as shown in Figure 11. The model network (MNN) is used to approximate the dynamic system to estimate the state at the next moment, the action network (ANN) is used to approximate the optimal control law by mapping the relationship between state and control variables, and the critic network (CNN) is used to estimate the optimal performance index function.



**Figure 11.** The description of the structure of ADP.

According to the different input and output of the critic network (CNN), the adaptive dynamic programming (ADP) can be roughly subdivided into six categories. Among them, heuristic dynamic programming (HDP) [103] is the most basic and widely used structure among ADP algorithms, CNN output of which is the cost function. When the output of CNN is the partial derivative of the cost function, the structure turns into dual heuristic programming (DHP) [104]. To improve calculation accuracy and efficiency, researchers have developed two improved structures: action-dependent heuristic dynamic programming (ADHDP) and action-dependent dual heuristic programming (ADDHP) [105], whose inputs of CNN are the state and control variables. Based on the above two structures, globalized dual heuristic programming (GDHP) [106] and action-dependent globalized dual heuristic programming (ADGDHP) have been developed to further improve the computational accuracy.

For the optimal control of nonlinear discrete-time systems, there are several recent theoretical developments related to ADP. The multi-step heuristic dynamic programming [107] method has been developed to solve the optimal control problem, and it can speed up value iteration and avoid the requirement of initial admissible control policy in policy iteration at the same time. The iterative GDHP algorithm is implemented by constructing three neural networks to approximate the error system dynamics, the cost function with its derivative, and the control policy in each iteration, respectively.

The energy management systems (EMS) of HEVs belong to a class of nonlinear systems that is complicated, multivariable and difficult to control. How to effectively allocate the ratio between fuel and electricity from a global perspective and achieve real-time/online control through updating operating conditions based on data is the focus for researchers. Traditional optimization control is not fully able to meet these requirements; meanwhile, with the rapid development of the ADP algorithm and artificial neural networks, data-driven ADP algorithms for hybrid powertrain systems can be utilized to achieve the largest energy-saving potential, reduce the computational burden and achieve practical application in HEVs. In recent years, various organizations have carried out related studies in ADP-based energy management control strategies.

When ADP is applied in the energy management of HEVs, the optimal control sequence based on DP needs to be obtained in advance, and is then used as the training sample data to compare with DP results to verify the similarity of the ADP-based strategy to the DP-based strategy. The vehicle speed and battery SOC can be selected as state variables, while the engine speed, engine torque, motor speed and motor torque can be selected as control variables.

The focus of the ADP algorithm is to determine the structure of ADP, the form of the utility function, the error definition, and the weight updating method of each network, and the training procedure of ADP. Therefore, the main steps of ADP-based strategy to achieve energy management for HEVs are shown in Figure 12, and the main procedures of algorithm design are as follows:

- Step1: Normalize the state variables and optimal control variables, which are obtained by the DP method.
- Step2: Structure selection of ADP. If the dynamic system of vehicle is unknown, the model network is used to approximate the nonlinear system; therefore, HDP-based strategies would be developed. In contrast, if we have knowledge of the dynamic system of vehicle, there is no need to use MNN to estimate the next state.

Step3: Determine the utility function. Generally speaking, the utility function  $U(x, u, k)$  can be chosen as the quadratic form of the states and the control variables:

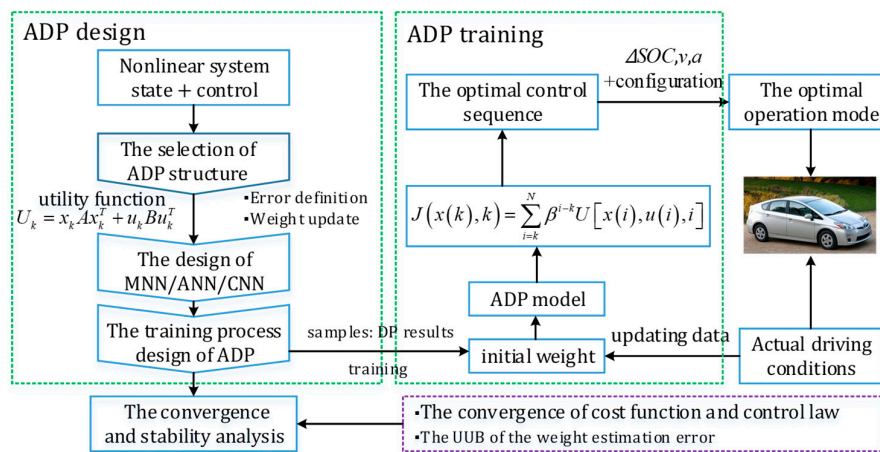
$$U(x, u, k) = x_k \cdot A \cdot x_k^T + u_k \cdot B \cdot u_k^T \quad (23)$$

where  $A$  and  $B$  are symmetric positive definite matrices with appropriate dimensions. If there is a particular constraint on SOC value, the square of the error between SOC and the standard value can be added to the utility function.

Step4: Design the ADP algorithm. The error definition and weight update are the focal points in it.

Step5: Design the algorithm procedure of ADP. The key procedure is to train ANN and CNN until the stopping rule is satisfied;

Step6: Analyze the convergence and stability of the algorithm, which includes theoretical proof and analysis of the simulation results. The uniform ultimate boundedness (UUB) of the associated weight estimation errors is proved using a (positive) Lyapunov approach [108].



**Figure 12.** Depiction of the main steps in applying ADP to achieve energy management control.

In the implementation of HDP, the ADP structure consists of an action neural network (ANN), a critic neural network (CNN) and a model neural network (MNN), which are established by a three-layer neural network (BP-NN) and all contain forward calculation and backward error back propagation to update weight. To identify the system dynamics based on the input–output data, the model neural network is trained in advance to obtain the next state, which takes the current state and control inputs as the inputs of MNN. The forward calculation of can be expressed as:

$$\begin{cases} inputM = [u_t^T, x_t^T]^T \\ \hat{x}_{t+1} = W_{m2}^T(t) \cdot \varphi_m(W_{m1}^T(t) \cdot inputM(t) + \xi_{m1}) + \xi_{m2} \end{cases} \quad (24)$$

where  $W_{m1}$  is the input-to-hidden weight vector of MNN,  $W_{m2}$  is hidden-to-output weight vector of MNN,  $\xi_{m1}$  is the threshold of the hidden layer of MNN,  $\xi_{m2}$  is the threshold of the output layer of MNN,  $\hat{x}_{t+1}$  is the next system state vector, and  $\varphi_m(\cdot)$  is the activation function, which can be selected as  $\varphi_m(\cdot) = \frac{1-e^{-ax}}{1+e^{-ax}}$ .

The identification error of MNN is defined as  $e_m = x_{t+1} - \hat{x}_{t+1}$ , where  $x_{t+1}$  is the expected value. The quadratic form of the identification error is denoted as the model network error:  $E_m = \frac{1}{2} e_m^T e_m$ .

The gradient-based descent approach is used for weight updating during the back-propagation process, which is:

$$W_m(t+1) = W_m(t) - \eta_m \cdot \frac{\partial E_m}{\partial e_m} \cdot \frac{\partial e_m}{\partial \hat{x}_{t+1}} \cdot \frac{\partial \hat{x}_{t+1}}{\partial W_m} \quad (25)$$

The critic network is used to approximate the iterative cost function, and there are two critic networks in each iteration, which have the same input-to-hidden vector  $V_{c1}$  and hidden-to-output weight vector  $V_{c2}$ . In the forward calculation of CNN, the first critic network (CNN1) takes current state  $x_t$  as the input and outputs the cost function of the current moment  $\hat{J}(x_t)$ , while the second critic network (CNN2) outputs the cost function of the next moment  $\hat{J}(x_{t+1})$  and takes the next state  $\hat{x}_{t+1}$  as the input, which is the output of MNN. The cost function is formulated as:

$$\hat{J}(x_t) = V_{c2}^T \varphi_c(V_{c1}^T x_t + \xi_{c1}) + \xi_{c2} \quad (26)$$

where  $\xi_{c1}$  is the threshold of the hidden layer of CNN1,  $\xi_{c2}$  is the threshold of the output layer, and  $\varphi_c(\cdot)$  is the activation function, which can be selected as  $\varphi_c(\cdot) = \frac{1-e^{-ax}}{1+e^{-ax}}$ .

The quadratic form of the error between the estimated cost function and the expected value is defined as the critic network error. The expected cost function in each iteration is denoted as:

$$J(x_t) = (x_t \cdot A \cdot x_t^T + u_t \cdot A \cdot u_t^T + \beta \hat{J}(x_{t+1})) \quad (27)$$

Similarly, the weight vectors are regulated based on the gradient-based adaptation algorithm [109].

$$W_c(t+1) = W_c(t) - \eta_c \cdot \frac{\partial E_c}{\partial e_c} \frac{\partial e_c}{\partial \hat{J}_t} \frac{\partial \hat{J}_t}{\partial W_c} \quad (28)$$

The action network with three-layer neural network is constructed to approach the iterative control law. In the forward calculation, ANN takes the current state as the input and outputs the estimated control law  $\hat{u}_t$ :

$$\hat{u}_t = W_{a2}^T(t) \cdot \varphi_a(W_{a1}^T(t) \cdot x_t + \xi_{a1}) + \xi_{a2} \quad (29)$$

where  $W_{a1}$  is the input-to-hidden weight vector of ANN,  $W_{a2}$  is the hidden-to-output weight vector of ANN,  $\xi_{a1}$  is the threshold of the hidden layer,  $\xi_{a2}$  is the threshold of the output layer of MNN, and  $\varphi_a(\cdot)$  is the activation function, which can be selected as  $\varphi_a(\cdot) = \frac{1-e^{-ax}}{1+e^{-ax}}$ .

Similarly, the action network error is defined as the quadratic form of the error between the estimated control and the expected value, which is obtained when the partial derivative of the cost function to the control variable is equal to 0, that is,

$$\frac{\partial J(x_t)}{\partial u_t} = 2B \cdot \hat{v}_t + \beta \frac{\partial \hat{J}(\hat{x}_{t+1})}{\partial \hat{x}_{t+1}} \frac{\partial \hat{x}_{t+1}}{\partial \hat{u}_t} = 0 \quad (30)$$

The gradient-based descent approach is used for weight updating during the back-propagation process, which is:

$$W_m(t+1) = W_m(t) - \eta_m \cdot \frac{\partial E_m}{\partial e_m} \frac{\partial e_m}{\partial \hat{x}_{t+1}} \frac{\partial \hat{x}_{t+1}}{\partial W_m} \quad (31)$$

When the cost function error of the adjacent iteration is extremely small, the iteration stops, and the optimal control sequence and optimal performance index function are obtained [110]. Based on the above analysis, the training process of HDP can be briefly summarized as in Algorithm 1. ( $I_{max}$  is the maximum number of iterations for ADP,  $N_{crit}$  is the maximum number of iterations for CNN training, and  $N_{act}$  is the maximum number of iterations for ANN training).



**Algorithm 1.** The training process of HDP.

- 
1. **For**  $k = 1$ : Time
  2. **While** ( $abs(J(k, i) - J(k, i - 1)) > \epsilon$ ) and ( $i < I_{max}$ ) **do**
    - 2.1.  $x_k \rightarrow ANN \rightarrow \hat{v}_k^{(i)}$
    - 2.2.  $x_k, \hat{v}_k^{(i)} \rightarrow \text{system/MNN} \rightarrow \hat{x}_{k+1}^{(i)}$
    - 2.3.  $\hat{x}_{k+1}^{(i)} \rightarrow CNN2 \rightarrow \hat{f}_{k+1}^{(i)}$
    - 2.4.  $\hat{U}_k^{(i)} = \hat{x}_k \cdot A \cdot \hat{x}_k^T + \hat{v}_k^{(i)} \cdot B \cdot \hat{v}_k^{(i)T}$
    - 2.5.  $i = i + 1$ ;
    - 2.6.  $x_k, \hat{v}_k^{(i)} \rightarrow CNN1 \rightarrow \hat{f}_k^{(i)}$
    - 2.7.  $J(k, i) = \beta \hat{f}_{k+1}^{(i-1)} + \hat{U}_k^{(i-1)}$
    - 2.8. **While** ( $E_c^{(i)}(k) > T_c$ ) and ( $cyc < N_{crit}$ ) **do**
      - 2.8.1. Update  $W_{c1}, W_{c2}$
      - 2.8.2.  $x_k \rightarrow CNN1 \rightarrow \hat{f}_k^{(i)}$
      - 2.8.3.  $e_c^{(i)}(k) = \hat{f}_k^{(i)} - J(k, i) \rightarrow E_c^{(i)}(k) = \frac{1}{2} e_c^2(k)$
    - 2.9. **Copy** CNN1 to CNN2
    - 2.10.  $v_{target}^{(i-1)}(k) = -\frac{1}{2} \cdot \beta \cdot B^{-1} \cdot \frac{\partial \hat{f}_{k+1}^{(i-1)}}{\partial \hat{v}_{k+1}^{(i-1)}}$
    - 2.11. **While** ( $E_a^{(i-1)}(k) > T_c$ ) and ( $clc < N_{act}$ ) **do**
      - 2.11.1. Update  $W_{a1}, W_{a2}$
      - 2.11.2.  $x_k \rightarrow ANN \rightarrow \hat{v}_k^{(i-1)}$
      - 2.11.3.  $x_k, \hat{v}_k^{(i-1)} \rightarrow \frac{\text{system}}{\text{MNN}} \rightarrow \hat{x}_{k+1}^{(i-1)}$
      - 2.11.4.  $e_a^{(i-1)}(k) = \hat{v}_k^{(i-1)} - v_{target}^{(i-1)}(k) \rightarrow E_a^{(i-1)}(k) = \frac{1}{2} e_a^2(k)$
  3. **Output:**  $x_k^*, u_k^*, J_k^*, x_{k+1}^*$
  4. **End**
- 

To achieve more effective approximation of nonlinear systems, the multi-resolution wavelet neural network (MRWNN) based on multi-resolution theory, Meyer scaling, and wavelet functions can be introduced to construct the critic network, whose activation function is orthogonal in order to achieve faster convergence and higher approximate accuracy. A conventional wavelet neural network (CWNN) using a wavelet function as the activation function in the hidden layer does not need clustering preprocessing on the input data, which can be utilized to construct the action network. Similar to BP-NN or RBF-NN, the adaptation of weights and parameters of both networks can be achieved by error back propagation and a gradient descent algorithm. A neuro-dynamic programming (NDP) method using MRWNN as the critic network and the CWNN as the action network was detailed in [111], and it was able to optimize fuel economy for online application and optimize SOC without previewing future traffic information. In addition, a correction model for the output of the action network was used to discretize the continuous gear ratio to the real vehicle gear ratio of the vehicle.

For multi-source hybrid electric vehicles (HEVs), the novel concepts of adaptive dynamic programming (ADP) and the progressive optimal search (POS) were proposed in [112]. From the perspectives of trip cost minimization, charge sustenance, and real-time applicability, the advantages of each method were compared with previously developed methods of rule-based strategy and adaptive rule-based strategy. In [113], real-time optimal control of multisource HEVs by ADP was proposed, which was based on drive state recognition according to physics-based parameters. This included the

implementation of ADP to obtain optimal EMS and the optimization of vehicle operating conditions for each state, which is optimized offline using the NSGA-II optimization tool. As a consequence, the total cost function decreases in terms of fuel consumption and on-board charge sustaining.

Since the ADP algorithm is still being developed, the ADP-based energy management strategy is currently in a stage of development and practical application. Some organizations have realized its experimental application based on a hardware-in-the-loop (HIL) test-rig, and the next step is to realize application in real vehicles. Therefore, the method of the ADP-based strategy for energy management in HEVs has been a research hot spot in universities and the automotive industry. With the rapid development of big data technology, artificial intelligence algorithms and the parallel computing technology of supercomputers, the data-driven energy management control strategy is able to be realized in real-vehicle applications to achieve the lowest fuel consumption in a global scope.

## 4. Algorithm Analysis

### 4.1. Complexity Analysis of Algorithms

The complexity of the various algorithms refers to the time resources and memory resources required by the algorithm once it has been written into an executable program. Therefore, the algorithm's complexity is analyzed based on time complexity and space complexity. The time complexity of an algorithm refers to the computational burden required to execute the algorithm, which is measured by the number of executions of the statement in the algorithm, namely,  $T(n)$ . The common order of magnitude is:  $O(1) < O(\log n) < O(n) < O(n \log n) < O(n^2) < O(n^3) < O(2^n) < O(n!) < O(n^n)$ .

There are several rules in calculating the time complexity of an algorithm:

1. For some simple input and output statements or assignment statements, the order of magnitude is considered to be  $O(1)$ .
2. For the sequential structure, the time required to execute a series of statements in turn can use the "summation rule".
3. For the selection structure, the required time mainly considers the time to execute the "then" or "else" statement.
4. For the loop structure, the required time is mainly reflected in the time required to execute the loop body and check the loop condition in multiple iterations, which usually uses the "multiplication rule".
5. For a complex algorithm, it can be divided into several easy-to-estimate parts, and then use the summation rule and multiplication law to calculate the time complexity of the algorithm.

The time complexity of common algorithms is detailed in Table 1.

**Table 1.** Time complexity of common algorithms.

Orders ( $n^1$ )	Affordable Scale	Common Algorithms
$O(1)$	✓	directly output results
$O(\log n)$	✓	binary search
$O(n^1)$	Millions	greedy, scanning, traversing
$O(n \log n)$	Hundreds of thousands	division algorithm (dichotomy)
$O(n^2)$	thousands	enumeration, DP
$O(n^3)$	Less than two hundred	dynamic programming
$O(2^n)$	24	searching
$O(n!)$	10	Full array
$O(n^n)$	8	violence law to crack passwords

<sup>1</sup>  $n$  represents the scale of the problem; ✓ represents any situation.

The space complexity is a measure of the amount of storage space temporarily occupied by an algorithm during its operation. The storage space occupied by the algorithm includes the storage

space occupied by the algorithm itself, the storage space occupied by the input and output data of the algorithm, and the storage space temporarily occupied by the algorithm during the running process. The storage space occupied by the algorithm itself is proportional to the length of the algorithm statement. The storage space occupied by the input and output data of the algorithm is determined by the problem to be solved, which is transmitted through the parameter table, and it does not change with the algorithm. Generally, the number of temporary working units that the algorithm needs to occupy is related to the magnitude of the problem, which increases as  $n$  increases.

Based on the above analysis, we can calculate the complexity of the above-mentioned algorithms, as shown in Table 2:

**Table 2.** The complexity of the above-mentioned algorithms.

Algorithm	Time Complexity ( $n^1$ )	Space Complexity
RB	$O(1) = O(1) + O(1)$	$m \cdot O(n)$
ECMS	$O(n)$	$m \cdot O(n^2)$
MPC	$O(n^3) \sim DP$ or $O(n^2) \sim GA/SA$	$O(n^2)$
PMP	$O(n^2)$	$m \cdot O(n^3)$
DDP	$O(n^3) = O(n(n+1)(n-1)/6)$	$m \cdot O(n^3)$
SDP	$O(n^3) = O(n^3) + O(n^2)$	$O(n^3)$
NN-DP	$O(n^2) = O(m \cdot n^2)$	$m \cdot O(n^2) \sim \text{train}$
RL	$O(n^2)$	$m \cdot O(n^2) \sim \text{train}$
ADP	$O(n^2) = O(m \cdot n^2)$	$O(n^2)$

<sup>1</sup>  $n$  represents the scale of the problem, which can be regarded as the dimension of input vector;  $m$  represents the number of samples.

#### 4.2. Comprehensive Performance Analysis of the Strategies

The performance of energy management control strategies can be analyzed with respect to the following aspects: fuel economy, real-time performance, computational time, computational burden, and degree of realization at present.

According to the space complexity of the algorithm, the computational burden of various strategies can be calculated, while computational time corresponds to the time complexity of the algorithms. Simultaneously, the complexity of an algorithm can be analyzed by combining the time complexity with the space complexity. According to Table 2, the computational time of the various strategies sorted from low to high is as follows: RB < ECMS < ADP/PMP < NN-DP/RL < MPC/SDP < DDP. Similarly, the computational burden of the various strategies, sorted from lowest to highest, is as follows: RB < ADP/MPC < RL/ECMS/NN-DP < SDP < PMP/DDP.

For a given driving cycle, the fuel economy of the various strategies can be compared according to the simulation results under the same vehicle configuration and the same driving cycle. The DP-based strategy is able to obtain a theoretical global optimal fuel economy for HEVs, which provides a benchmark for assessing the optimality of other energy management strategies, despite the off-line optimization. If the performance of the minimum fuel consumption of the various strategies is divided into five levels from high to low, the DP-based strategy can be regarded as the first level, while the rule-based strategy can be regarded as the last level, due to the fact that it cannot obtain optimal fuel economy. The PMP-based strategy can obtain the global optimum solution by using the minimum principle when the obtained local optimal trajectory is a unique trajectory. Similarly, the ADP-based strategy (RL as a special structure of ADP) can obtain the approximation results for DP by approximating the performance index function and control strategy based on the function approximation structure to satisfy the optimality principle. The accuracy of the velocity prediction will influence the optimal results of the MPC-based strategy and the SDP-based strategy, and robustness and accuracy cannot be guaranteed when the actual driving cycle is quite different from the training driving cycle. The results of ECMS are very sensitive to the equivalent factor, while the results of the

NN-DP strategy are influenced by the comprehensiveness of the training samples. According to the approximate degree between the results of each control strategy and the DP results, the fuel economy of the various strategies, sorted from lowest to highest, is as follows:  $RB < ECMS/NN-DP < SDP/MPC < ADP/RL < PMP/DDP$ .

Similarly, if the real-time performance of the EMS in HEVs is divided into five levels from high to low, the rule-based strategy can be regarded as the first level, due to its convenience in adjusting parameters, its simplicity, and its practicality, while the DP-based strategy can be regarded as the last level, due to its being time consuming and requiring a tremendous amount of memory. ECMS and the MPC-based strategy, as instantaneous strategies, can be considered as the second level. Meanwhile, the ADP-based strategy (containing the RL-based strategy and the NDP-based strategy) can be implemented based on updating data, and can therefore be regarded as the second level. The SDP-based strategy can be regarded as the third level, while the PMP-based strategy can be considered as the fourth level due to the difficulty it has in solving the Hamiltonian function.

Based on the above analysis, we can obtain the Figure 13. The axis represents the level of performance. The larger the value, the better the performance.

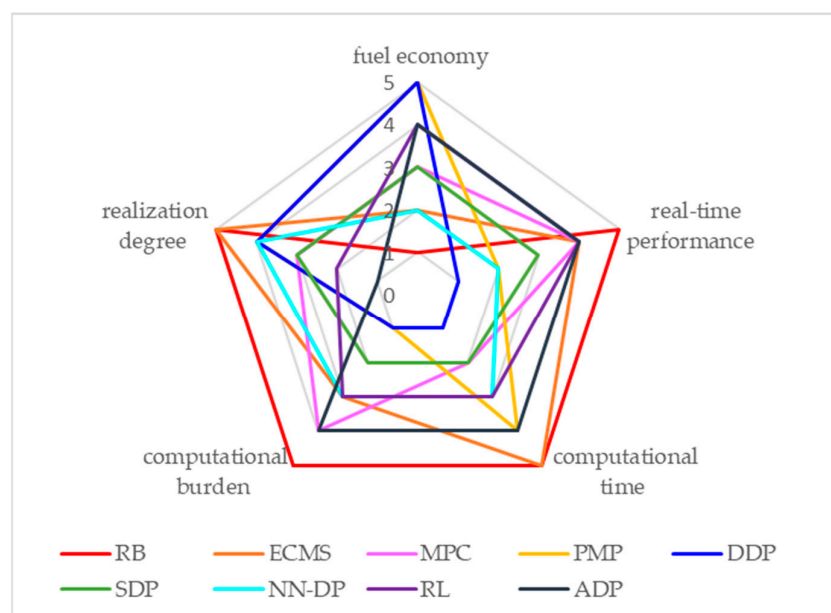


Figure 13. Depiction of the performance of each control strategy.

## 5. Conclusions and Suggestions

Control strategies are keys to the energy management system for HEVs, and can effectively distribute the power between the engine and the motors to improve fuel economy as much as possible. This review introduces the main research ideas, strengths and weaknesses of all of the existing control strategies in the available literature, categorizing them as rule-based strategies and optimization-based strategies. Rule-based control strategies, as the most basic strategies, have been widely used due to their simplicity and practical application. However, they tend to fall into local optimal solutions. Meanwhile, optimization-based control strategies are able to obtain global optimal or near-optimal solutions, and include instantaneous optimization, global information-driven optimization, and data-driven strategies.

Regarding each optimization strategy, the research focus and future research hotspots are as follows. The focus of ECMS is to determine the optimal equivalent factor, and the adaptive-ECMS with velocity prediction is the new research direction to adaptively adjust the equivalence factor in real time. As for the PMP-based strategy, the keys include the solution to the boundary value problem and the determination of the covariant variable. To implement online control, developing an approximate PMP

is necessary for enhancing computational efficiency and reducing computational burden. Deterministic dynamic programming (DDP) in EMS of HEVs provides a benchmark for assessing the optimality of other energy management strategies despite the off-line optimization. The MPC-based strategy includes velocity prediction within a prediction horizon, and the optimization of the short-term horizon based on PMP, DP, or intelligent algorithm. Based on the existing standard driving cycles or historical driving data, the SDP-based strategy establishes a statistical model of driver demand, and then uses dynamic programming to solve the energy management problem. For the above two control strategies, the key points are to improve the accuracy of speed prediction and to guarantee the robustness when the actual driving cycle is quite different from the training driving cycle.

Based on the DP results, the neural network is utilized to construct a prediction model of the road environment and driving conditions in the NN-DP strategy; however, it cannot guarantee robustness. Reinforcement learning, as a new research hotspot, uses neural networks to approximate the Q function. To solve the issue of the “curse of dimensionality”, the deep reinforcement learning (DRL/DQL) algorithm is a breakthrough, and is the research hotspot for achieving practical application and handling high-dimensional data well. The ADP-based strategy can be implemented in real time based on updating the driving conditions, which is the next step to overcome in EMS for HEVs. With the rapid development of artificial intelligence algorithms and large data processing technology, an optimal control strategy based on data-driven strategies will be the most promising research direction due to high computing efficiency and global optimality, especially the DNN-based and ADP-based strategies. It is worth deeply considering how to effectively combine intelligent network technology with energy management systems and make full use of the collected geographic information based on the technology of ITSs, GISs or GPSs in order to globally optimize energy consumption.

**Author Contributions:** N.X. and Y.K. conceived the topic and frame. Y.K., N.X., L.C., H.J., Z.Y., Z.X. (Zhe Xu) and Z.X. (Zhuoqi Xu) wrote the full manuscript; all the authors were involved in conducting the literature survey and revising the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under Grant 51805201, the Jilin Province Science and Technology Development Fund under Grant 20180101062JC, and the Energy Administration of Jilin Province 2017.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Lukic, S.M.; Cao, J.; Bansal, R.C.; Rodriguez, F.; Emadi, A. Energy storage systems for automotive applications. *IEEE Trans. Ind. Electron.* **2008**, *55*, 2258–2267. [[CrossRef](#)]
2. Baisden, A.C.; Emadi, A. An ADVISOR based model of a battery and an ultra-capacitor energy source for hybrid electric vehicles. *IEEE Trans. Veh. Technol.* **2004**, *53*, 199–205. [[CrossRef](#)]
3. Chu, L.; Jia, Y.-F.; Chen, D.-S.; Xu, N. Research on control strategies of an open-end winding permanent magnet synchronous driving motor (OW-PMSM)-Equipped Dual Inverter with a switchable winding mode for electric vehicles. *Energies* **2017**, *10*, 616. [[CrossRef](#)]
4. Wirasingha, S.G.; Emadi, A. Classification and review of control strategies for plug-in hybrid electric vehicles. *IEEE Trans. Veh. Technol.* **2011**, *60*, 111–122. [[CrossRef](#)]
5. Gao, J.; Sun, F.; He, H.; Zhu, G.G.; Strangas, E.G. A comparative study of supervisory control strategies for a series hybrid electric vehicle. In Proceedings of the Asia-Pacific Power and Energy Engineering Conference (APPEEC 2009), Wuhan, China, 27–31 March 2009; pp. 1–7.
6. Malikopoulos, A.A. Supervisory power management control algorithms for hybrid electric vehicles: A survey. *IEEE Trans. Intell. Transp. Syst.* **2014**, *15*, 1869–1885. [[CrossRef](#)]
7. Torres, J.L.; Gonzalez, R.; Gimenez, A.; Lopez, J. Energy management strategy for plug-in hybrid electric vehicles. A comparative study. *Appl. Energy* **2014**, *113*, 816–824. [[CrossRef](#)]
8. Solouk, A.; Shakiba-Herfeh, M.; Arora, J.; Shahbakhti, M. Fuel consumption assessment of an electrified powertrain with a multi-mode high-efficiency engine in various levels of hybridization. *Energy Convers. Manag.* **2018**, *155*, 100–115. [[CrossRef](#)]



9. Moulik, B.; Söffker, D. Optimal Rule-Based Power Management for Online, Real-Time Applications in HEVs with Multiple Sources and Objectives: A Review. *Energies* **2015**, *8*, 9049–9063. [\[CrossRef\]](#)
10. Martinez, C.M.; Hu, X.; Cao, D.; Velenis, E.; Gao, B.; Wellers, M. Energy Management in Plug-in Hybrid Electric Vehicles: Recent Progress and a Connected Vehicles Perspective. *IEEE Trans. Veh. Technol.* **2017**, *66*, 4534–4549. [\[CrossRef\]](#)
11. Zhang, Y.; Chu, L.; Fu, Z.; Xu, N. Optimal energy management strategy for parallel plug-in hybrid electric vehicle based on driving behavior analysis and real time traffic information prediction. *Mechatronics* **2017**, *46*, 177–192. [\[CrossRef\]](#)
12. Banvait, H.; Anwar, S.; Chen, Y. A rule-based energy management strategy for plug-in hybrid electric vehicle (PHEV). In Proceedings of the 2009 American Control Conference, St. Louis, MO, USA, 10–12 June 2009; pp. 3938–3943.
13. Chen, B.C.; Wu, Y.Y.; Tsai, H.C. Design and analysis of power management strategy for range extended electric vehicle using dynamic programming. *Appl. Energy* **2014**, *113*, 1764–1774. [\[CrossRef\]](#)
14. Ansarey, M.; Shariat Panahi, M.; Ziarati, H.; Mahjoob, M. Optimal energy management in a dual-storage fuel-cell hybrid vehicle using multidimensional dynamic programming. *J. Power Sources* **2014**, *250*, 359–371. [\[CrossRef\]](#)
15. Yang, Y.; Hu, X.; Pei, H. Comparison of power-split and parallel hybrid powertrain architectures with a single electric machine: Dynamic programming approach. *Appl. Energy* **2016**, *168*, 683–690. [\[CrossRef\]](#)
16. Kim, N.; Cha, S.; Peng, H. Optimal control of hybrid electric vehicles based on Pontryagin’s minimum principle. *IEEE Trans. Control Syst. Technol.* **2011**, *19*, 1279–1287.
17. Zhang, S.; Xiong, R.; Zhang, C. Pontryagin’s minimum principle-based power management of a dual-motor-driven electric bus. *Appl. Energy* **2015**, *159*, 370–380. [\[CrossRef\]](#)
18. Teng, L.; Yuan, Z.; Liu, D.X.; Sun, F. Real-time Control for a Parallel Hybrid Electric Vehicle Based on Pontryagin’s Minimum Principle. In Proceedings of the 2014 IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific), Beijing, China, 31 August–3 September 2014.
19. Xia, C.Y.; Du, Z.M.; Zhang, C. A Single-Degree-of-Freedom Energy Optimization Strategy for Power-Split Hybrid Electric Vehicles. *Energies* **2017**, *10*, 896. [\[CrossRef\]](#)
20. Wang, X.; He, H.; Sun, F. Application Study on the Dynamic Programming Algorithm for Energy Management of Plug-in Hybrid Electric Vehicles. *Energies* **2015**, *8*, 3225–3244. [\[CrossRef\]](#)
21. Song, Z.; Hofmann, H.; Li, J.; Han, X.; Ouyang, M. Optimization for a hybrid energy storage system in electric vehicles using dynamic programming approach. *Appl. Energy* **2015**, *139*, 151–162. [\[CrossRef\]](#)
22. Li, G.; Zhang, J.; He, H. Battery SOC constraint comparison for predictive energy management of plug-in hybrid electric bus. *Appl. Energy* **2017**, *194*, 578–587. [\[CrossRef\]](#)
23. Martel, F.; Kelouwani, S.; Dube, Y.; Agbossou, K. Optimal economy-based battery degradation management dynamics for fuel-cell plug-in hybrid electric vehicles. *J. Power Sources* **2015**, *274*, 367–381. [\[CrossRef\]](#)
24. Xie, S.; He, H.; Peng, J. An energy management strategy based on stochastic model predictive control for plug-in hybrid electric buses. *Appl. Energy* **2017**, *196*, 279–288. [\[CrossRef\]](#)
25. Li, L.; You, S.; Yang, C. Driving-behavior-aware stochastic model predictive control for plug-in hybrid electric buses. *Appl. Energy* **2016**, *162*, 868–879. [\[CrossRef\]](#)
26. Li, L.; Yan, B.; Song, J. Two-step optimal energy management strategy for single-shaft series-parallel powertrain. *Mechatronics* **2016**, *36*, 147–158. [\[CrossRef\]](#)
27. Wang, J.; Wang, Q.-N.; Zeng, X.-H.; Wang, P.-Y. Research on the optimal power management strategy for a hybrid electric bus. *Automob. Eng.* **2015**, *229*, 1529–1542. [\[CrossRef\]](#)
28. Lewis, F.L.; Vrabie, D. Reinforcement learning and adaptive dynamic programming for feedback control. *IEEE Circuits Syst. Mag.* **2009**, *9*, 32–50. [\[CrossRef\]](#)
29. Lewis, F.L.; Vamvoudakis, K.G. Reinforcement learning for partially observable dynamic processes: Adaptive dynamic programming using measured output data. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2011**, *41*, 14–25. [\[CrossRef\]](#)
30. Werbos, P.J. Approximate dynamic programming for real-time control and neural modeling. In *Handbook of Intelligent Control*; Van Nostrand: New York, NY, USA, 1992.
31. Bertsekas, D.P. *Dynamic Programming and Optimal Control*, 4th ed.; Athena Scientific: Belmont, MA, USA, 2017; Volume I, ISBN 1-886529-43-4.

32. Moura, S.J.; Callaway, D.S. Tradeoffs between battery energy capacity and stochastic optimal power management in plug-in hybrid electric vehicles. *J. Power Sources* **2010**, *195*, 2979–2988. [\[CrossRef\]](#)
33. Hu, X.; Murgovski, N.; Johannesson, L.; Bo, E. Energy efficiency analysis of a series plug-in hybrid electric bus with different energy management strategies and battery sizes. *Appl. Energy* **2013**, *111*, 1001–1009. [\[CrossRef\]](#)
34. Zhang, B.; Mi, C.C.; Zhang, M. Charge-Depleting Control Strategies and Fuel Optimization of Blended-Mode Plug-In Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2011**, *60*, 1516–1525. [\[CrossRef\]](#)
35. Sharer, P.B.; Rousseau, A.; Karbowski, D.; Pagerit, S. *Plug-in Hybrid Electric Vehicle Control Strategy: Comparison between EV and Charge-Depleting Options*; SAE Technical Paper Series; SAE International: Warrendale, PA, USA, 2008; Volume 32, pp. 1996–2014.
36. Zhang, M.; Yang, Y.; Mi, C.C. Analytical Approach for the Power Management of Blended-Mode Plug-In Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2012**, *61*, 1554–1566. [\[CrossRef\]](#)
37. Li, L.; Zhang, Y.H.; Chao, Y.G.; Jiao, X.H.; Zhang, L.P.; Song, J. Hybrid Genetic Algorithm-based Optimization of Powertrain and Control Parameters of Plug-in Hybrid Electric Bus. *J. Frankl. Inst.* **2014**, *352*, 776–801. [\[CrossRef\]](#)
38. Hao, J.; Yu, Z.; Zhao, Z.; Shen, P.; Zhan, X. Optimization of Key Parameters of Energy Management Strategy for Hybrid Electric Vehicle Using DIRECT Algorithm. *Energies* **2016**, *9*, 997. [\[CrossRef\]](#)
39. Liu, Y.; Gao, J.; Qin, D.; Zhang, Y.; Lei, Z. Rule-corrected energy management strategy for hybrid electric vehicles based on operation-mode prediction. *J. Clean. Prod.* **2018**, *188*, 796–806. [\[CrossRef\]](#)
40. Navale, V.; Havens, T.C. Fuzzy Logic Controller for Energy Management of Power Split Hybrid Electrical Vehicle Transmission. In Proceedings of the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing, China, 6–11 July 2014.
41. Sabri, M.F.M.; Danapalasingam, K.A.; Rahmat, M.F. Improved Fuel Economy of Through-the-Road Hybrid Electric Vehicle with Fuzzy Logic-Based Energy Management Strategy. *Int. J. Fuzzy Syst.* **2018**, *20*, 2677–2692. [\[CrossRef\]](#)
42. Denis, N.; Dubois, M.R.; Desrochers, A. Fuzzy-based blended control for the energy management of a parallel plug-in hybrid electric vehicle. *IET Intell. Transp. Syst.* **2015**, *9*, 30–37. [\[CrossRef\]](#)
43. Wei, Z.; Xu, Z.; Halim, D. Study of HEV power management control strategy Based on driving pattern recognition. *Energy Procedia* **2016**, *88*, 847–853. [\[CrossRef\]](#)
44. Wiczeorek, M.; Lewandowski, M. A mathematical representation of an energy management strategy for hybrid energy storage system in electric vehicle and real time optimization using a genetic algorithm. *Appl. Energy* **2017**, *192*, 222–233. [\[CrossRef\]](#)
45. Zhu, B.; Liu, Z.; Zhao, J.; Chen, Y.; Deng, W. Driver Behavior Characteristics Identification Strategies Based on Bionic Intelligent Algorithms. *IEEE Trans. Hum.-Mach. Syst.* **2018**, *48*, 572–581. [\[CrossRef\]](#)
46. Syed, F.U.; Filev, D.; Ying, H. Fuzzy Rule-Based Driver Advisory System for Fuel Economy Improvement in a Hybrid Electric Vehicle. In Proceedings of the NAFIPS 2007—2007 Annual Meeting of the North American Fuzzy Information Processing Society, San Diego, CA, USA, 24–27 June 2007.
47. McQueen, B.; McQueen, J. *Intelligent Transportation Systems Architectures*; Artech House: Norwood, MA, USA, 2003.
48. Peng, Z.-R.; Tsou, M.-H. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Network*; Wiley: Hoboken, NJ, USA, 2003.
49. Gong, Q.; Li, Y.; Peng, Z. Trip-Based Optimal Power Management of Plug-in Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2008**, *57*, 3393–3401. [\[CrossRef\]](#)
50. Rezaei, A.; Burl, J.B.; Zhou, B.; Rezaei, M. A New Real-Time Optimal Energy Management Strategy for Parallel Hybrid Electric Vehicles. *IEEE Trans. Control Syst. Technol.* **2017**, *27*, 830–837. [\[CrossRef\]](#)
51. Rezaei, A.; Burl, J.B.; Zhou, B. Estimation of the ECMS Equivalent Factor Bounds for Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2018**, *26*, 2198–2205. [\[CrossRef\]](#)
52. Rezaei, A.; Burl, J.B.; Solouk, A.; Zhou, B.; Rezaei, M.; Shahbakhti, M. Catch energy saving opportunity (CESO), an instantaneous optimal energy management strategy for series hybrid electric vehicles. *Appl. Energy* **2017**, *208*, 655–665. [\[CrossRef\]](#)
53. Yang, S.; Wang, W.; Zhang, F.; Hu, Y.; Xi, J. Driving-Style-Oriented Adaptive Equivalent Consumption Minimization Strategies for HEVs. *IEEE Trans. Veh. Technol.* **2018**, *67*, 9249–9261. [\[CrossRef\]](#)

54. Sun, C.; Sun, F.; He, H. Investigating adaptive-ECMS with velocity forecast ability for hybrid electric vehicles. *Appl. Energy* **2017**, *185*, 1644–1653. [\[CrossRef\]](#)
55. Han, J.; Kum, D.; Park, Y. Synthesis of Predictive Equivalent Consumption Minimization Strategy for Hybrid Electric Vehicles Based on Closed-Form Solution of Optimal Equivalence Factor. *IEEE Trans. Veh. Technol.* **2017**, *66*, 5604–5616. [\[CrossRef\]](#)
56. Guo, L.; Gao, B.; Gao, Y.; Chen, H. Optimal Energy Management for HEVs in Eco-Driving Applications Using Bi-Level MPC. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 2153–2162. [\[CrossRef\]](#)
57. Borhan, H.; Vahidi, A.; Phillips, A.M.; Kuang, M.L.; Kolmanovsky, I.V.; Di Cairano, S. MPC-Based Energy Management of a Power-Split Hybrid Electric Vehicle. *IEEE Trans. Control Syst. Technol.* **2012**, *20*, 593–603. [\[CrossRef\]](#)
58. Zhang, S.; Luo, Y.; Wang, J.; Wang, X.; Li, K. Predictive Energy Management Strategy for Fully Electric Vehicles Based on Preceding Vehicle Movement. *IEEE Trans. Intell. Transp. Syst.* **2017**, *18*, 3049–3060. [\[CrossRef\]](#)
59. Lim, H.; Su, W.; Mi, C.C. Distance-Based Ecological Driving Scheme Using a Two-Stage Hierarchy for Long-Term Optimization and Short-Term Adaptation. *IEEE Trans. Veh. Technol.* **2017**, *66*, 1940–1949. [\[CrossRef\]](#)
60. Lin, X.; Wang, Y.; Bogdan, P.; Chang, N.; Pedram, M. Optimizing Fuel Economy of Hybrid Electric Vehicles Using a Markov Decision Process Model. In Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV), Seoul, Korea, 28 June–1 July 2015.
61. Xiang, C.; Ding, F.; Wang, W.; He, W. Energy management of a dual-mode power-split hybrid electric vehicle based on velocity prediction and nonlinear model predictive control. *Appl. Energy* **2017**, *189*, 640–653. [\[CrossRef\]](#)
62. Ramadan, H.S.; Becherif, M.; Claude, F. Energy Management Improvement of Hybrid Electric Vehicles via Combined GPS/Rule-Based Methodology. *IEEE Trans. Autom. Sci. Eng.* **2017**, *14*, 586–597. [\[CrossRef\]](#)
63. Yu, K.; Xu, X.; Liang, Q.; Hu, Z.; Yang, J.; Guo, Y.; Zhang, H. Model predictive control for connected hybrid electric vehicles. *Math. Probl. Eng.* **2015**, *2015*, 318025. [\[CrossRef\]](#)
64. He, H.; Zhang, J.; Li, G. Model predictive control for energy management of a plug-in hybrid electric bus. *Energy Procedia* **2016**, *88*, 901–907. [\[CrossRef\]](#)
65. Chen, Z.; Xiong, R.; Wang, C. An on-line predictive energy management strategy for plug-in hybrid electric vehicles to counter the uncertain prediction of the driving cycle. *Appl. Energy* **2017**, *185*, 1663–1672. [\[CrossRef\]](#)
66. Chen, Z.; Xia, B.; You, C. A novel energy management method for series plug-in hybrid electric vehicles. *Appl. Energy* **2015**, *145*, 172–179. [\[CrossRef\]](#)
67. Huang, Y.; Wang, H.; Khajepour, A.; He, H.; Ji, J. Model predictive control power management strategies for HEVs: A review. *J. Power Sources* **2017**, *341*, 91–106. [\[CrossRef\]](#)
68. Xie, S.; Peng, J.; He, H. Plug-In Hybrid Electric Bus Energy Management Based on Stochastic Model Predictive Control. *Energy Procedia* **2017**, *105*, 2672–2677. [\[CrossRef\]](#)
69. Di Cairano, S.; Bernardini, D.; Bemporad, A.; Kolmanovsky, I.V. Stochastic MPC with Learning for Driver-Predictive Vehicle Control and its Application to HEV Energy Management. *IEEE Trans. Control Syst. Technol.* **2014**, *22*, 1018–1031. [\[CrossRef\]](#)
70. Zeng, X.; Wang, J. A Parallel Hybrid Electric Vehicle Energy Management Strategy Using Stochastic Model Predictive Control with Road Grade Preview. *IEEE Trans. Control Syst. Technol.* **2015**, *23*, 2416–2423. [\[CrossRef\]](#)
71. Lee, T.-K.; Adornato, B.; Filipi, Z.S. Synthesis of real-world driving cycles and their use for estimating PHEV energy consumption and charging opportunities: Case study for midwest/U.S. *IEEE Trans. Veh. Technol.* **2011**, *60*, 4153–4163. [\[CrossRef\]](#)
72. Xie, S.; Hu, X.; Xin, Z.; Brighton, J. Pontryagin's Minimum Principle based model predictive control of energy management for a plug-in hybrid electric bus. *Appl. Energy* **2019**, *236*, 893–905. [\[CrossRef\]](#)
73. Du, J.; Zhang, X.; Wang, T.; Song, Z.; Yang, X.; Wang, H.; Ouyang, M.; Wu, X. Battery degradation minimization oriented energy management strategy for plug-in hybrid electric bus with multi-energy storage system. *Energy* **2018**, *165*, 153–163. [\[CrossRef\]](#)
74. Wang, Y.; Wu, Z.; Chen, Y. Research on energy optimization control strategy of the hybrid electric vehicle based on Pontryagin's minimum principle. *Comput. Electr. Eng.* **2018**, *72*, 203–213. [\[CrossRef\]](#)

75. Chen, Z.; Mi, C.C.; Xia, B.; You, C. Energy management of power-split plug-in hybrid electric vehicles based on simulated annealing and Pontryagin's minimum principle. *J. Power Sources* **2014**, *272*, 160–168. [[CrossRef](#)]
76. Zhang, J.; Zheng, C.; Cha, S.W.; Duan, S. Co-State Variable Determination in Pontryagin's Minimum Principle for Energy Management of Hybrid Vehicles. *Int. J. Precis. Eng. Manuf.* **2016**, *17*, 1215–1222. [[CrossRef](#)]
77. Hou, C.; Ouyang, M.; Xu, L.; Wang, H. Approximate Pontryagin's minimum principle applied to the energy management of plug-in hybrid electric vehicles. *Appl. Energy* **2014**, *115*, 174–189. [[CrossRef](#)]
78. Kum, D.; Peng, H.; Bucknor, N.K. Supervisory Control of Parallel Hybrid Electric Vehicles for Fuel and Emission Reduction. *J. Dyn. Sys. Meas. Control* **2011**, *133*, 061010. [[CrossRef](#)]
79. Hovgard, M.; Jonsson, O.; Murgovski, N.; Sanfridson, M.; Fredriksson, J. Cooperative energy management of electrified vehicles on hilly roads. *Control Eng. Pract.* **2018**, *73*, 66–78. [[CrossRef](#)]
80. Bellman, R. *Dynamic Programming*; Princeton University Press: Princeton, NJ, USA, 1957.
81. Peng, J.; He, H.; Xiong, R. Rule based energy management strategy for a series-parallel plug-in hybrid electric bus optimized by dynamic programming. *Appl. Energy* **2017**, *185*, 1633–1643. [[CrossRef](#)]
82. Hou, C.; Xu, L.; Wang, H. Energy management of plug-in hybrid electric vehicles with unknown trip length. *J. Frankl. Inst.* **2015**, *352*, 500–518. [[CrossRef](#)]
83. Vinot, E. Time reduction of the Dynamic Programming computation in the case of hybrid vehicle. *Int. J. Appl. Electromagn. Mech.* **2017**, *53*, 213–227. [[CrossRef](#)]
84. Mansour, C.; Clodic, D. Optimized energy management control for the Toyota Hybrid system using dynamic programming on a predicted route with short computation time. *Int. J. Autom. Technol.* **2012**, *13*, 309–324. [[CrossRef](#)]
85. Zhang, S.; Xiong, R. Adaptive energy management of a plug-in hybrid electric vehicle based on driving pattern recognition and dynamic programming. *Appl. Energy* **2015**, *155*, 68–78. [[CrossRef](#)]
86. Powell, W.B. *Introduction to Markov Decision Processes. Approximate Dynamic Programming: Solving the Curses of Dimensionality*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011; pp. 57–109.
87. Vagg, C.; Akehurst, S.; Brace, C.J.; Ash, L. Stochastic Dynamic Programming in the Real-World Control of Hybrid Electric Vehicles. *IEEE Trans. Control Syst. Technol.* **2016**, *24*, 853–866. [[CrossRef](#)]
88. Zhou, M.; Liu, Z.; Feng, J. Research on the energy management of composite energy storage system in electric vehicles. *Int. J. Electr. Hybrid Veh.* **2018**, *10*, 41–56. [[CrossRef](#)]
89. Xiao, R.; Li, T.; Zou, G. Energy management strategy for series-parallel hybrid electric vehicle based on stochastic dynamic programming. *Autom. Eng.* **2013**, *35*, 317–321.
90. Jiao, X.; Shen, T. SDP Policy Iteration-Based Energy Management Strategy Using Traffic Information for Commuter Hybrid Electric Vehicles. *Energies* **2014**, *7*, 4648–4675. [[CrossRef](#)]
91. Tian, H.; Li, S.E.; Wang, X.; Huang, Y.; Tian, G. Data-driven hierarchical control for online energy management of plug-in hybrid electric city bus. *Energy* **2018**, *142*, 55–67. [[CrossRef](#)]
92. Tian, H.; Lu, Z.; Wang, X.; Zhang, X.; Huang, Y.; Tian, G. A length ratio based neural network energy management strategy for online control of plug-in hybrid electric city bus. *Appl. Energy* **2016**, *177*, 71–80. [[CrossRef](#)]
93. Zhu, B.; Jiang, Y.; Zhao, J.; He, R.; Bian, N.; Deng, W. Typical Driving Styles Oriented Personalized Adaptive Cruise Control Design Based on Human Driving Data. *Transp. Res. Part C-Emerg. Technol.* **2019**, *100*, 274–288. [[CrossRef](#)]
94. Ibrahim, M.; Jemei, S.; Wimmer, G.; Hissel, D. Nonlinear autoregressive neural network in an energy management strategy for battery/ultra-capacitor hybrid electrical vehicles. *Electr. Power Syst. Res.* **2016**, *136*, 262–269. [[CrossRef](#)]
95. Liu, J.; Chen, Y.; Zhan, J.; Shang, F. An On-Line Energy Management Strategy Based on Trip Condition Prediction for Commuter Plug-In Hybrid Electric Vehicles. *IEEE Trans. Veh. Technol.* **2018**, *67*, 3767–3787. [[CrossRef](#)]
96. Murphey, Y.L.; Park, J.; Chen, Z.; Kuang, M.L.; Masrur, M.A.; Phillips, A.M. Intelligent Hybrid Vehicle Power Control—Part I: Machine Learning of Optimal Vehicle Power. *IEEE Trans. Veh. Technol.* **2012**, *61*, 3519–3530. [[CrossRef](#)]
97. Chen, Z.; Hu, H.; Wu, Y.; Xiao, R.; Shen, J.; Liu, Y. Energy Management for a Power-Split Plug-In Hybrid Electric Vehicle Based on Reinforcement Learning. *Appl. Sci.* **2018**, *8*, 2494. [[CrossRef](#)]
98. Wu, J.; He, H.; Peng, J.; Li, Y.; Li, Z. Continuous reinforcement learning of energy management with deep Q network for a power split hybrid electric bus. *Appl. Energy* **2018**, *222*, 799–811. [[CrossRef](#)]



99. Zhao, P.; Wang, Y.; Chang, N.; Zhu, X.; Lin, X. Deep Reinforcement Learning Framework for Optimizing Fuel Economy of Hybrid Electric Vehicles. In Proceedings of the 2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC), Jeju, Korea, 22–25 January 2018.
100. Xiong, R.; Cao, J.; Yu, Q. Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle. *Appl. Energy* **2018**, *211*, 538–548. [[CrossRef](#)]
101. Hu, Y.; Li, W.; Xu, K.; Zahid, T.; Qin, F.; Li, C. Energy Management Strategy for a Hybrid Electric Vehicle Based on Deep Reinforcement Learning. *Appl. Sci.* **2018**, *8*, 187. [[CrossRef](#)]
102. Liu, T.; Hu, X.; Li, S.E.; Cao, D. Reinforcement Learning Optimized Look-Ahead Energy Management of a Parallel Hybrid Electric Vehicle. *IEEE/ASME Trans. Mechatron.* **2017**, *22*, 1497–1507. [[CrossRef](#)]
103. Luo, B.; Liu, D.; Huang, T.; Yang, X.; Ma, H. Multi-step heuristic dynamic programming for optimal control of nonlinear discrete-time systems. *Inf. Sci.* **2017**, *411*, 66–83. [[CrossRef](#)]
104. Zhang, H.; Qin, C.; Luo, Y. Neural-Network-Based Constrained Optimal Control Scheme for Discrete-Time Switched Nonlinear System Using Dual Heuristic Programming. *IEEE Trans. Autom. Sci. Eng.* **2014**, *11*, 839–849. [[CrossRef](#)]
105. Huang, Y.; Liu, D. Neural-network-based optimal tracking control scheme for a class of unknown discrete-time nonlinear systems using iterative ADP algorithm. *Neurocomputing* **2014**, *125*, 46–56. [[CrossRef](#)]
106. Mu, C.; Wang, D.; He, H. Data-Driven Finite-Horizon Approximate Optimal Control for Discrete-Time Nonlinear Systems Using Iterative HDP Approach. *IEEE Trans. Cybern.* **2017**, *48*, 2948–2961. [[CrossRef](#)]
107. Mu, C.; Sun, C.; Song, A. Iterative GDHP-based approximate optimal tracking control for a class of discrete-time nonlinear systems. *Neurocomputing* **2016**, *214*, 775–784. [[CrossRef](#)]
108. Mu, C.; Wang, D.; He, H. Novel iterative neural dynamic programming for data-based approximate optimal control design. *Automatica* **2017**, *81*, 240–252. [[CrossRef](#)]
109. Chen, H.-X.; Nan, Y.; Yang, Y. A Two-Stage Method for UCAV TF/TA Path Planning Based on Approximate Dynamic Programming. *Math. Probl. Eng.* **2018**, *2018*, 1092092. [[CrossRef](#)]
110. Li, W.; Xu, G.; Xu, Y. Online learning control for hybrid electric vehicle. *Chin. J. Mech. Eng.* **2012**, *25*, 98–106. [[CrossRef](#)]
111. Qin, F.; Li, W.; Hu, Y.; Xu, G. An Online Energy Management Control for Hybrid Electric Vehicles Based on Neuro-Dynamic Programming. *Algorithms* **2018**, *11*, 33. [[CrossRef](#)]
112. Ali, A.M.; Soffker, D. Realtime application of progressive optimal search and adaptive dynamic programming in multi-source HEVS. In Proceedings of the ASME 10th Annual Dynamic Systems and Control Conference, Tysons, VA, USA, 11–13 October 2017.
113. Ali, A.M.; Soffker, D. Realtime power management of a multi-source hev using adaptive dynamic programming and probabilistic drive state. In Proceedings of the ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, Cleveland, OH, USA, 6–9 August 2017; pp. 2159–7383.

