ORIGINAL RESEARCH **Open access**

Reinforcement Learning–Driven Energy Management of Series–Parallel Hybrid Electric Vehicles Using Q-Learning

Karolina Nowak[1*], Tomasz Wrobel[2], Julia Mazur[1]

Abstract

This work explores an energy management framework grounded in reinforcement learning for series-parallel hybrid vehicle architectures. The intricate nature of power distribution in hybrid electric vehicles facilitates the deployment of more sophisticated control policies. Capitalizing on this characteristic, a Q-Learning algorithm is developed to formulate an energy management strategy. Departing from prior investigations, an online reward function is constructed to optimize fuel economy and battery service life concurrently. Another distinguishing aspect is that the developed method does not call for any foreknowledge of the driving cycle or a high-fidelity vehicle representation. The strategy's performance is evaluated across four driving cycles via co-simulation between MATLAB and ADVISOR. The resulting data reveal that, under the HWFET cycle, fuel usage drops by 1.25%, while battery longevity improves by 65% compared to the rule-based baseline controller embedded in ADVISOR. Outcomes from additional driving cycles further substantiate the strategy's inherent self-improvement capability. It is also shown that the approach sustains its performance and surpasses the rule-based controller even when the driving cycle is altered.

Keywords Reinforcement learning, Hybrid electric vehicles, Series-parallel configuration, Power management policy, Q-Learning

*Correspondence:
Karolina Nowak
karolina.nowak@pw.edu.pl
[1] Department of Intelligent Energy Engineering, Faculty of Electrical Engineering, Warsaw University of Technology, Warsaw, Poland
[2] Department of AI and Smart Infrastructure, Faculty of Engineering, AGH University of Science and Technology, Krakow, Poland

Introduction

The progression of modern civilization has been substantially shaped by automobiles. In recent times, scholarly attention has turned toward key parameters that reduce fuel consumption and emissions, aiming to produce more efficient vehicles. Within this sphere, Hybrid Electric Vehicles (HEVs) offer advantages in regulating both fuel use and pollutant outputs. HEVs incorporate a hybrid propulsion architecture consisting of an internal combustion engine (ICE) serving as the principal power source, supplemented by an electric motor that provides additional torque by transforming stored electrical energy. This arrangement enables more sophisticated power management strategies designed to realize energy savings and curtail emissions [1, 2].

Power management control strategies in HEVs fall into two broad classifications: rule-based and optimization-based methodologies. Rule-based schemes are designed to confine the state of charge (SoC) to a designated window and keep engine operation within a high-efficiency band. Optimization-based schemes, by contrast, formulate an objective cost function tied to either fuel consumption or SoC to minimize equivalent fuel consumption and boost overall vehicular efficiency [3, 4]. In the context of SoC monitoring, many approaches employ battery models coupled with diverse algorithms. Genetic algorithm techniques are deployed for model parametrization to refine both accuracy and complexity

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0), which permits non-commercial use, distribution, reproduction, and adaptation in any medium or format, provided that appropriate credit is given to the original author(s) and source, a link to the license is provided, and any changes made are indicated. Adapted material must be distributed under the same license.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

[5]. A neural network-based solution is documented in [6] to address nonlinear behaviors inherent in battery characteristics. Reference [7] proposes a model-based observer for estimating the SOC of lithium-ion batteries used in hybrid electric vehicles, and the method is tested under fluctuating load scenarios. A few other studies opt for independent methodologies. Specifically, the wavelet transformation matrix serves as a tool for SoC estimation of lithium-ion batteries in electric vehicles [8]. An enhanced feedforward-long short-term memory network is presented in Ref. [9] to deliver precise whole-life-cycle SOC predictions for lithium-ion batteries by incorporating fluctuations in temperature, current, and voltage. In that work, an optimized sliding balance mechanism is adopted to harmonize the current and refine SOC prediction accuracy. A deep convolutional neural network founded on process data modeling is introduced to predict SOC, thereby ensuring a reliable predictive state for the battery. To sustain engine operation within the high-efficiency zone, deterministic [10] and fuzzy [11] methods have been advanced. Rule-based strategies maintain both the state of charge (SoC) boundaries and the engine's efficient operating region through deterministic and fuzzy logic schemes [10-12]. To boost powertrain and associated element efficiency, a rule-based supervisory control framework [13] and an optimal control scheme for the fuel cell coupled with the battery charging sequence [14] have been suggested, capable of securing up to 95% of the achievable fuel economy. Optimization-based strategies deploy an objective cost function formulated around fuel consumption to reduce equivalent fuel usage and elevate vehicle efficiency. More recently, Wang *et al.* [15] introduced a multi-objective energy management scheme aimed at optimizing both fuel consumption reduction and torsional vibration suppression. Nonetheless, the strategies mentioned above remain uniformly contingent on the availability of an accurate control system model.

Over the past several years, a growing body of research [16, 17] has examined artificial intelligence-driven techniques for orchestrating power management in HEVs. Reinforcement Learning (RL), functioning as an intelligent algorithmic framework, facilitates power management regulation without dependence on a precise system representation by learning from reward and penalty feedback. Representative RL algorithms applied to HEV energy management encompass Deep Deterministic Policy Gradient (DDPG) [18], TD(λ)-learning [19], Deep Q-network (DQN) [20], and Trust region policy optimization (TRPO) [21]. Teng *et al*. [22] proposed an RL-optimized energy management algorithm, using data-driven methods to enhance fuel savings. Yue *et al*. [23] described a deep Q-learning strategy for minimizing fuel expenditure in hybrid vehicles under sparse reward conditions. Yet, the sluggish coverage during the training phase associated with sparse reward structures compels investigators [24] to adopt shaped reward functions for steering energy management in HEVs.

Earlier works within the domain of AI-based energy management strategies have predominantly targeted singular objectives—either fuel consumption minimization or battery lifetime extension—and their efficacy has typically hinged upon having precise driving cycle information. By comparison, the current study centers on a Q-learning strategy that employs a reward function engineered to optimize both fuel consumption and battery life cycle simultaneously. In essence, this approach exploits the interrelationship between fuel usage and battery durability. The principal objective is to achieve minimal fuel consumption alongside maximal battery life across any driving cycle, all without requiring advance knowledge of those cycles in HEV applications.

Vehicle Modeling

The research object selected is a Toyota Prius featuring a series-parallel hybrid architecture [25]. The series-parallel layout merges the strengths of both series and parallel topologies. This arrangement allows the engine and the electric motor to supply power either separately or in combination. Put differently, the ICE can either turn the wheels directly (following the parallel drivetrain principle) or operate while the electric motor alone delivers motive force (following the series drivetrain principle). **Figure 1** shows the layout of a series-parallel hybrid vehicle. As shown, a battery pack feeds the electric motor (green line), and a fuel tank feeds the IC engine (violet line). Depending on the power-delivery pathway (series drivetrain or parallel drivetrain), the wheels receive propulsion either directly from the ICE or exclusively from the electric motor (orange line). The parameter set for a hybrid Toyota Prius is listed in **Table 1** [25].

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

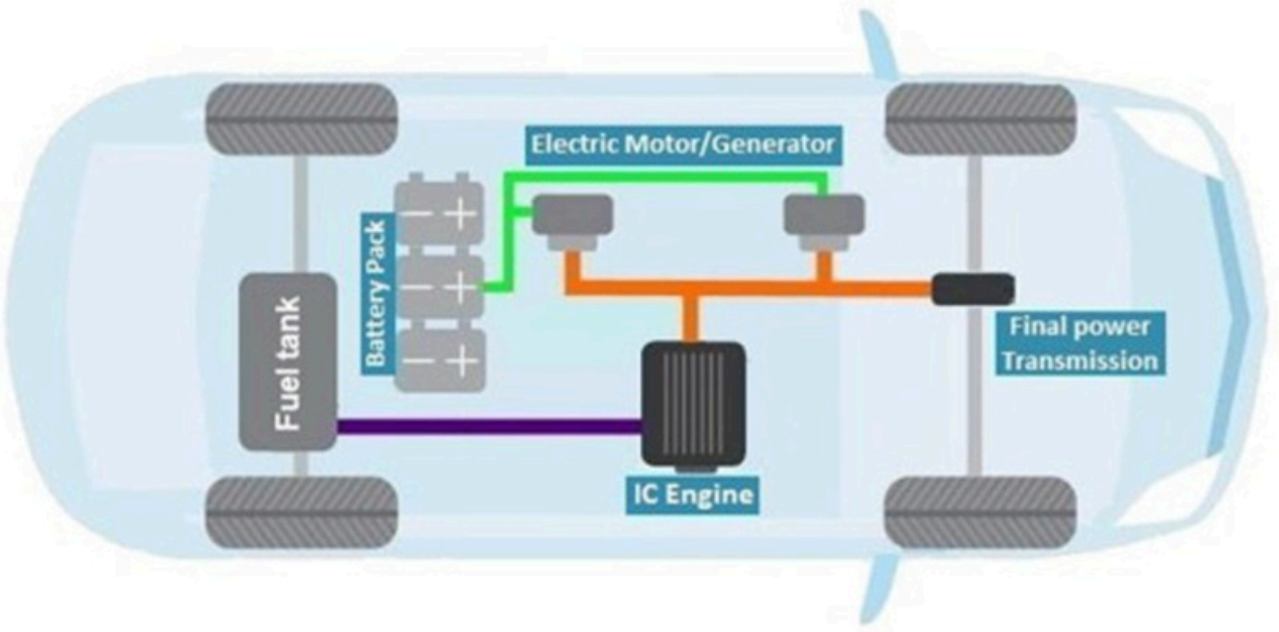


Figure 1. Series-parallel hybrid configuration of a Toyota Prius vehicle.

Table 1. Prius parameters.

Item	Specification	Value
Combustion engine	Configuration	
	Maximum output in KW/rps	1.8e-3 m^3 gasoline
	Maximum torque in Nm/rps	73/544.5
		142/418.88
Electric motor	Configuration	
	Maximum output in kW/rps	Synchronous AC
	Maximum torque in Nm/rps	60/109–586.43
		350/0–400
Generator	Configuration	
	Maximum output in kW	Synchronous AC
	Maximum speed in rps	42
		1047.2
Battery	Configuration	
	Maximum output kW	Nickel-metal hydride
		23.4
Max speed (km/h)	–	48.2

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Vehicle dynamics

The power required by the vehicle, or the traction force, can be derived from the force equilibrium equation grounded in Newton's laws, expressed as follows [26]:

$$F_T = F_g + F_R + F_A + ma_v\, F_g = mg\, sin\, \sin\theta\, F_R = C_R.mg\, cos\, \cos\theta\, F_A = \frac{1}{2}C_D\rho A_F v^2 \quad (1)$$

Where m, av, FT and CR stand for the vehicle mass, vehicle acceleration, total traction force, and rolling resistance coefficient, respectively. Furthermore, AF, CD, ρ, v and θ correspond to the frontal area, the drag coefficient, the air density, vehicle velocity, and the road slope, with the slope being omitted from consideration in the present work.

The planetary gear constitutes the heart of the Prius, facilitating power coupling among the engine, motor, and generator before delivery to the wheels. The final traction torque is obtained as [26]:

$$T_{fd} = \frac{T_d}{\eta_{fd} z_{fd}} \quad (2)$$

Where Td denotes the wheel torque, and ηfd and zfd represent the efficiency and the gear ratio of the final traction torque, respectively.

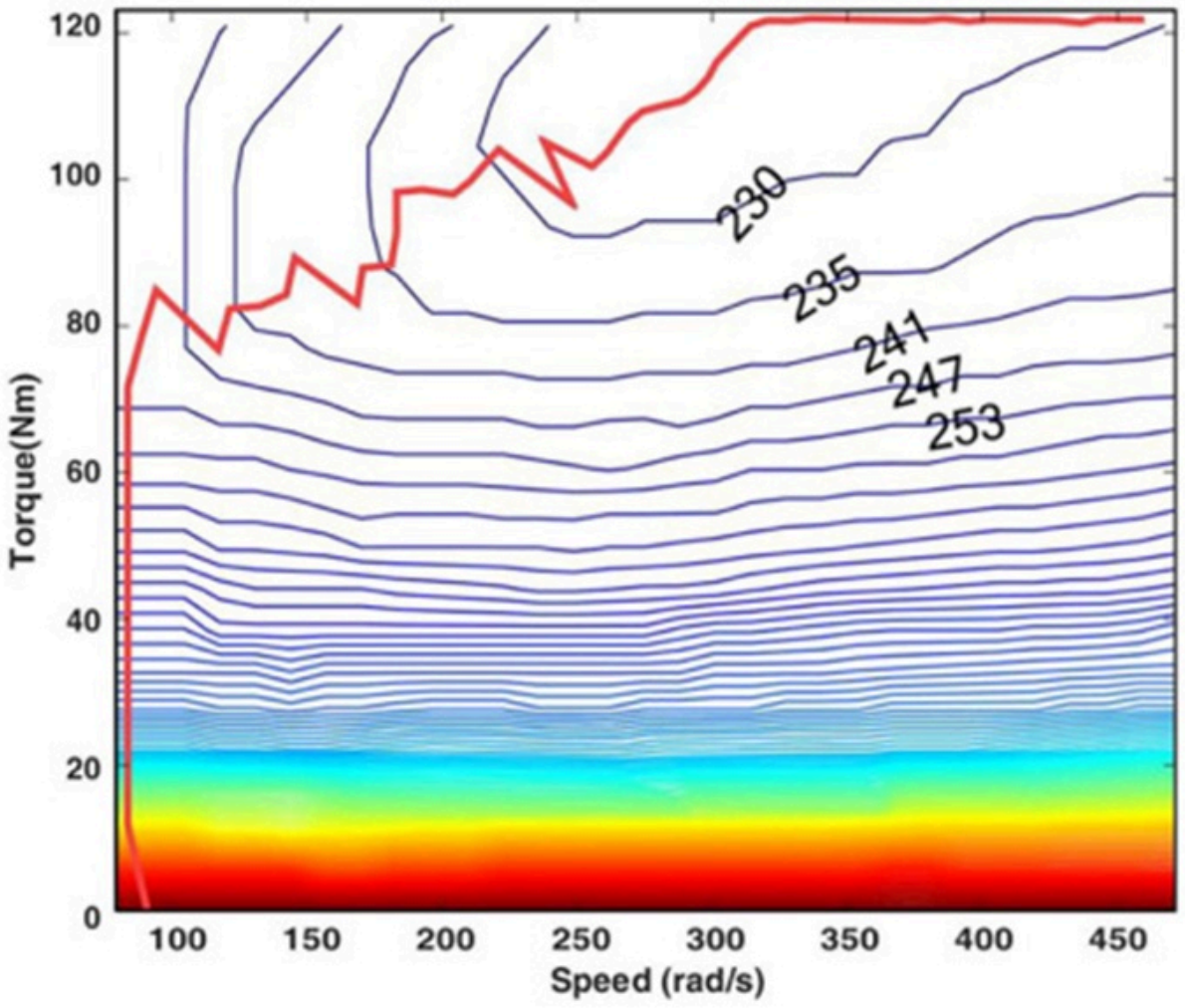
Model of powertrain system

The IC engine, generator, and motor are modeled using their respective efficiency maps [27]. The battery fitted in the Toyota Prius is composed of Ni-MH cells, whose behavior can be captured by:

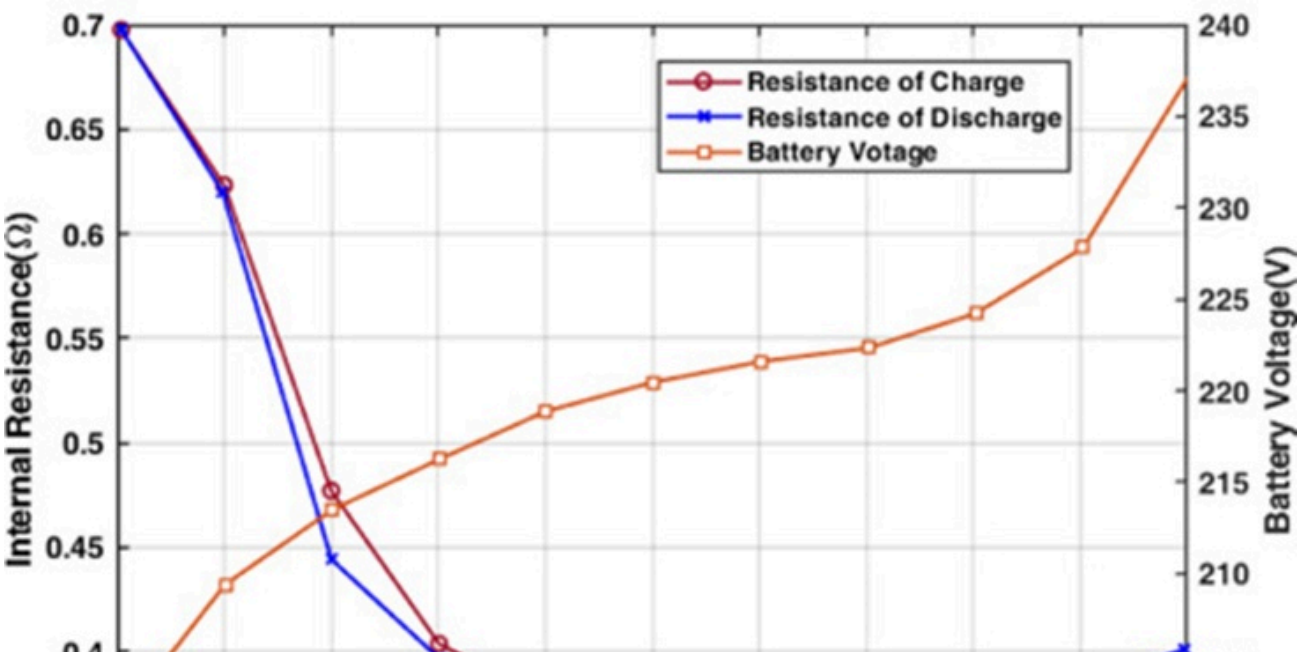
$$P_s = \left(\eta^{+}_{bat}(SoC, P_{bat}), \frac{1}{\eta^{-}_{bat}(SoC, P_{bat})}\right)$$
$$SoC = \frac{\int_0^t I_b(t)dt}{Q_0} \quad (3)$$

Where Ps is net power, η^{+}_{bat} battery power in the charge state, η^{-}_{bat} battery power in the discharge state, Pbat is battery power, SoC is the state of charge, Q is the nominal battery capacity, and Ib is battery current, which is the function of output open-circuit voltage and internal resistance. The experimentally obtained data for the battery, covering internal resistance, output open-circuit voltage, and the engine map, are displayed in **Figure 2.**

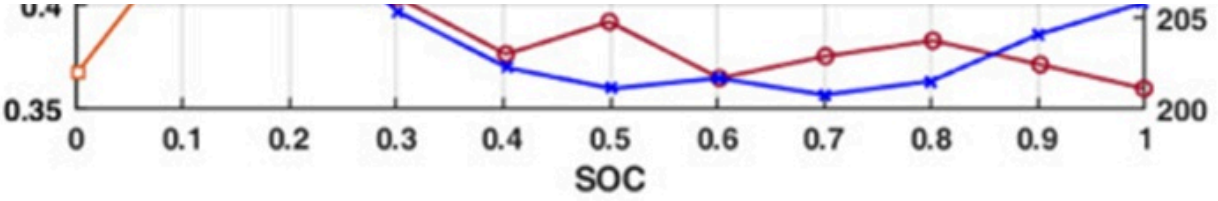
© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press



(a) Engine map



© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press



(b) Experimental data of battery

measured relative to the rated cell capacity, as opposed to the state of charge of the cell with respect to a fully charged condition. DoD is formulated as [28]:

$$d_{eff} = \left(\frac{D_A}{D_R}\right)^{u_0} e^{u_1\left(\left(\frac{D_A}{D_R}\right)-1\right)} d_{actual} \quad (4)$$

Where d_{eff}, D_A, d_{actual} and D_R denote effective discharge, the minimum battery charge level, actual discharge, and depth of discharge, respectively.

The second factor governing battery life is Discharge Rate. Raising this factor extends battery life by mitigating conductivity losses. The factor is given by [28]:

$$d_{ff} = \frac{C_R}{C_A} d_{actual} \quad (5)$$

Where CA and CR correspond to the actual discharge current of interest and the rated amp-hour capacity at the rated discharge current, respectively; accordingly, the effective discharge can be computed by combining **Eq. 4** and **Eq. 5** as:

$$d_{eff} = \left(\frac{D_A}{D_R}\right)^{u_0} e^{u_1\left(\left(\frac{D_A}{D_R}\right)-1\right)} \left(\frac{C_R}{C_A}\right) d_{actual} \quad (6)$$

This expression yields the effective discharge for a single discharge occurrence. To estimate the life prediction, summing the effective discharge across N driving cycles is necessary, as shown below:

$$L_{time} = \sum_{0}^{N} d_{eff} \quad (7)$$

Q-Learning for HEV Power Management

When compared to rule-based control methodologies, Q-Learning brings forth the following compelling characteristics that render it suitable for power management regulation in HEVs,

(i) Altering the system states requires consideration of both the present state and the requisite actions. In HEVs, actions encompass vehicle speed, power demand, and battery charge level, whereas the battery life state is tied to the charging/discharging activity in the current condition.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

(ii) The cumulative reward undergoes optimization, which in HEV terminology translates to the joint optimization of fuel consumption and battery life throughout an entire driving cycle.

(iii) Neither a detailed system representation nor prior knowledge of the states is needed. In the context of the current study, the driving cycle is not known beforehand in its entirety.

(iv) It proves beneficial for non-stationary systems. Driving cycles are inherently non-stationary in nature [29].

Q-learning implementation

State space

The state space for the power management task is formed by discretizing the continuous state variables, which comprise power demand, vehicle speed, and the battery state of charge, and is expressed as:

$$S = \{s = [p_{dem}, v, soc] |\ p_{dem} \in P_{dem},\ v \in V,\ soc \in SOC\} \quad (8)$$

Where p_{dem} is power demand, soc is the battery state of charge, and v is vehicle speed. P_{dem}, V and SOC are power demand levels, vehicle speed levels, and state of charge levels, respectively. To carry out the implementation, it is essential to discretize the finite sets introduced in **Eq.8**. For illustration, SOC is obtained by partitioning the state-of-charge range within the bounds [soc_{min}, soc_{max}] into a finite collection:

$$\mathrm{SOC} = \{\mathrm{soc}_1, \mathrm{soc}_2, \ldots, \mathrm{soc}_N\} \quad (9)$$

Where, $\mathrm{soc}_{min} \leq \mathrm{soc}_1 < \mathrm{soc}_2 < \ldots < \mathrm{soc}_{max}$. An analogous discretization procedure is applied to P_{dem} and V.

Action space

The action space consists of a finite collection of power values, defined as:

$$A = \{a = [\mathrm{P_{bat}}] | \mathrm{P_{bat}} \in \mathrm{P_{bat}}\} \quad (10)$$

Where a is the action chosen for charging or discharging with a designated value of Pbat. The set Pbat encompasses a finite number of power levels spanning the interval [-Pbat, max, Pbat, max]. Pbat > 0 signifies battery discharging, while Pbat < 0 indicates battery charging. Framing the action space in this manner allows the reinforcement learning method to function without engaging the intricate specifics of HEV modeling.

Reward

Figure 3 portrays the interaction between the agent and the environment in the reinforcement learning framework. The variable r denotes the reward the agent receives upon executing action a.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

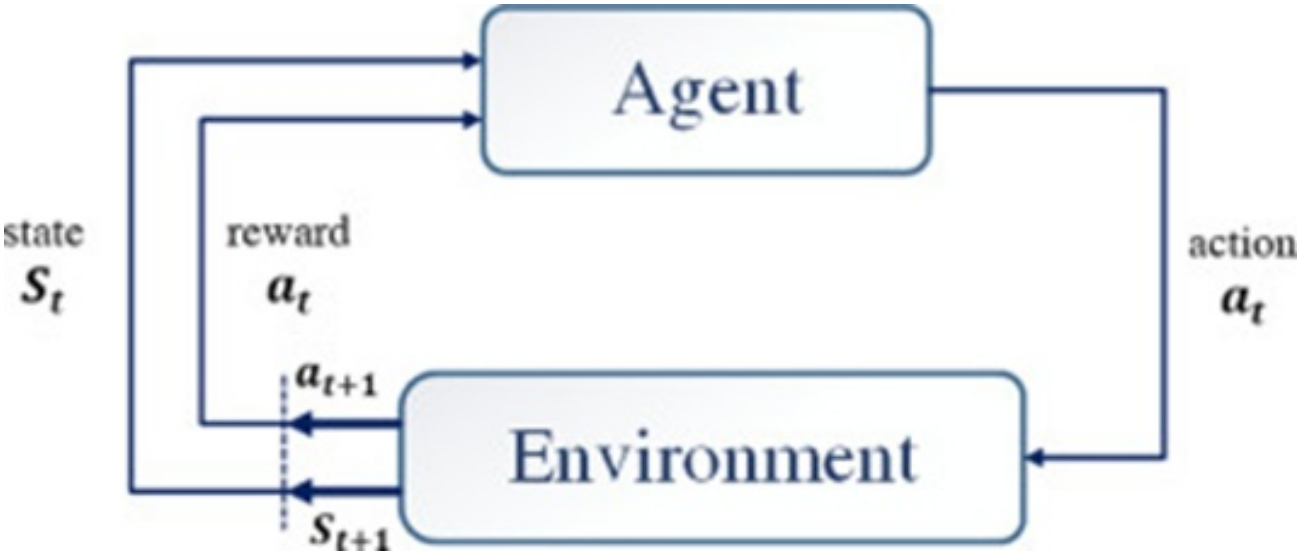


Figure 3. Agent-environment interaction in RL.

For the HEV examined in this study, the reward is specified as:

$$r = -K_f \dot{m}_f \Delta T - K_{bat} d_{eff} \Delta T \qquad (11)$$

Where $\dot{m}_f \Delta T$ and $d_{eff} \Delta T$ represent the fuel consumption and the effective discharge occurring within the time step ΔT, respectively. By treating the negative values of fuel consumption and effective discharge over a time step as the reward, both fuel consumption and battery life are optimized by balancing Kf and K_{bat}.

Q-Learning algorithm for HEV

The implementation of the HEV power management strategy is accomplished through the application of a Q-learning algorithm [30]. The algorithm assigns a Q-value, written as Q(s, a), to every pair of a state s and an action a. This value represents an estimate of the total accumulated reward anticipated when the system executes action a from state s. The algorithm's execution revolves around two core procedures: determining which action to take and revising the stored Q-values.

Action selection

The mechanism for choosing an action operates by consulting and picking the entry with the greatest Q-value. This process guides the system toward the most advantageous choice among the complete catalogue of admissible actions. This decision-making procedure incorporates two distinct operating phases: exploration and exploitation. Throughout the exploration phase, the system adheres to an ϵ-greedy policy, meaning the optimal action is adopted with likelihood 1 - ϵ; during exploitation, the action that directly yields the maximum Q-value is invariably selected.

Q-Value update

Following the interaction flow depicted in **Figure 3**, the system acts at while situated in state st at the temporal index t. Upon transitioning to the next moment, t + 1, the environment returns an immediate reward signal rt+1 together with the newly occupied state st+1. With these observations available, the Q-value tied to each state-action combination (s, a) undergoes an update at step t + 1 governed by the expression below:

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

$$Q(s,a) \leftarrow Q(s,a) + \alpha(r_{t+1} + \gamma Q(s_{t+1}, a') - Q(s,a)) \quad (12)$$

Where the parameter α corresponds to the learning rate, and γ represents the discount factor.

Algorithm procedure

A complete description of the Q-learning algorithm devised for HEV power management is laid out in the following steps [30].

Q-Learning Algorithm for HEV Power Management

Initialize Q(s, a).

Repeat (for each episode):

Initialize s

Repeat (for each step of the episode):

Select a from s using Q policy

Take action a, observe r, s′

Q(s,a)←Q(s,a) + α(rt+1 + γQ(st+1,a′) - Q(s,a))s←s′

Until s is terminal.

A schematic illustration of how the algorithm is deployed for the HEV is given in **Figures 3 and 4.**

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

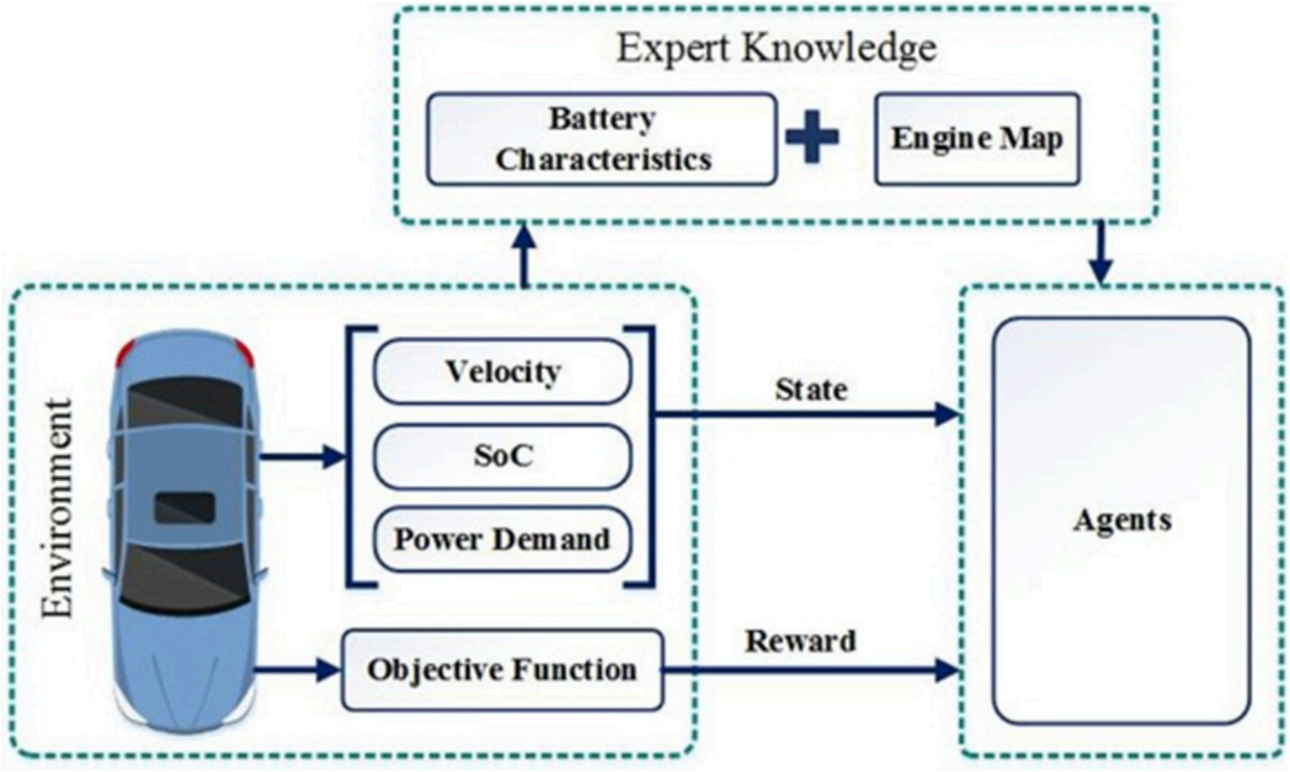


Figure 4. The schematic of the Q-learning algorithm for hybrid Toyota Prius.

Simulation Results

The strategy proposed in this investigation is tested through co-simulation involving ADVISOR and MATLAB Simulink on an HEV platform modelled after the Toyota Prius. More concretely, the vehicle representation native to ADVISOR is coupled to an optimization routine constructed in MATLAB Simulink. The key model characteristics that serve as ADVISOR inputs are collated in **Table 2** [31]. The exploration parameter ϵ is assigned an initial magnitude of 0.7 at the launch of the algorithm and thereafter undergoes an exponential reduction consistent with a greedy scheme, progressively steering the behavior toward the optimal policy. Before each driving cycle, the cumulative reward counter is reset to zero, and the battery SoC is restored to its predefined starting level.

Table 2. Main parameters of hybrid electric vehicle.

Symbol	Description	Value
m	Vehicle mass	1380 kg
A	Vehicle frontal area	2.6005 m^2
Cd	Vehicle drag coefficient	0.25
Rtire	Wheel radius	0.317 m
fr	Wheel friction coefficient	0.008
ρair	Density of air	1.29 kg/m^3
fbfd	Brake force distribution	0.93
CR	Rated amp-hour capacity of battery	6.5 Ah

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

DR	Depth of Discharge	0.3
u0	–	0.8038
u1	–	0.4091

The simulation exercise targets the HWFET driving cycle. **Figure 5** captures the evolution of the Q-learning algorithm over the course of this cycle. From **Figures 5b and 5f**, it is evident that the velocity reference is accurately tracked, and the aggregate reward reaches its highest level after 8200 episodes. The trajectories for fuel consumption and battery life cycle are plotted in **Figures 5a and 5c**. A clear trend emerges: as episode 8200 concludes, fuel usage exhibits a notable reduction while the battery life cycle increases. The temporal variation of battery SoC is additionally provided in **Figure 5e.**

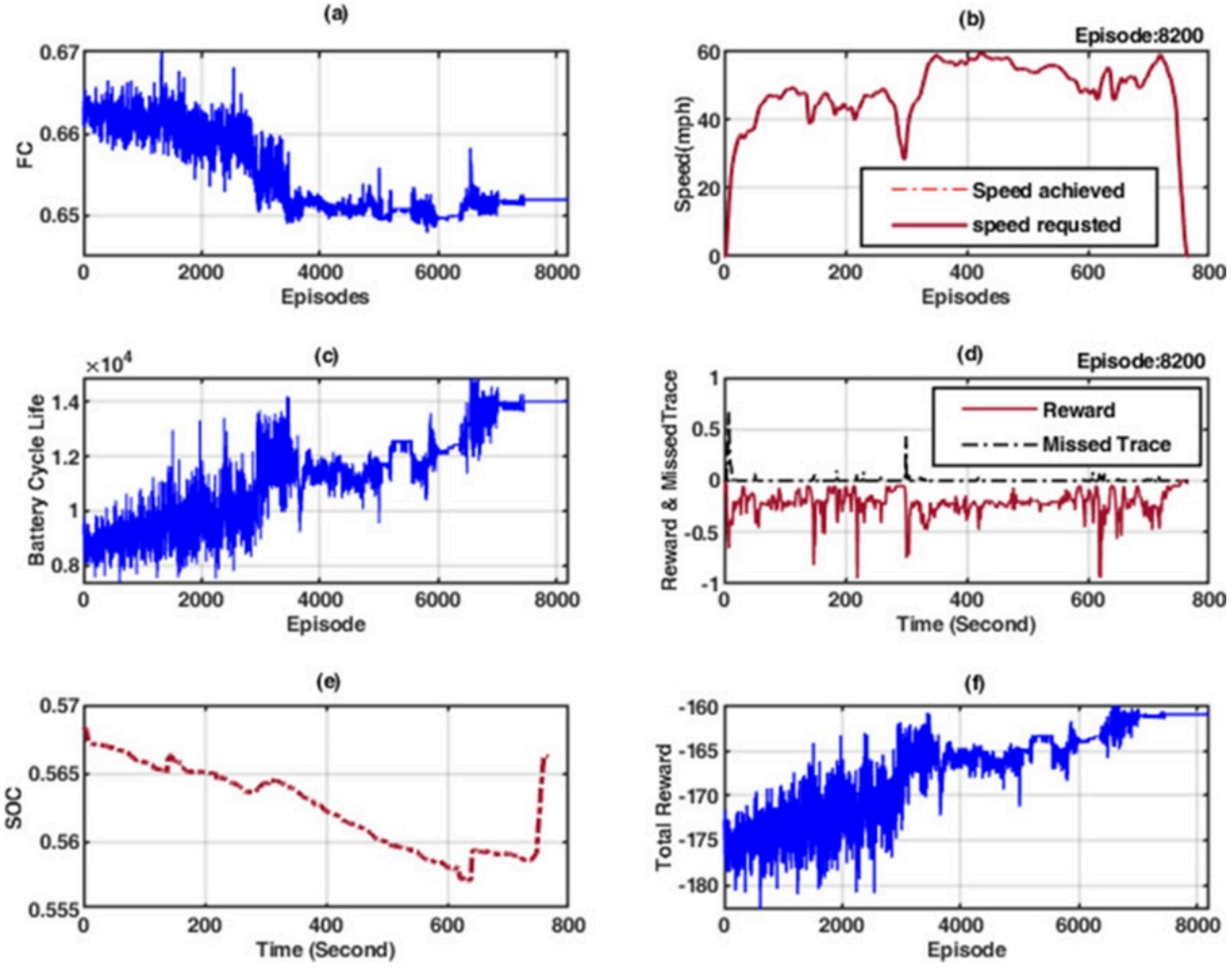


Figure 5. Results of Q-learning algorithm in HWFET cycle.

The control variable manipulated by the algorithm is the battery power. A comparison is drawn between the power demand dictated by the Q-learning strategy and that prescribed by the rule-based controller embedded within ADVISOR, both evaluated on the HWFET cycle. The side-by-side results appear in **Figure 6**. The comparison reveals that the Q-learning-based power demand is consistently lower in magnitude and displays considerably less oscillatory behavior than

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

its rule-based counterpart. This observation suggests that the Q-learning formulation helps extend the battery's operational lifespan in a hybrid electric vehicle. Within the reinforcement learning paradigm, the optimal behavioral policy emerges through continuous interaction between the learning agent and its operational environment. In the specific case of Q-learning, this emergence is mediated by a state-action value matrix denoted Q. Starting from an initial guess Q0, the matrix is iteratively refined according to the steps of the proposed algorithm and the stream of environmental feedback. Once convergence is reached, the final matrix encodes the optimal policy. The dimensions of this matrix are determined by the total count of states and the total count of permissible control actions. The converged Q matrix resulting from the power management task on the HWFET cycle is visualized in **Figure 7.**

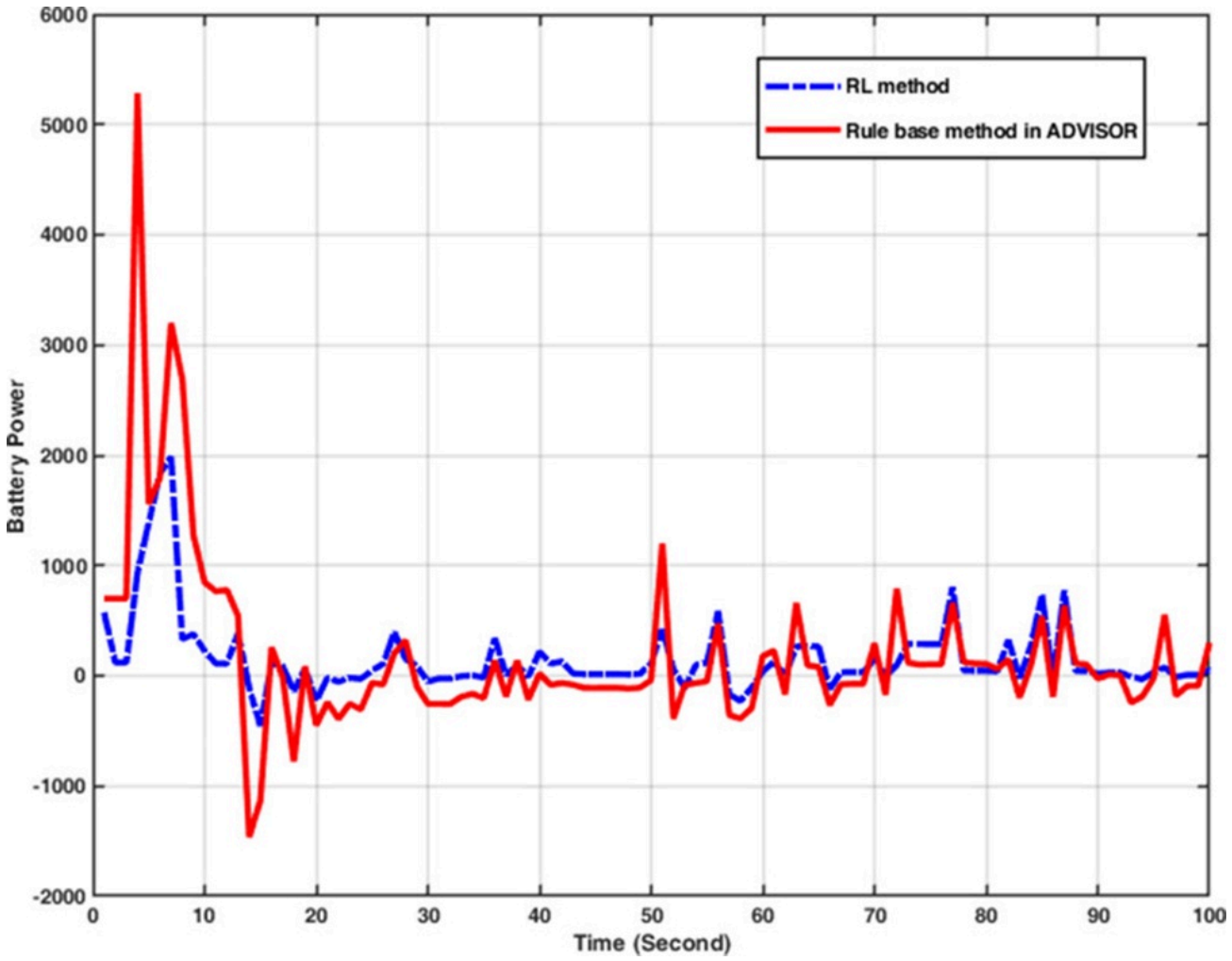


Figure 6. Power demand of battery.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

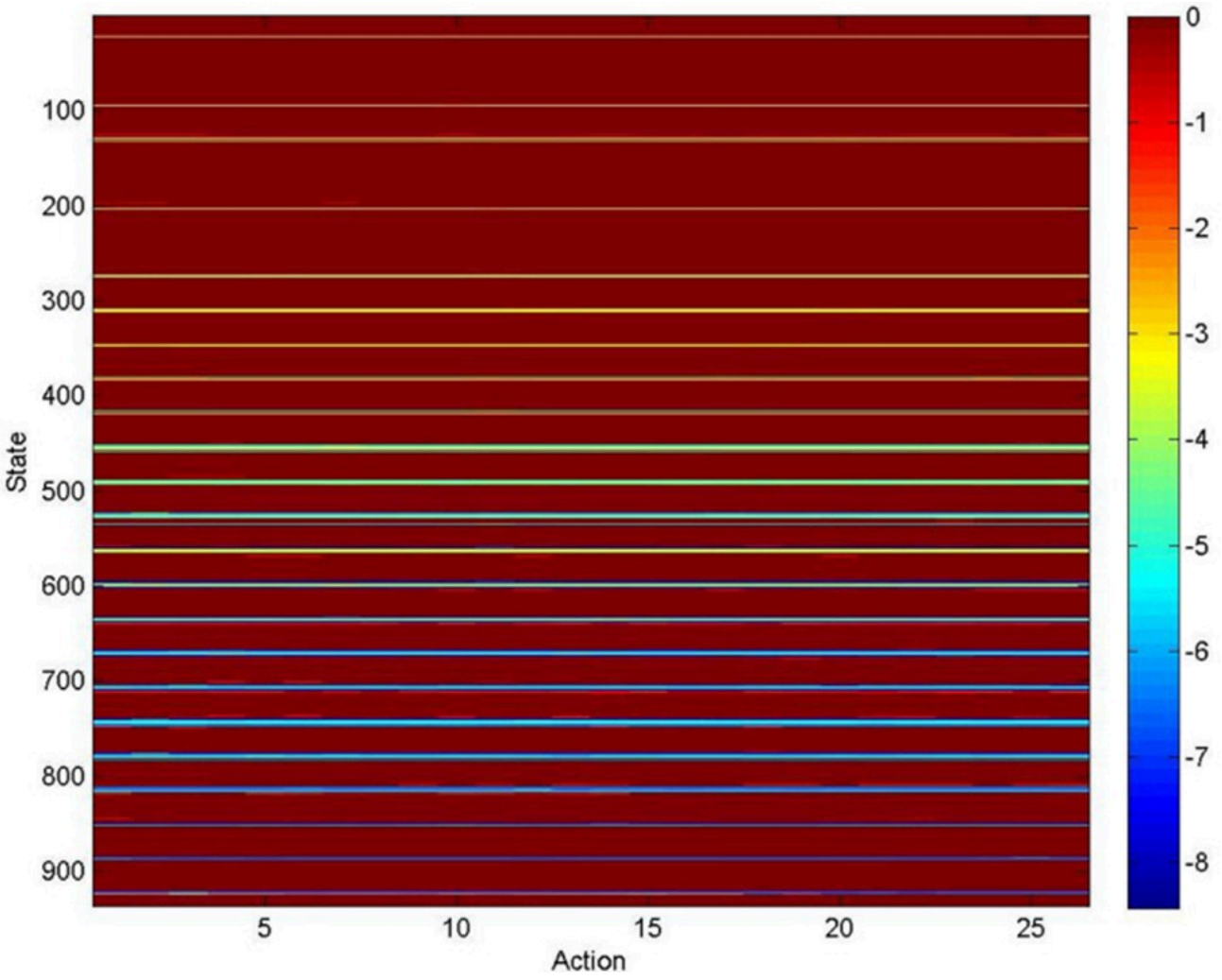


Figure 7. Q matrix for HWFET Cycle.

The evolution of the battery State of Charge (SOC) throughout the HWFET cycle, as produced by both the Q-learning algorithm and the rule-based approach, is contrasted in **Figure 8.**

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

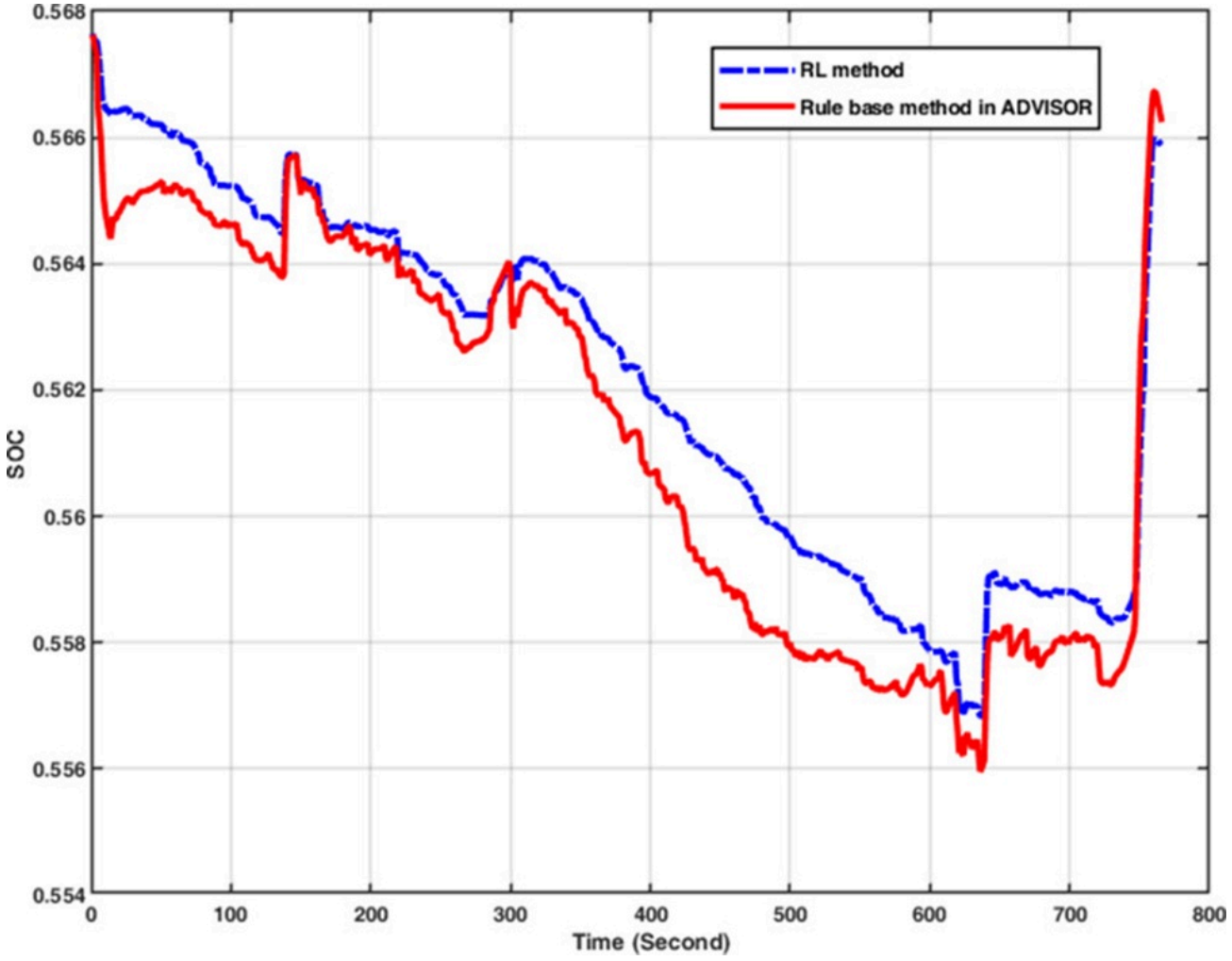


Figure 8. The state of the battery.

Figure 9 provides a view of the operating points visited by the electric motor and the IC engine during the cycle. **Figure 10** shows similar cycles to HWFET.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

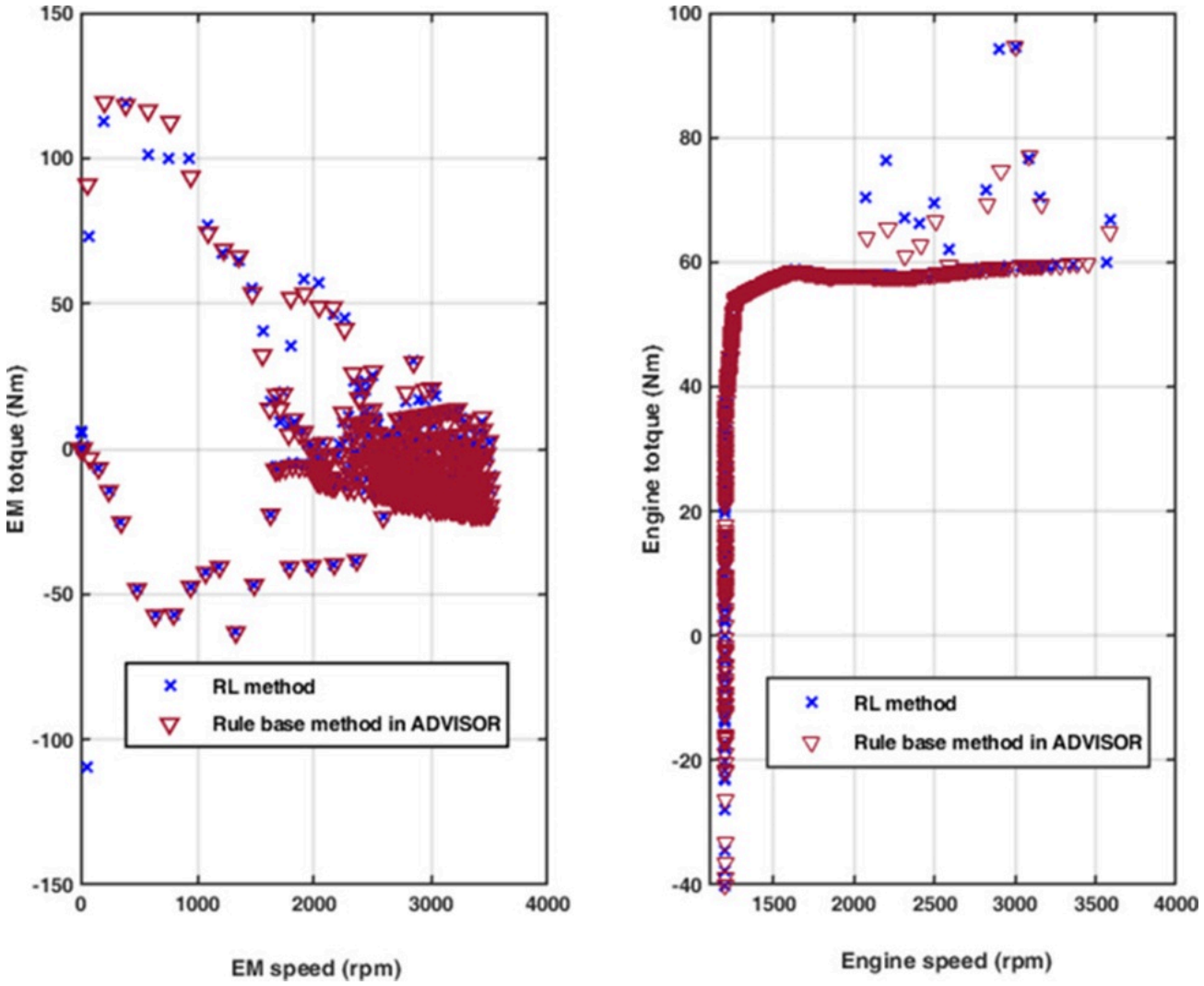


Figure 9. Operating points for electric motor and IC engine.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

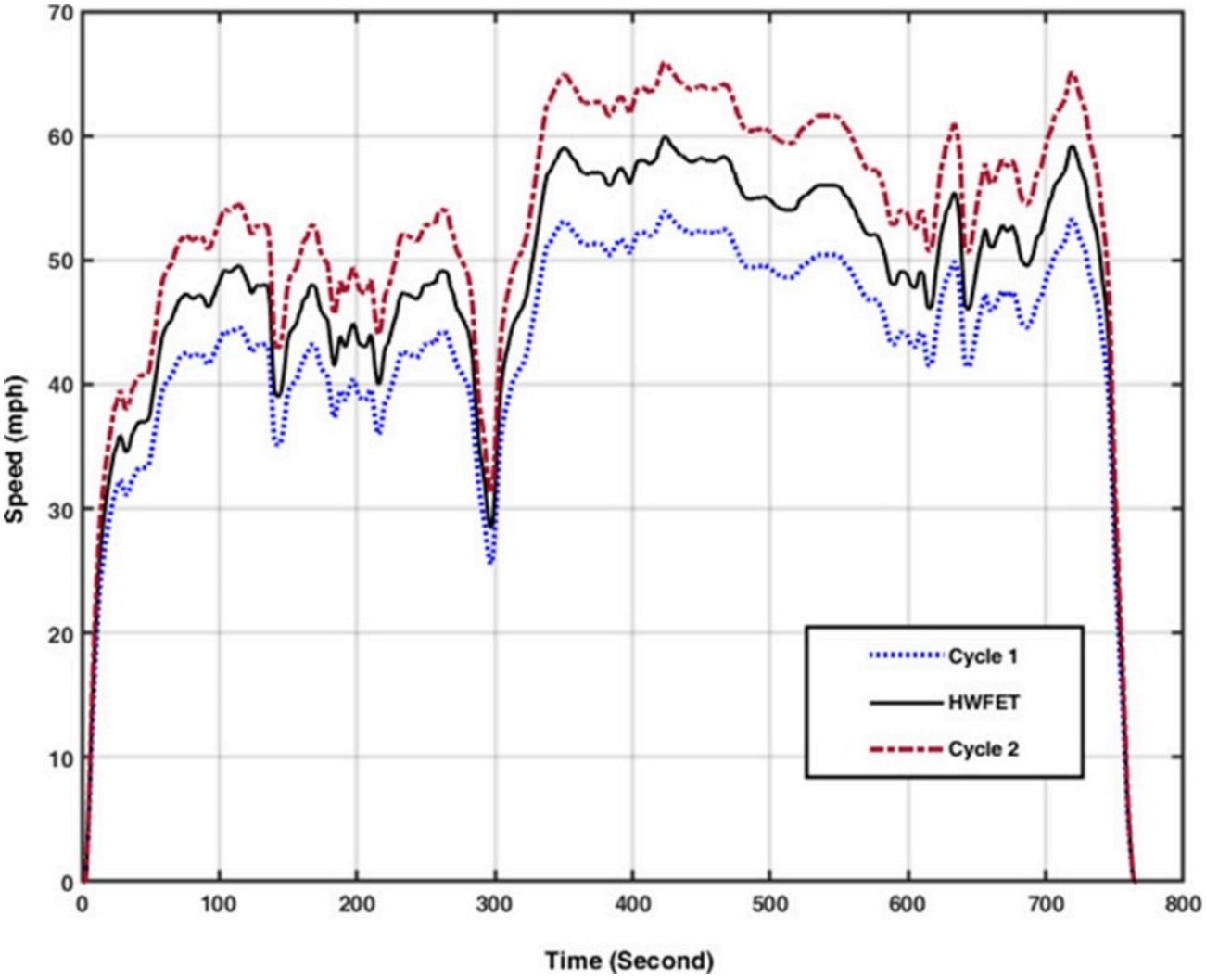


Figure 10. Similar cycles to HWFET.

The proposed technique is further tested on three supplementary driving cycles: IM240, UDDS, and NEDC. Across all these cycles, the Q-learning method is directly benchmarked against the rule-based controller available in ADVISOR. The quantitative outcomes for all cycles are assembled in **Table 3.** The data show that on the HWFET cycle, fuel consumption decreases by 1.25% while battery life increases by 65% compared with the rule-based method. On the IM240 cycle, fuel consumption decreases by 0.68% and battery life improves by 50%. Turning to the UDDS and NEDC cycles, battery life enhancements of 47% and 36% are respectively realized.

Table 3. Compared results for different cycles.

Cycle test	Parameter/Unit	Variation (%)	Rule-based ADVISOR	Q-learning approach
HWFET	Fuel consumption (L/100 km^{-1})	−1.25	4	3.95
	Battery life (cycle)	62.5	0.88×10^4	1.43×10^4
IM240	Fuel consumption (L/100 km^{-1})	−0.68	5.9	5.86
	Battery life (cycle)	50	1.04×10^4	1.56×10^4
UDDS	Fuel consumption (L.100km^{-1})	1.78	5.6	5.7

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

	Battery life (cycle)	46.77	1.24×10^3	1.82×10^3
NEDC	Fuel consumption (L.100km^{-1})	2	5.43	5.54
	Battery life (cycle)	35.48	3.1×10^4	4.2×10^4

To further corroborate the validity of these findings, the method is evaluated on two extra cycles that were purposefully constructed to share similarities with HWFET. The velocity profiles of these cycles are rendered in **Figure 9**, and a concise summary of the corresponding numerical outcomes is tabulated in **Table 4**. The results demonstrate that the Q-learning method preserves its effectiveness when faced with a shift in the driving cycle, consistently yielding gains in battery longevity coupled with reductions in fuel consumption.

Table 4. Compared results for the cycles similar to HWFET.

Cycle test	Parameter/Unit	Variation (%)	Rule-based ADVISOR	Q-learning approach
HWFET	Fuel Consumption (L/100 km-^{1})	-1.25	4	3.95
	Battery life (cycle)	62.5	0.88×10^4	1.43×10^4
Cycle 1	Fuel Consumption (L/100 km-^{1})	0	4.3	4.3
	Battery life (cycle)	64	0.62×10^4	1.02×10^4
Cycle 2	Fuel Consumption (L/100 km-^{1})	-2.6	3.8	3.7
	Battery life (cycle)	46.77	1.27×10^4	1.78×10^4

Conclusion

The present study has put forward a reinforcement learning-centric energy management methodology intended for series-parallel hybrid vehicles. The developed strategy simultaneously optimizes fuel economy and battery service life, operating without any requirement for prior awareness of the driving cycle or a high-fidelity mathematical model of the vehicle. The optimization is driven by a reward function crafted to suppress fuel consumption while promoting maximal battery longevity. Simulation experiments were carried out within a MATLAB and ADVISOR co-simulation environment covering the HWFET, IM240, UDDS, and NEDC driving cycles, and the generated outcomes were compared against those of a conventional rule-based strategy. According to the comparative analysis, fuel consumption for the HWFET and IM240 cycles fell by 1.25% and 0.68%, respectively, whereas battery life rose by 65% and 50%. Battery life also increased by 47% for the UDDS cycle and by 36% for the NEDC cycle. To verify robustness, the method was applied to two alternative driving cycles, with the results affirming the characteristic self-improvement capability of reinforcement learning.

Acknowledgements

None

Conflict of interest

None

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

Financial support

None

Ethics statement

None

Received: 07 Mar 2024 Revised: 13 May 2024 Accepted: 30 May 2024
Published online: 15 November 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at https://creativecommons.org/licenses/by-nc-sa/4.0/.

References

1. Rajaei Salmasi F. Control strategies for hybrid electric vehicles: evolution, classification, comparison, and future trends. IEEE Trans Veh Technol. 2007;56(5):2393-404.

2. Tian G, Shi J, Sun L, Long X, Guo B. Dynamic changes in the energy–carbon performance of Chinese transportation sector: a meta-frontier non-radial directional distance function approach. Nat Hazards. 2017;89(2):585-607.

3. 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;115:174-89.

4. Sun C, Sun F, He H. Investigating adaptive-ECMS with velocity forecast ability for hybrid electric vehicles. Appl Energy. 2017;185:1644-53.

5. Liu G, Lu L, Fu H, Hua J, Li J, Ouyang M, et al. A comparative study of equivalent circuit models and enhanced equivalent circuit models of lithium-ion batteries with different model structures. In: IEEE Conference and Expo Transportation Electrification Asia-Pacific (ITEC Asia-Pacific); 2014.

6. Chaoui H, Ibe-Ekeocha CI, Gualous H. Aging prediction and state of charge estimation of a LiFePO4 battery using input time-delayed neural networks. Electr Power Syst Res. 2017;146:189-97.

7. Wu L, Pang H, Geng Y, Liu X, Liu J, Liu K. Low-complexity state of charge and anode potential prediction for lithium-ion batteries using a simplified electrochemical model-based observer under variable load condition. Int J Energy Res. 2022;46(12):16845-62.

8. Zhang ZL, Cheng X, Lu ZY, Gu DJ. SOC estimation of lithium-ion batteries with AEKF and wavelet transform matrix. IEEE Trans Power Electron. 2017;32(10):7626-34.

9. Wang S, Takyi-Aninakwa P, Jin S, Yu C, Fernandez C, Stroe DI. An improved feedforward-long short-term memory modeling method for the whole-life-cycle state of charge prediction of lithium-ion batteries considering current-voltage-temperature variation. J Energy. 2022;254:124282.

10. Tang XZ, Yang SJ, Xia HC. Control strategy of hybrid electric vehicles based on driving style identification. Adv Mater Res. 2014;945:1587-96.

11. Zhou ML, Lu DK, Li WM, Xu HF. Optimized fuzzy logic control strategy for parallel hybrid electric vehicle based on genetic algorithm. Appl Mech Mater. 2013;274:345-9.

12. Montazeri M, Mahmoodi M. Development a new power management strategy for power split hybrid electric vehicles. Transp Res Part D Transp Environ. 2015;73:79-96.

13. Dokuyucu HI, Cakmakci M. Concurrent design of energy management and vehicle traction supervisory control algorithms for parallel hybrid electric vehicles. IEEE Trans Veh Technol. 2015;65(2):555-65.

14. Liu Y, Liu J, Qin D, Li G, Chen Z, Zhang Y. Online energy management strategy of fuel cell hybrid electric vehicles based on rule learning. J Clean Prod. 2020;260:121017.

15. Wang F, Xia J, Xu X, Cai Y, Ni S, Que H. Torsional vibration-considered energy management strategy for power-split hybrid electric vehicles. J Clean Prod. 2021;296:126399.

16. Li J, Liu Y, Zhang Y, Lei Z, Chen Z, Li G. Data-driven based eco-driving control for plug-in hybrid electric vehicles. J Power Sources. 2021;498:229916.

17. Qi C, Zhu Y, Song C, Cao J, Xiao F, Zhang X, et al. Self-supervised reinforcement learning-based energy management for a hybrid electric vehicle. J Power Sources. 2021;514:230584.

18. Silver D, Lever G, Heess N, Degris T, Wierstra D, Riedmiller M. Deterministic policy gradient algorithms. In: Proceedings of the 31st International Conference on Machine Learning (ICML); 2014. p. 387-95.

19. Lin X, Wang Y, Bogdan P, Chang N, Pedram M. Reinforcement learning based power management for hybrid electric vehicles. In: 2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD); 2014. p. 33-8. IEEE.

20. Duan Y, Schulman J, Chen X, Bartlett PL, Sutskever I, Abbeel P. RL^2: Fast reinforcement learning via slow reinforcement learning. arXiv preprint arXiv:1611.02779. 2016.

21. Schulman J, Levine S, Moritz P, Jordan MI, Abbeel P. Trust region policy optimization. arXiv preprint arXiv:1502.05477. 2015.

22. Liu TT, Hu X, Li SE, Cao D. Reinforcement learning optimized look-ahead energy management of a parallel hybrid electric vehicle. IEEE/ASME Trans Mechatron. 2017;22(4):1497-507.

23. 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(2):187.

24. Qi C, Zhu Y, Song C, Yan G, Xiao F, Zhang X, et al. Hierarchical reinforcement learning based energy management strategy for hybrid electric vehicle. Energy. 2022;238:121703.

25. Kim N, Rousseau A, Rask E. Autonomie model validation with test data for 2010 Toyota Prius. Fuel. 2012;48(45):46.

26. Liu J, Peng H, Filipi Z. Modeling and analysis of the Toyota hybrid system. TIC. 2005;200(3).

27. Zhang X. Design of power split hybrid powertrains with multiple planetary gears and clutches [dissertation]. 2015.

28. Drouilhet S, Johnson B. A battery life prediction method for hybrid power applications. In: 35th Aerospace Sciences Meeting and Exhibit; 1997. p. 948.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press

29. US Environmental Protection Agency. Dynamometer drive schedules. Available from: https://www.epa.gov/vehicle-and-fuel-emissions-testing/dynamometer-drive-schedules.

30. Sutton RS, Barto AG. Introduction to reinforcement learning. 1998. p. 551283.

31. Kermani S, Delprat S, Guerra TM, Trigui R. Predictive control for HEV energy management: experimental results. In: 2009 IEEE Vehicle Power and Propulsion Conference; 2009. p. 364-9. IEEE.

© 2025 The Author(s). | Published by Institute of Smart Energy Systems Press