

Article

Deep Reinforcement Learning-Based Real-Time Joint Optimal Power Split for Battery–Ultracapacitor–Fuel Cell Hybrid Electric Vehicles

Daniel Kim , Seokjoon Hong , Shengmin Cui  and Inwhae Joe *

Department of Computer Science, Hanyang University, Seoul 04763, Korea; danielkim96@hanyang.ac.kr (D.K.); daniel379@hanyang.ac.kr (S.H.); shengmincui@hanyang.ac.kr (S.C.)

* Correspondence: iwjoe@hanyang.ac.kr; Tel.: +82-02-2220-1088

Abstract: Hybrid energy storage systems for hybrid electric vehicles (HEVs) consisting of multiple complementary energy sources are becoming increasingly popular as they reduce the risk of running out of electricity and increase the overall lifetime of the battery. However, designing an efficient power split optimization algorithm for HEVs is a challenging task due to their complex structure. Thus, in this paper, we propose a model that jointly learns the optimal power split for a battery/ultracapacitor/fuel cell HEV. Concerning the mechanical system of the HEV, two propulsion machines with complementary operation characteristics are employed to achieve higher efficiency. Additionally, to train and evaluate the model, standard driving cycles and real driving cycles are employed as input to the mechanical system. Then, given the inputs, a temporal attention long short-term memory model predicts the next time step velocity, and through that velocity, the predicted load power and its corresponding optimal power split is computed by a soft actor–critic deep reinforcement learning model whose training phase is aided by shaped reward functions. In contrast to global optimization techniques, the local velocity and load power prediction without future knowledge of the driving cycle is a step toward real-time optimal energy management. The experimental results show that the proposed method is robust to different initial states of charge values, better allocates the power to the energy sources and thus better manages the state of charge of the battery and the ultracapacitor. Additionally, the use of two motors significantly increases the efficiency of the system, and the prediction step is shown to be a reliable way to plan the HESS power split in advance.

Keywords: hybrid electric vehicle; deep reinforcement learning; optimal power split; real-time

Citation: Kim, D.; Hong, S.; Cui, S.; Joe, I. Deep Reinforcement Learning-Based Real-Time Joint Optimal Power Split for Battery–Ultracapacitor–Fuel Cell Hybrid Electric Vehicles. *Electronics* **2022**, *11*, 1850. <https://doi.org/10.3390/electronics11121850>

Academic Editors: Erik Scholtz and Minh-Khai Nguyen

Received: 1 May 2022

Accepted: 8 June 2022

Published: 10 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The basic operating principle of internal combustion engine (ICE) vehicles involves transforming energy from fossil fuels into thermal energy. In this combustion, the gases generated are released into the atmosphere with a negative impact on the environment and human health. In particular, the transportation sector contributed 29% of greenhouse gas emissions (GHG) to total United States GHG emissions in 2019 [1]. Furthermore, there is a growing concern regarding the scarcity of fossil fuels and the need to implement a sustainable economy based on renewable energy sources. The Sustainable Development Goals (SDGs) established in 2015 by the United Nations General Assembly have sustainable energy at its core. The seventh goal specifically emphasizes the need for renewable energy sources [2]. In light of the growing concern of society with environmental issues, the development of battery electric vehicles (BEVs) is a step toward the fulfillment of the seventh SDG.

The engine is the primary distinction between a traditional internal combustion engine (ICE) car and an electric vehicle (EV). The latter is powered by an electric motor that works by transforming chemical energy stored in rechargeable batteries into electrical energy. The

generated electricity is then transformed into mechanical and kinetic energy. These new models of vehicles also have multiple advantages when compared to ICE ones. EVs are efficient, do not emit tailpipe pollutants, are quiet and produce little noise pollution. In addition, recharging an EV is cheaper than refueling an ICE car [3,4].

Electric vehicles include battery electric vehicles (BEVs), hybrid electric vehicles (HEVs), plug-in hybrid electric vehicles (PHEVs), and fuel cell hybrid electric vehicles (FCHEVs), depending on the energy source for the vehicle [5,6].

Fuel cells are considered an ideal energy source for electric vehicles because of their high efficiency and zero emission [7]. However, there is very little charging infrastructure for fuel cells compared to charging stations for EVs and gas stations for ICEs. Moreover, the cost of charging a fuel cell is much higher than that of charging an electric vehicle. Because of this shortcoming of fuel cells, FCHEVs require additional energy sources.

Batteries are the most widely used energy source in EVs because they have high energy density [8,9]. However, they have disadvantages such as long charging times, high prices, and high temperature sensitivity. Therefore, recently, solid-state type batteries with better performance and shorter charging time than conventional batteries have been continuously studied [10,11].

Ultracapacitors (UC) have high power density and very long lifetimes and are not affected by temperature; so they are suitable for devices with high peak currents [12]. However, low energy density, the need for voltage balancing and high self-discharging are major drawbacks. One of the most promising solutions proposed for FCHEVs is a hybrid energy storage system (HESS) consisting of three energy sources: a fuel cell (FC), a battery, and an ultracapacitor [13].

Since the HESS structure for EVs is diverse, a different energy management system (EMS) control strategy is required according to the HESS structure [14]. One of the most common HESS structures consists of a battery and a UC [15]. Among other existing studies, there are studies that reviewed a HESS consisting of an FC and a battery [16] and an FC and a UC [17]. As a more complex structure, there have been studies on hybrid energy storage systems that include all three of them: a battery, a UC, and an FC [18].

In addition, there are some studies [19,20] of EMS considering a complex HESS structure in which not only multiple energy storage units but also more than one propulsion machine are connected. To improve power performance and propulsion efficiency, two propulsion machines with complementary torque–speed characteristics can be used in the EV powertrain.

The energy management systems of electric vehicles with hybrid energy storage systems are largely classified into two types: rule-based and optimization-based [21]. Rule-based EMSs are divided into deterministic rule-based EMSs and fuzzy rule-based EMSs according to the characteristics of the rules used. In addition, optimization-based EMSs can be classified into global optimization and real-time optimization EMSs according to the level of information of the driving condition. Global optimization-based methods have the advantage of finding the global optimum of the problem, but they demand previous knowledge of the entire driving cycle, which would be unfeasible in a real-time optimization scenario.

Many previous studies focused on convex optimization techniques [19,20,22,23]. However, convex optimization of complex systems is very difficult due to the excessive number of parameters to consider and the difficulty of linearizing all systems. Therefore, a neural network (NN)-based machine learning algorithm was proposed by [20] to solve the multi-purpose energy problem for a dual-motor battery/FC/UC. The control was performed through two steps. First, the load power was divided between the two complementary motors, and then the load power was split between the battery pack, the FC and the UC in the HESS. For the HESS power split, a convex optimization technique was used to generate optimal target values which became the training data for the NN. When employing convex optimization, the results showed that the battery life was extended by about 5 years in contrast to the NN model which extended the battery life by about 92.5% of 5 years.

Another technique that has been emerging in the area of HEV EMSs is deep reinforcement learning (DRL) [24–28]. They explore different DRL techniques such as deep Q-learning (DQN), soft actor–critic (SAC) and deep deterministic policy gradient (DDPG) for different HEV topologies. Particularly, by applying reinforcement learning using real driving data, energy savings of about 16% compared to the existing binary control strategy were confirmed [28]. The energy efficiency was improved by looking at the existing driving cycle of one’s vehicle and dividing the power output level of the ICE into 24 levels through reinforcement learning, instead of just following the rules set by the EMS.

However, we found only one work that addresses the dual-motor battery/UC/FC HEV EMS control strategy, due to their complex structure [20], which makes use of convex optimization and machine learning techniques, not reinforcement learning. Additionally, even though the authors in [27] develop a reinforcement learning-based model for an FC/battery/UC HEV, they use a deep-Q network model and do not explore the use of multiple motors nor the forecast of the future load power. Thus, in this paper, a real-time deep reinforcement learning (SAC)-based EMS control strategy for a HESS of a vehicle that has three energy sources, battery, FC and UC, and two complementary motors is proposed. First, a method for real-time velocity and load power prediction using SDCs and RDCs is proposed. This prediction step allows the EMS to better plan the use of resources of the HEV and prevent it from running out of electricity. Given the predicted load power, the SAC model is then used to efficiently distribute energy in the HESS. Through the experimental results, it was confirmed that compared to traditional rule-based methods the proposed method can better allocate the energy sources power and thus achieve good FC efficiencies and better manage the SOCs of the battery and the ultracapacitor. The model is also shown to be robust, as it can handle different initial values of SOC while satisfying all the constraints of the system. Moreover, the addition of reward shaping to the training phase of the SAC agent accelerated its convergence. Additionally, it can be seen that the use of two complementary motors leads to a big improvement in the efficiency of the vehicle compared to a single-motor architecture. Finally, the results of the prediction step show that the method is reliable to forecast the future speeds and load power and, therefore, aid the model in allocating the resources of the HEV.

2. Preliminaries

2.1. Overall Procedure for a Reinforcement Learning-Based Energy Management Strategy (EMS)

Figure 1 illustrates the proposed method. The inputs of the neural network are the five past velocity and acceleration values of either standard driving cycles (SDCs) or real driving cycles (RDC). Given those values, the neural network predicts the future velocity and acceleration for one time step ahead, which will be used to compute the load powers of the two electrical motors. The load powers of these motors will be sent into the EMS, which will distribute the load power between the battery, the fuel cell, and the ultracapacitor using deep reinforcement learning (DRL).

2.2. Input Data

For the input, both SDCs and RDCs were used. Particularly, the worldwide harmonized light vehicles test cycles (WLTC) and the urban dynamometer driving schedule (UDDS) were used. Despite their ease of use, they do not represent the reality well enough; so they were mainly used for model validation and performance evaluation purposes. On the other hand, RDCs overcome these limitations because of the higher degree of complexity associated to them. There are personal elements, such as the driver’s behavior, which is hard to model because of its time-varying nature. There are also factors that do not depend on the driver, such as the weather conditions (snowstorm, heavy rain, flooding), traffic signals and traffic conditions (construction and accidents may slow down the traffic). The RDCs data, collected by an On-Board Diagnostics (OBD)-II dongle connected to one of the author’s vehicles, are the same as the data used in Hong et al. [29]. The four RDCs employed in this paper are shown in Figure 2.

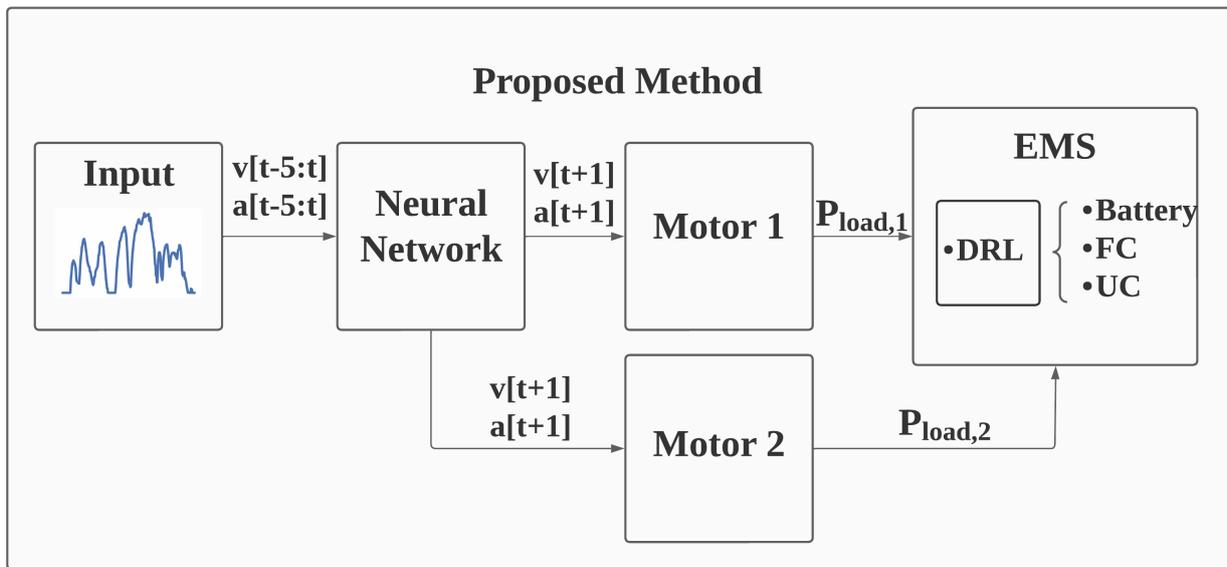


Figure 1. Procedure of reinforcement learning-based EMS.

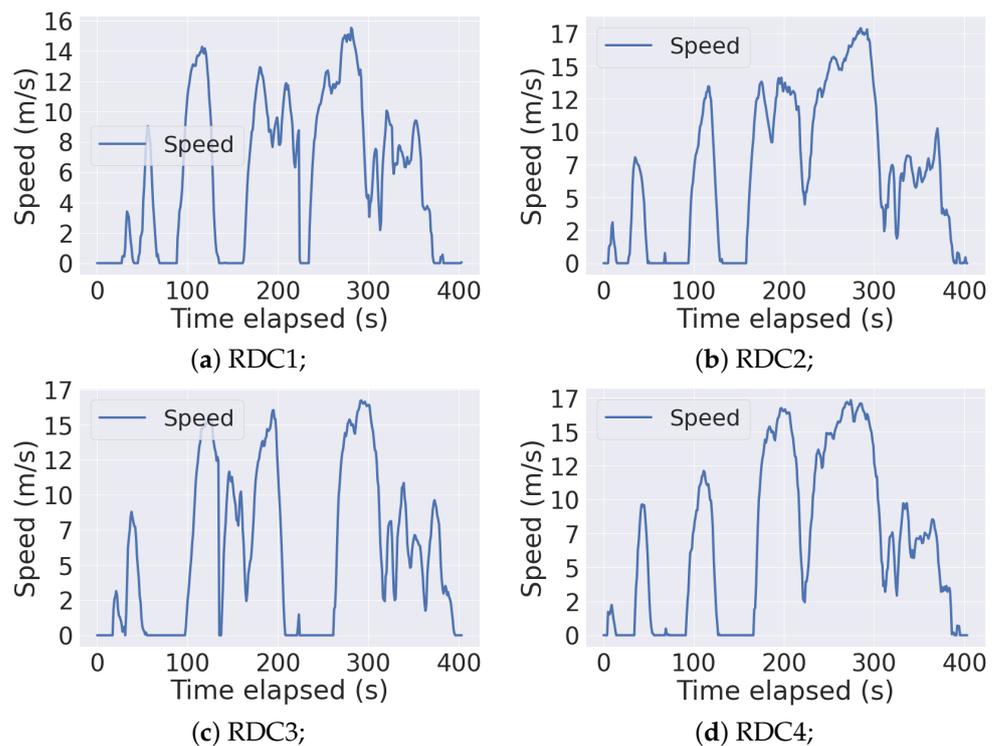


Figure 2. Speed and acceleration over time of the different time series collected by Hong et al. [29].

2.3. Neural Network for Velocity Forecast

The problem that the neural network is trying to solve can be modeled as a univariate time series; that is, there is only one time-varying variable, which is the velocity of the vehicle. Despite the existing challenges related to time series forecast, it is an important field of study because knowing the future allow us to better plan short-term and long-term goals. Particularly, in our work, we predict the velocity to optimally allocate the available resources so that the energy storage components may achieve maximum efficiency, their lifetime may be maximized and the vehicle may not run out of electricity during the trip.

Based on the previous work of Hong et al. [29], which used the same RDC dataset as this present paper and evaluated different time series and deep learning models for velocity

prediction (seasonal auto regressive integrated moving average (SARIMA), recurrent neural network (RNN), gated recurrent unit (GRU) and long short-term memory (LSTM)), we decided to use the attention-based LSTM model with step size $T = 5$ and hidden state size $m = 32$ proposed by the authors, as it outperformed the other models that were tested.

2.4. Vehicle System Overview

The vehicle is mainly composed of the HESS and the mechanical components shown in Figure 3. The HESS is composed of three energy sources (battery/ultracapacitor/fuel cell), and the HEV has two complementary propulsion machines.

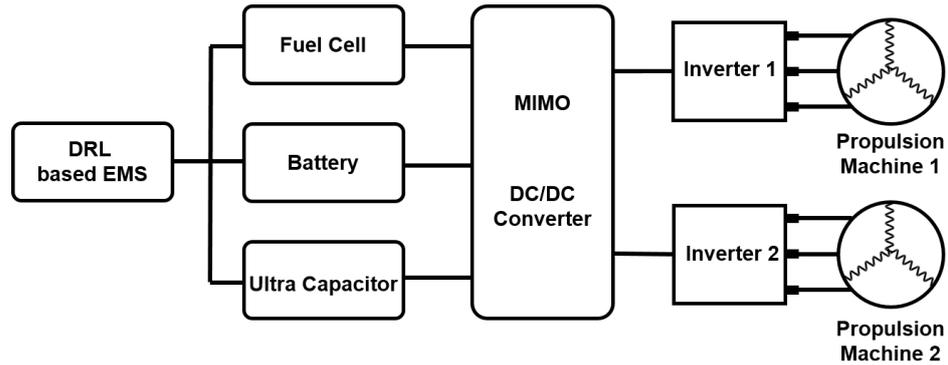


Figure 3. HESS architecture.

2.5. Power Split Optimization for the Propulsion Machines

The mechanical model of the vehicle may be computed by Equations (1)–(14), and its parameters are given in Table 1. The acceleration values needed for the equations come from Section 2.3 by differentiating the forecast velocity values. Such acceleration values are then used to compute the total force F_t acting on the vehicle, given by (1), consisting of the rolling resistance F_{rr} , the aerodynamic drag F_{ad} , the grading resistance F_{gr} and the F_{la} . Given the total force, the load power and the power loss for each motor may be computed according to the motor efficiencies. Then, combining all the Equations (1)–(14) with the power split optimization procedure proposed by [19], the optimal power split and losses for the two motors may be computed.

In this paper, the motors chosen are the Toyota Camry MG1 [30] and the UQM PowerPhase 125 [31]. Their parameters are shown in Table 2 and their efficiency curves may be seen in Figure 4. Finally, Figure 5 shows the output of the method (load power and losses) for the two motors for the four RDCs shown in Figure 2.

$$F_t = F_{rr} + F_{ad} + F_{gr} + F_{la} \tag{1}$$

$$F_{rr} = C_r mg \cos(\alpha) \tag{2}$$

$$C_r = a_0 + a_1 \cdot v + a_2 \cdot v^2 + a_3 \cdot v^3 + a_4 \cdot v^4 + a_5 \cdot v^5 \tag{3}$$

$$F_{ad} = \frac{1}{2} \rho AC_d v^2 \tag{4}$$

$$F_{gr} = mg \sin(\alpha) \tag{5}$$

$$F_{la} = \delta ma \tag{6}$$

$$T_t = F_t \times r_w \tag{7}$$

$$P_t = F_t \times v \tag{8}$$

$$\omega_w = \frac{v}{r_w} \tag{9}$$

$$\tau = \frac{r_{wh} \cdot F_t}{G_r} \tag{10}$$

$$\omega_m = \frac{v}{r_{wh}} G_r \tag{11}$$

$$P_{Load} = \tau \omega_m \tag{12}$$

$$P_{loss,i}(t) = \left\{ \begin{array}{ll} \omega(t) \cdot \tau_i(t) \cdot (1 - \eta_i(t)) & \text{if } \tau_i(t) \geq 0 \\ \omega(t) \cdot \tau_i(t) \cdot \frac{(1 - \eta_i(t))}{\eta_i(t)} & \text{if } \tau_i(t) \leq 0 \end{array} \right\}, \text{ i} = \{1,2\} \text{ is the motor index} \tag{13}$$

$$\text{Efficiency}(\eta) = \text{lookup2d}(\text{Torque(Nm)}, \text{Speed(rpm)}) \tag{14}$$

Table 1. Mechanical model parameters.

Parameters	Description	Values
m	Mass of the vehicle (kg)	1650
r_{wh}	Wheel radius (m)	0.33602
g	Gravity acceleration (m/s ²)	9.81
ρ	Air density	1.2
A	Front area of vehicle (m ²)	2.12344
α	Angle of driving surface (rad)	0
a_0	Rolling resistance coefficient	$8.8 \cdot 10^{-3}$
a_1	Rolling resistance coefficient	$-6.42 \cdot 10^{-5}$
a_2	Rolling resistance coefficient	$9.27 \cdot 10^{-6}$
a_3	Rolling resistance coefficient	$-3.3 \cdot 10^{-7}$
a_4	Rolling resistance coefficient	$6.68 \cdot 10^{-11}$
a_5	Rolling resistance coefficient	$-4.46 \cdot 10^{-11}$
C_d	Aerodynamic drag coefficient	0.24
G_r	Gearbox ratio	5.5

Table 2. Parameters of the motors.

Motor	Parameters	Values
Toyota Camry MG1	Motor peak power rating (kW)	105
	Motor peak torque rating (Nm)	270
	Top rotational speed (rpm)	14,000
	Motor mass (kg)	41.7
UQM PowerPhase 125	Motor peak power rating (kW)	125
	Motor peak torque rating (Nm)	300
	Top rotational speed (rpm)	8000
	Motor mass (kg)	41

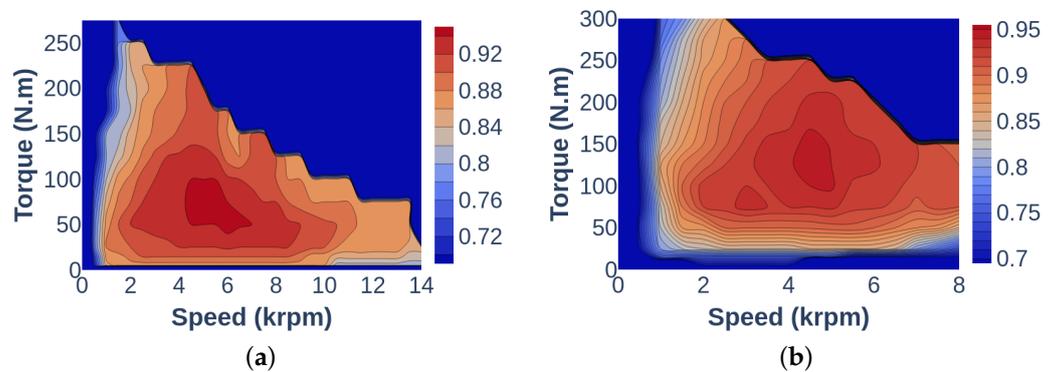


Figure 4. Efficiency maps: (a) Motor 1; (b) Motor 2. Adapted from [30,31].

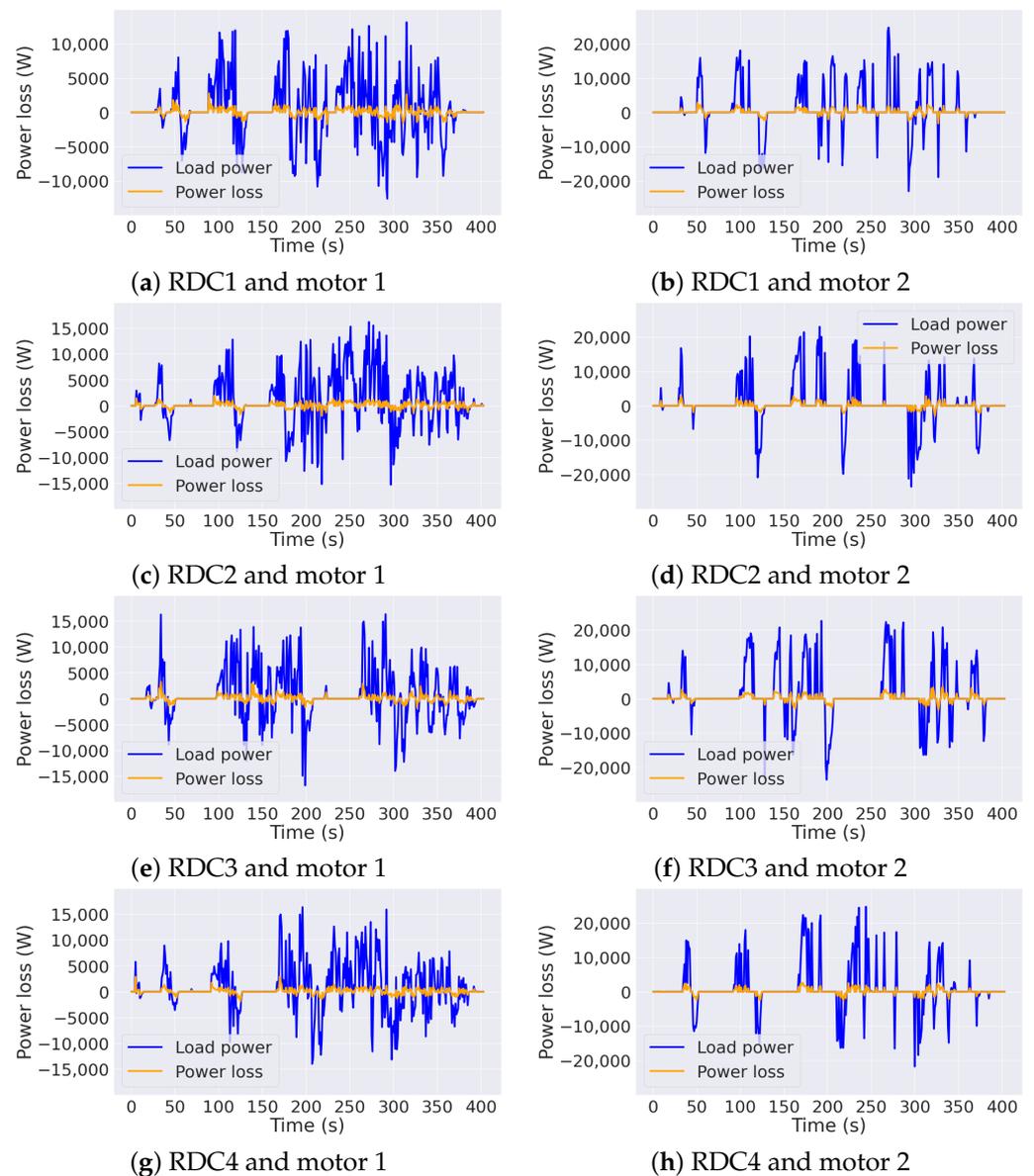


Figure 5. Load power and losses for the two motors and the four RDCs: (a) load power and loss for motor 1 and RDC1; (b) load power and loss for motor 2 and RDC1; (c) load power and loss for motor 1 and RDC2; (d) load power and loss for motor 2 and RDC2; (e) load power and loss for motor 1 and RDC3; (f) load power and loss for motor 2 and RDC3; (g) load power and loss for motor 1 and RDC4; (h) load power and loss for motor 2 and RDC4.

2.6. Electrical Model of the Hybrid Energy Storage System (HESS)

The battery and the ultracapacitor packs adopted are based on the ones proposed by [19]. They are composed of, respectively, multiple K2 High Capacity Lithium iron phosphate (LiFePO₄) 22650P battery cells [32] and multiple Maxwell BCAP1500 ultracapacitors [33]. Their basic models are given, respectively, by Figures 6 and 7; the parameters are given by Tables 3 and 4, and the equations used to compute the parameters are given by Equations (15)–(22) and (23)–(29).

Finally, the fuel cell (FC) and the DC/DC converter were modelled based on the system proposed in [20]. The DC has a maximum output of 35kW, the parameters for the FC are given in Table 5 and the FC stack's efficiency and power loss may be computed according to Equations (30) and (31).

Two inverters for two propulsion machines were modelled based on the datasheet [34]. These inverters are insulated gate bipolar transistors (IGBTs) with a based pulse width modulated (PWM) inverter. We assumed that inverter efficiency is 97%.

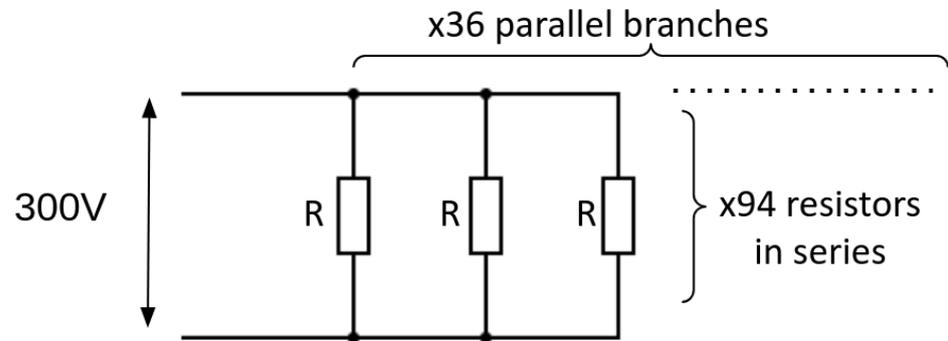


Figure 6. Basic model of the battery.

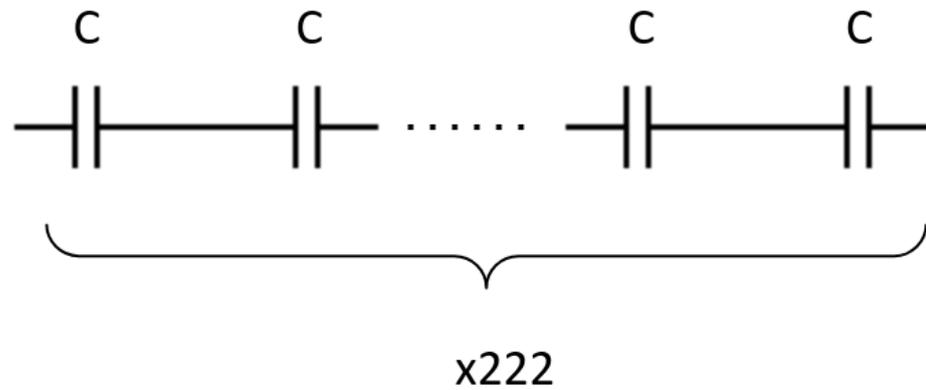


Figure 7. Basic model of the ultracapacitor.

Table 3. Battery parameters.

Parameters	Description	Values
R	Internal resistance	9 mΩ
N_s	Number of serial cells	94
N_p	Number of parallel branches	36
V_T	Battery voltage (V)	300
$R_n, n = 1, 2, \dots, N_s$	Equivalent resistance of serial cells	846 mΩ
R_T	Total equivalent resistance	23.5 mΩ
V_{cell}	Cell voltage (V)	3.2
E_{cell}	Maximum energy stored in a cell (Wh)	8.32
C_{cell}	Capacity of a cell (Ah)	2.6
E_T	Total energy stored (kWh)	28.2
C_T	Total capacity (Ah)	93.6
SOC_{bat}	SOC of the battery	-
i_{bat}	Current of the battery	-

$$V_T = N_s \cdot V_{cell} \tag{15}$$

$$R_n = N_s \cdot R; n = 1, \dots, 94 \tag{16}$$

$$R_T = \frac{R_n}{N_p} \tag{17}$$

$$E_{cell} = V_{cell} \cdot C_{cell} \tag{18}$$

$$C_T = C_{cell} \cdot N_p \tag{19}$$

$$E_T = E_{cell} \cdot N_p \cdot N_s \tag{20}$$

$$i_{bat} = \frac{V_T - \sqrt{V_T^2 - 4R_T}}{2R_T} \tag{21}$$

$$SoC_{bat} = SoC_{bat,prev} + \frac{\beta \int i_{bat} dt}{C_T} \tag{22}$$

Table 4. Ultracapacitor parameters.

Parameters	Description	Values
R	Internal resistance	0.47 mΩ
N_s	Number of serial cells	222
V_T	Battery voltage (V)	600
R_T	Total equivalent resistance	103.34 mΩ
C_{cell}	Capacitance of a cell (F)	1500
V_{cell}	Cell voltage (V)	2.7
E_{cell}	Maximum energy stored in a cell (Wh)	1.52
E_T	Total energy stored (kWh)	0.338
C_T	Total capacitance (F)	6.76
SOC_{UC}	SOC of the battery	-
i_{UC}	Current of the battery	-

$$V_T = N_s \cdot V_{cell} \tag{23}$$

$$R_T = N_s \cdot R \tag{24}$$

$$C_T = \frac{C_{cell}}{N_s} \tag{25}$$

$$E_{cell} = 0.5 \cdot C_{cell} \cdot V_{cell}^2 \tag{26}$$

$$E_T = E_{cell} \cdot N_s \tag{27}$$

$$i_{UC} = \frac{V_T - \sqrt{V_T^2 - 4R_T}}{2R_T} \tag{28}$$

$$SoC_{UC} = SoC_{UC,prev} + \frac{\beta \int i_{UC} dt}{C_T} \tag{29}$$

Table 5. FC and DC/DC converter parameters.

Parameters	Description	Values
$P_{nom,fc}$	FC nominal power	8 kW
a_1	Slope (if $\frac{P_{fc}}{P_{fc,nom}} \leq 0.15$)	2
b_1	Intercept (if $\frac{P_{fc}}{P_{fc,nom}} \leq 0.15$)	0.4
a_2	Slope (if $\frac{P_{fc}}{P_{fc,nom}} \geq 0.15$)	-0.040625
b_2	Intercept (if $\frac{P_{fc}}{P_{fc,nom}} \geq 0.15$)	0.620625
η_{fc}	FC efficiency	Equation (30)
$P_{fc,loss}$	Loss in the FC	Equation (31)
η_{inv}	Inverter efficiency	0.97
$P_{inv,loss}$	Loss in the inverter	Equation (32)

$$\eta_{fc} = a \cdot \frac{P_{fc}}{P_{fc,nom}} + b \tag{30}$$

$$P_{loss,fc} = P_{fc} \cdot (1 - \eta_{fc}) \tag{31}$$

$$P_{loss,inv} = P_{inv} \cdot (1 - \eta_{inv}) \quad (32)$$

$$P_{PM} = P_{inv} \cdot \eta_{inv} \quad (33)$$

3. Soft Actor–Critic (SAC) for HESS Power Split Optimization

3.1. Problem Overview

Reinforcement learning (RL) at its core involves an agent trying to learn decision making and control by interacting with an environment. Such interaction happens through actions taken by the agent, which can modify the environment and cause a state transition. As every action has an impact on the environment, each yields rewards to encourage good actions and discourage bad ones. The choice of action for each state depends on the policy of the agent, which maps the states to the probabilities of choosing each action given a specific state.

Therefore, the problem of power split optimization for the presented HESS can be modeled as a deep reinforcement learning problem. As shown in the next sections, the agent of our proposed model must learn the best policy possible to maximize the rewards of the system as a whole. Moreover, as it is not possible to map all the possible state–action pairs of the environment to their respective rewards, we use deep neural networks (DNN) to approximate them. Finally, all the code related to the deep reinforcement learning-based model was implemented using the Python programming language and the TensorFlow’s TF-Agents framework.

3.2. Agent, Environment, Action and State

In our case, we have the HESS, which is the agent inside the HEV, which is the environment. The HEV is supposed to move for a specified amount of time according to a driving cycle while varying velocity, optimizing the fuel cell’s efficiency and minimizing the battery power magnitude and variation. However, initially the HEV does not know how to distribute the load power computed in Section 2.5 between the three different energy storage systems (battery, ultracapacitor and fuel cell). The task of power split is performed by the HESS, which must decide among infinite possible combinations of the optimal battery, ultracapacitor and fuel cell powers. Each of the combinations chosen is an action performed by the agent, which leads to a state transition. Thus, in this paper, the state and the action spaces are given, respectively, by $S = \{P_{load}; SOC_{bat}; SOC_{uc}\}$ and $A = \{P_{bat,\%}, P_{UC,\%}, P_{FC,\%}\}$. The actions are not absolute power values but continuous values ranging from zero to one, which are transformed into percentages summing up to one through the softmax function. For instance, if $P_{load} = 10$ kW and $\{P_{bat,\%}, P_{UC,\%}, P_{FC,\%}\} = \{0.8, 0.5, 0.2\}$, the softmax function outputs approximately $\{0.44, 0.32, 0.24\}$, which would represent in absolute power values $\{P_{bat}, P_{UC}, P_{FC}\} = \{4.4$ kW, 3.2 kW, 2.4 kW}.

3.3. Rewards and Penalties

The learning process of the agent is driven by rewards and penalties depending on how good the agent’s actions are. In our case, the agent should ideally satisfy all the following criteria (35)–(41), which are based on [20]. The parameters related to the constraints may be found in Table 6, and the objective function to be optimized is shown in Equation (34), where the coefficients (a) to (f) are positive penalty coefficients.

$$g = a \cdot P_{bat} + b \cdot P_{UC} + c \cdot P_{FC} + d \cdot SOC_{bat}^2 + e \cdot SOC_{UC}^2 + f \cdot \eta_{FC}^2 \quad (34)$$

$$P_{min,bat} \leq P_{bat}(t) \leq P_{max,bat} \quad (35)$$

$$P_{min,bat_diff} \leq P_{bat}(t) - P_{bat}(t-1) \leq P_{max,bat_diff} \quad (36)$$

$$P_{min,DC} \leq P_{DC}(t) \leq P_{max,DC} \quad (37)$$

$$P_{min,FC} \leq P_{FC}(t) \leq P_{max,FC} \quad (38)$$

$$P_{min,UC} \leq P_{UC}(t) \leq P_{max,UC} \quad (39)$$

$$P_{bat}(t) + P_{FC}(t) - P_{DC}(t) = P_{M_1}(t) \quad (40)$$

$$P_{UC}(t) + P_{DC}(t) = P_{M_2}(t) \quad (41)$$

Table 6. Optimization constraints parameters.

Parameters	Description	Values
$P_{min,bat}$	Minimum battery power	−30 kW
$P_{max,bat}$	Maximum battery power	30 kW
$SOC_{min,bat}$	Minimum battery SOC	0.2
$SOC_{max,bat}$	Maximum battery SOC	0.9
P_{min,bat_diff}	Minimum battery power difference	−9 kW
P_{max,bat_diff}	Maximum battery power difference	9 kW
$P_{min,DC}$	Minimum DC/DC converter power	−35 kW
$P_{max,DC}$	Maximum DC/DC converter power	35 kW
$P_{min,FC}$	Minimum FC power	0
$P_{max,FC}$	Maximum FC power	8 kW
$P_{min,UC}$	Minimum ultracapacitor power	−30 kW
$P_{max,UC}$	Maximum ultracapacitor power	30 kW
$SOC_{min,UC}$	Minimum ultracapacitor SOC	0.5
$SOC_{max,UC}$	Maximum ultracapacitor power	1

Considering the constraints (35)–(41), terminal states were designed. The terminal states and actions would be a SOC_{bat} , SOC_{uc} , P_{bat} , P_{uc} and P_{FC} smaller or greater than the minimum or maximum values allowed according to Table 6. Thus, whenever the vehicle reached an illegal state or tried to perform an illegal action, the environment was reset, and the agent received a penalty proportional to the length of the driving cycle. As the length of the driving cycles used were all less than 2000, the agent receives a penalty of 1000, and the environment is reset. On the other hand, if the states and the actions chosen are valid, the vehicle receives a reward according to $r = SOC_{bat}^2 + SOC_{uc}^2 + \eta_{fc}^2 + g(x)$, where $g(x) = 1 - \left(\frac{d}{d_{max}}\right)^p$ is the output of the shaped reward function as explained below.

Reward shaping is a technique that involves changing the structure of a sparse reward function to offer more regular feedback to the agent [35] and thus accelerate the learning process. Figure 8 shows an example of a sparse and a shaped reward function.

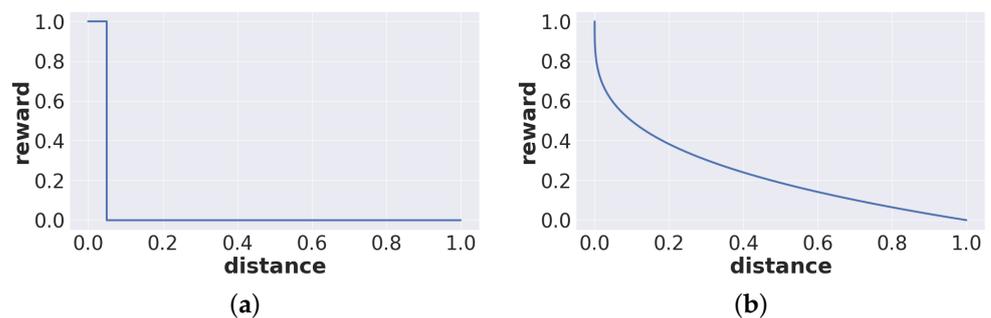


Figure 8. Reward shaping: (a) sparse reward function; (b) shaped reward function.

In our paper, we designed three different reward functions and one penalty function. Equations (42)–(44) show the reward functions, and Equation (45) represents the penalty function for different values of d and d_{max} of the equation $g(x) = 1 - \left(\frac{d}{d_{max}}\right)^p$. The value of p was set to 0.4 after performing hyperparameter tuning.

All the four functions were activated depending on the SOC of the ultracapacitor to encourage higher or lower power from the three energy sources. Their thresholds and descriptions are shown in Table 7.

$$g_1(x) = 1 - \left(\frac{P_{max,UC} - P_{UC}}{P_{max,UC}} \right)^{0.4} \quad (42)$$

$$g_2(x) = 1 - \left(\frac{P_{max,FC} - P_{FC}}{P_{max,FC}} \right)^{0.4} \quad (43)$$

$$g_3(x) = 1 - \left(\frac{P_{max,bat} - P_{bat}}{P_{max,bat}} \right)^{0.4} \quad (44)$$

$$g_4(x) = - \left(1 - \left(\frac{P_{max,UC} - P_{UC}}{P_{max,UC}} \right)^{0.4} \right) \quad (45)$$

Table 7. Shaped reward functions.

Function	Threshold	Description
$g_1(x)$	Activated when $SOC_{UC} > 0.65$ or ($SOC_{UC} > 0.65$ and $P_{load} < 0$)	Encourages high UC power or regenerative power
$g_2(x)$	Activated when $SOC_{UC} < 0.65$	Encourages high FC power
$g_3(x)$	Activated when $SOC_{UC} < 0.65$	Encourages high battery power
$g_4(x)$	Activated when $SOC_{UC} < 0.65$	Encourages low UC power

3.4. Soft Actor–Critic (SAC)

The essence of solving a reinforcement learning problem lies in optimizing the trade-off between exploration and exploitation. In contrast to supervised learning, in RL there are no labels, and the agent must learn to satisfy all the rules of the environment simply by exploration. During the training stage, the model will solely perform random actions and gradually find an optimal balance between exploration and exploitation, thus the optimal policy. On the other hand, during the testing stage, the model will only perform exploitation. That is, it will act according to the optimal policy learned during the training stage.

The soft actor–critic (SAC) model [36] is an off-policy method that uses DRL to find the optimal balance between exploration and exploitation by maximizing both the reward and the entropy of the system. By maximizing the entropy, the model is encouraged to keep exploring and thus assign similar probabilities to actions with similar action values and not assign excessively large probabilities to a specific set of actions. On the other hand, by maximizing the rewards, the model will strive toward finding the optimal policy. Therefore, given the large action and state spaces of our model (see Figure 5 and Table 6), we believe that the SAC model would be appropriate to learn the ideal power split algorithm for the HESS.

The architecture chosen for our SAC agent is a two-network design without shared features between them: one for the actor and one for the critic. Their parameters after hyperparameter tuning are shown in Table 8.

Table 8. SAC agent parameters.

Parameters	Values
Activation function	ReLU
Discount factor	0.999
Learning rate	3×10^{-4}
Batch size	2048
Number of hidden layers	2
Number of hidden units per layer	256
Target smoothing coefficient	0.005
Target update interval	1
Gradient steps	1
Replay buffer size	1,000,000

4. Experimental Results

4.1. SAC Agent Training Phase

The training phase of the SAC agent was performed using the WLTC class 3 data and the parameters shown in Table 8. Figure 9 shows the rewards over 1,400,000 iterations, while Figure 10 represents the average rewards over 10,000 steps. It may be seen that the average rewards converge to the value of about 800.

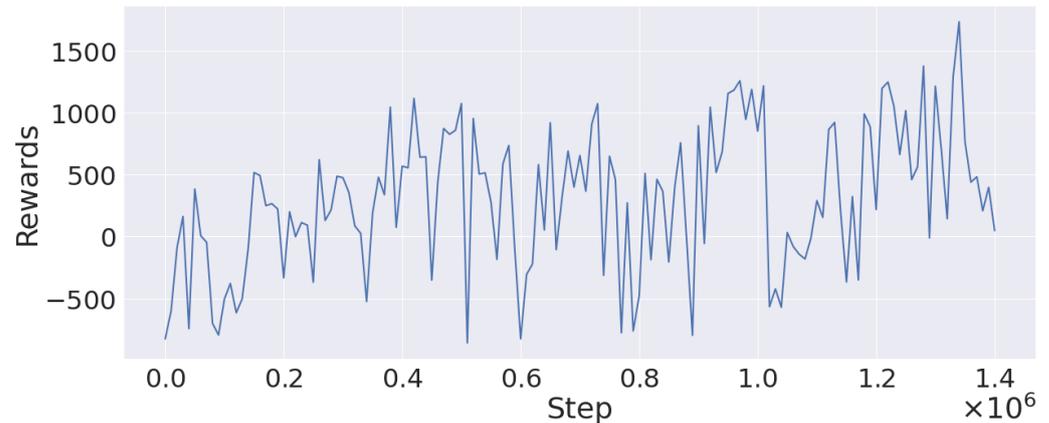


Figure 9. Rewards obtained by the SAC agent.

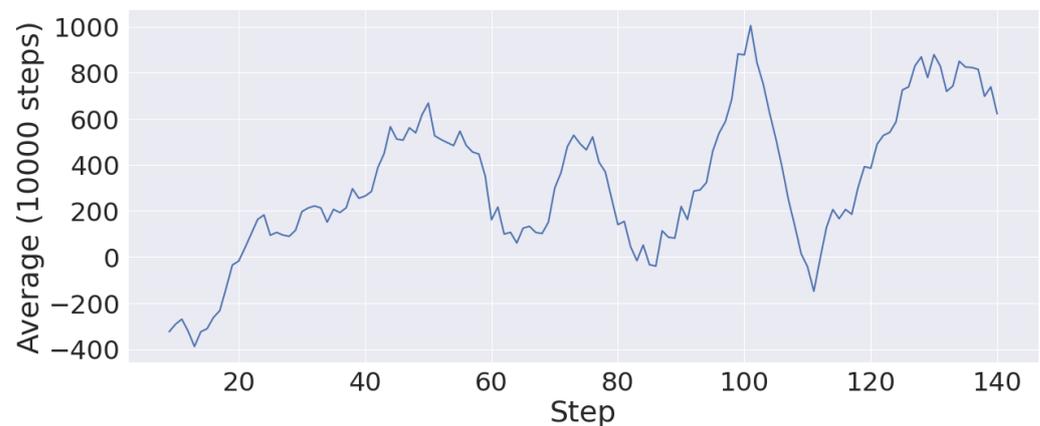


Figure 10. Average rewards computed every 10,000 steps.

4.2. SAC Agent Evaluation with SDCs

To evaluate the proposed method, a rule-based power split technique was employed in the HESS for the WLTC class 1 standard driving cycle. The power split rule changed based on the sign of the load power. For positive load powers, ratios of $\{P_{bat}, P_{UC}, P_{FC}\} = \{40\%P_{load}, 40\%P_{load}, 20\%P_{load}\}$ were used, and in the case of regenerative braking, ratios of $\{P_{bat}, P_{UC}, P_{FC}\} = \{50\%P_{load}, 50\%P_{load}, 0\%P_{load}\}$ were used. The results are shown in the Figures 11–13.

Both techniques have similar results in terms of power split, but it can be seen that the rule-based technique struggles with managing the SOCs and the currents of the ultracapacitor and the battery. The initial SOCs are 0.9, but after the 1023 s of simulation, the final SOCs of the ultracapacitor are about 62.8% and 46.9% for the DRL and rule-based technique, respectively, as shown in Figure 13. The latter value is below the acceptable value because the theoretically minimum allowed SOC for the ultracapacitor is 50%, as shown in Table 6 due to the approximately linear discharge of the ultracapacitor.

Additionally, the SAC model is far more robust than the rule-based one. Figure 14 shows the SOCs of the battery and the UC when their initial values are 0.7. The SAC model

is able to find the optimal power split while satisfying the constraints from Table 6, in contrast to the rule-based technique.

Finally, the DRL model also has the advantage of being able to perform well in new data without the need for retraining. It is shown in Sections 4.3 and 4.4 that the model can obtain good results, even for different data on which it was never trained.

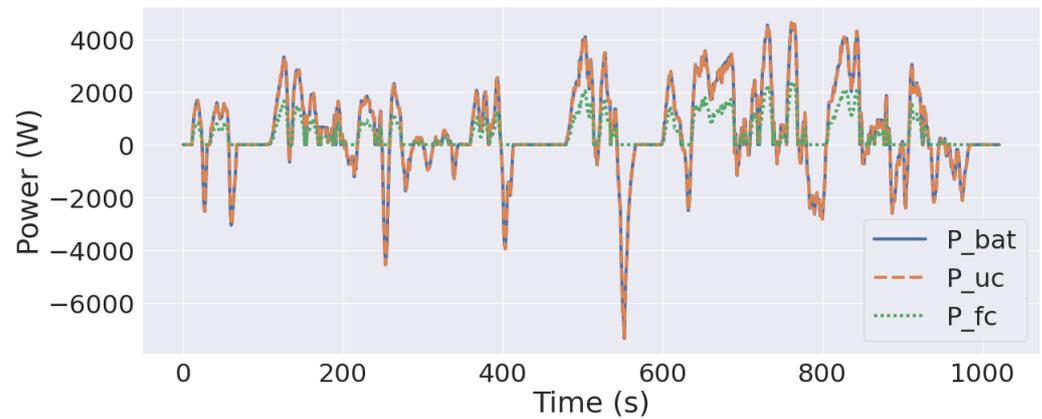


Figure 11. Battery, ultracapacitor and fuel cell powers by rule-based optimization.

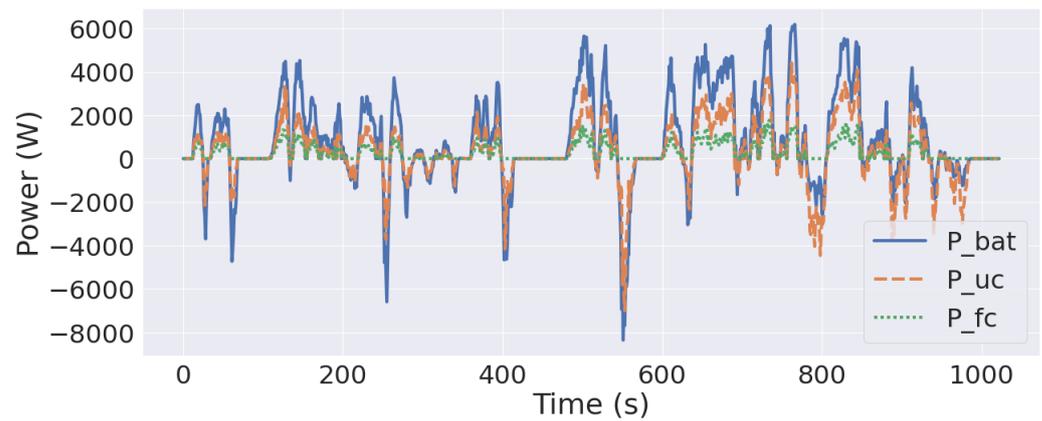


Figure 12. Battery, UC and FC powers by DRL.

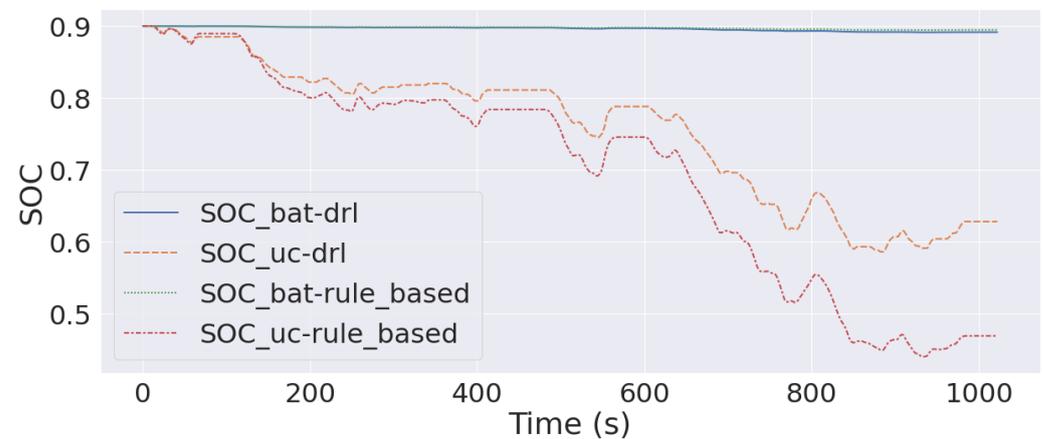


Figure 13. Comparison of SOC obtained by rule-based and DRL technique when the initial SOC is 0.9.

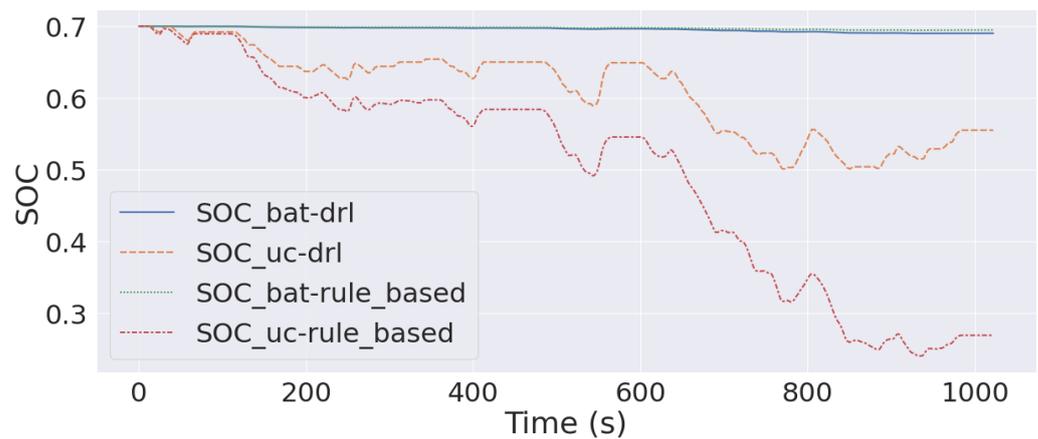


Figure 14. Comparison of SOC_s obtained by rule-based and DRL techniques when the initial SOC_s are 0.7.

4.3. SAC Agent Evaluation with RDCs

As explained in Section 2.2, four RDCs were employed to evaluate the model. Table 9 shows the results obtained by the proposed method when prediction of the speed and load power is not considered. First, it is interesting to analyze the SOC_{UC}. Two initial values were considered: 55% and 90%. In general, when the initial SOC_{UC} = 55%, the model could perform the optimal power split while respecting the constraints from Table 6, as the minimum SOC_{UC} did not go below 50%. Additionally, it may be seen that the model focused on recharging the UC through regenerative braking, as the final SOC_{UC} was greater than the initial. On the other hand, when the initial SOC_{UC} = 90%, the model allocated more power to the UC instead of recharging it, as the final SOC_{UC} of approximately 82% is smaller than the initial SOC_{UC}.

Second, considering that the minimum and maximum efficiencies η_{FC} are, respectively, 40% and 62.1%, the obtained results in the range of 55.8% to 57.3% were satisfactory. We have also plotted the FC powers related to the RDC1 for SOC_{UC} = 55% and SOC_{UC} = 90% in Figure 15 to better analyze the FC results. One thing to note is that there was no significant difference between the plots. However, one would expect that the model would allocate more power to the FC for low values of SOC_{UC} to reduce the use of the UC and prevent it from going below the minimum of 50%. This means that the model could be a little bit further optimized to improve the power split method.

Third, compared to the Toyota Camry MG1-only structure, there is a significant efficiency improvement when two motors are used. The high improvement of 17.6% is achieved because the two motors are complementary.

Table 9. Model results without prediction step.

RDC	SOC _{uc}				SOC _{bat}				Mean η_{FC}	Motors Efficiency		
	Initial	Final	Min	Max	Initial	Final	Min	Max		M1 Only	M1 + M2	Improvement
1	55%	61.2%	50.7%	61.2%	55%	54.5%	54.5%	55%	55.8%	69.6%	79.8%	14.7%
	90%	81.5%	77.3%	90%	90%	89.7%	89.6%	90%	56%			
2	55%	62.7%	50.4%	62.7%	55%	54.5%	54.4%	55%	56.2%	73.8%	80.5%	9.1%
	90%	82%	75.2%	90%	90%	89.7%	89.6%	90%	56.4%			
3	55%	62.5%	50.2%	62.9%	55%	54.4%	54.3%	55%	57.3%	68.3%	80.3%	17.6%
	90%	82%	75.3%	90%	90%	89.6%	89.5%	90%	57.3%			
4	55%	61.3%	50%	61.5%	55%	54.5%	54.4%	55%	55.9%	70%	81.3%	16.1%
	90%	81.5%	75.6%	90%	90%	89.7%	89.6%	90%	56.2%			

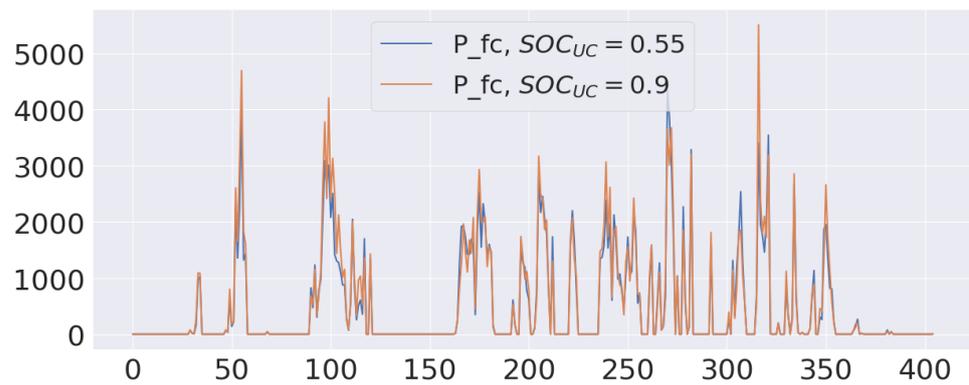


Figure 15. FC powers for different initial SOC_{UC} for the RDC1.

4.4. Evaluation of the Prediction Method with RDCs

In this section, we evaluate the velocity and load power prediction method using the model described in Sections 2.3 and 2.5, along with the SAC agent. Figures 16–18 show the results obtained. By analyzing the figures, it can be seen that the predicted powers and SOCs are not too far from the actual ones. Additionally, Table 10 was made to gain greater insights into the prediction method. It shows the mean absolute error (MAE) between the predicted and actual values for 10 different parameters: speed, load power, battery and ultracapacitor SOCs, fuel cell efficiency, battery and ultracapacitor currents and battery, ultracapacitor and fuel cell powers. In general, the MAE is good for the speed, SOCs and currents prediction. In the case of the load and energy sources powers, the mean deviation was relatively high (around 5000 W) because the prediction model is highly sensitive to differences in acceleration values. Analogously, the deviation in the fuel cell efficiency is also relatively high (around 11% to 14%) because it is highly sensitive to small changes in the fuel cell power values. For instance, there is a pair of points $(P_{FC,predicted}, P_{FC,label}) = (945.7W, 0) \implies (\eta_{FC,predicted}, \eta_{FC,label}) = (61.5\%, 0)$ where the big difference between the predicted and the actual efficiencies can be clearly seen.

Table 10. Model results with prediction step.

RDC	MAE									
	Speed (m/s)	$P_{load}(W)$	SOC_{bat}	SOC_{UC}	η_{FC}	$i_{bat}(A)$	$i_{UC}(A)$	$P_{bat}(W)$	$P_{UC}(W)$	$P_{FC}(W)$
1	0.261	4708.9	0.03%	0.65%	11.3%	0.007	0.11	2778.9	1618.7	458.1
2	0.273	5818.7	0.02%	0.68%	13.1%	0.007	0.13	3476.6	2012.5	553.9
3	0.293	5634.5	0.01%	0.61%	11.4%	0.006	0.11	3456.5	1861.9	480.5
4	0.281	5641.1	0.02%	0.32%	14.3%	0.003	0.14	3318.3	1912.1	522
Average	0.277	5450.8	0.02%	0.56%	12.5%	0.006	0.13	3257.6	1851.3	503.7

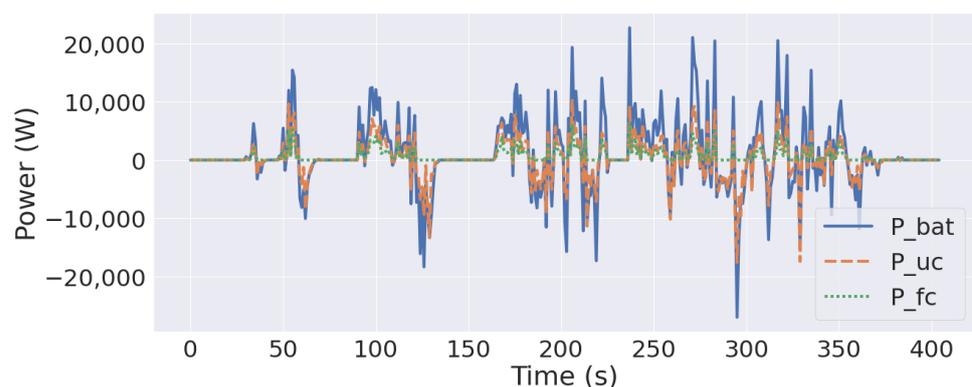


Figure 16. Predicted battery, ultracapacitor and fuel cell powers by DRL.

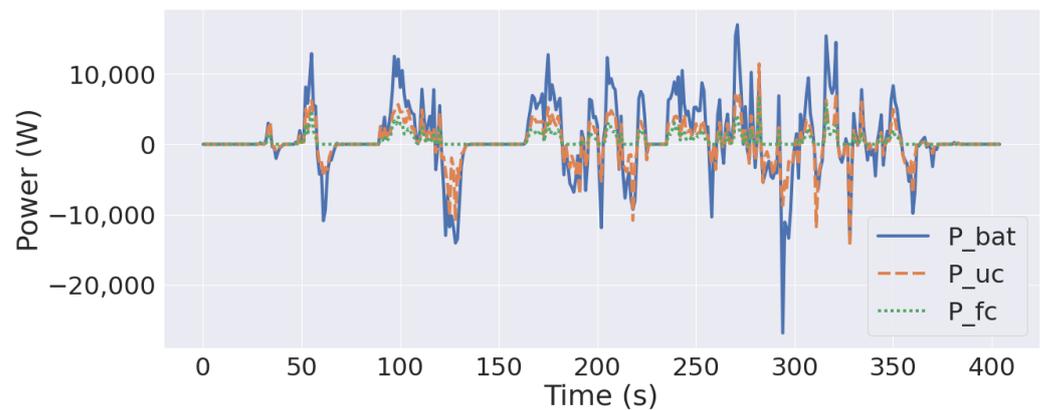


Figure 17. Actual battery, ultracapacitor and fuel cell powers by DRL.

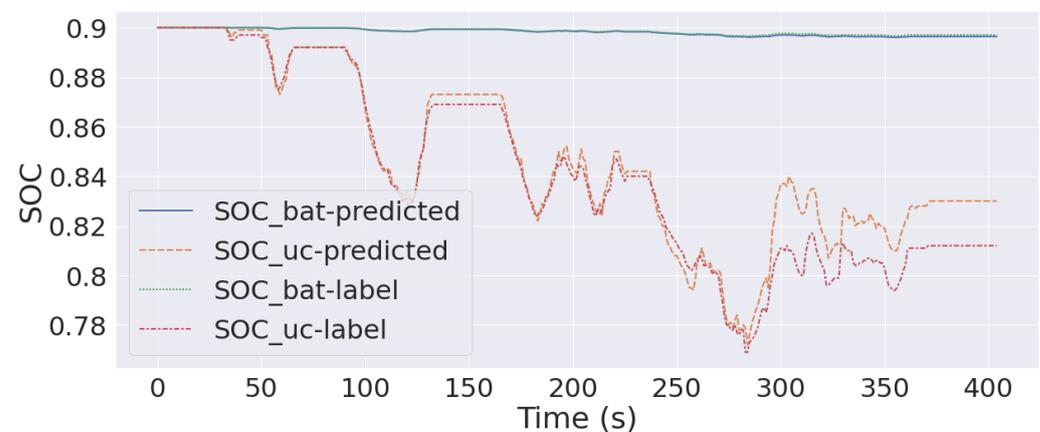


Figure 18. Comparison of the predicted and actual SOC obtained by DRL when the initial SOC values are 0.9.

5. Conclusions

In this paper, a DRL-based method for real-time joint power split optimization for a battery/UC/FC HEV was proposed. First, the TA-LSTM was responsible for predicting the future velocity, which was converted to required load power through the mechanical model. The load power was then optimally split between the two motors and also split between the battery/UC/FC by the proposed SAC agent, which makes use of shaped reward functions to accelerate the training process. Compared to traditional rule-based techniques, the proposed method is robust to different initial SOC values and is able to satisfy the system constraints. Moreover, the results show that the usage of two complementary motors greatly increase the efficiency of the system. Finally, the average MAEs of the prediction step are reliable; therefore, the method may be used to plan in advance the HESS power split.

In the future, a better model for the HEV, including auxiliary systems modeling (air conditioning, lights, sound, power-assisted seats, windows) could be designed to compute a more precise load power. Additionally, the velocity forecast method can currently predict velocity only for the next time step. A better forecast method would be able to forecast the velocity for more time steps, allowing the model to better allocate its resources. Finally, more emphasis could be given to the optimization of the fuel cell to increase its efficiency and its usage whenever the SOC of the ultracapacitor falls below the minimum operating value.

Author Contributions: Methodology, D.K. and S.H.; project administration, S.H.; software, D.K., S.H. and S.C.; supervision, I.J.; writing—original draft, D.K. and S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported partly by the Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. 2020-0-00107, Develop-

ment of the technology to automate the recommendations for big data analytic models that define data characteristics and problems), and partly by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2019R1I1A1A01058964).

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

References

1. United States Environmental Protection Agency and Office of Transportation and Air Quality and Standards Division. *Fast Facts: U.S. Transportation Sector Greenhouse Gas Emissions, 1990–2019*; 2021. Available online: <https://www.epa.gov/greenvehicles/fast-facts-transportation-greenhouse-gas-emissions> (accessed on 30 April 2022).
2. United Nations Department of Economic and Social Affairs. Ensure Access to Affordable, Reliable, Sustainable and Modern Energy for All. Available online: <https://sdgs.un.org/goals/goal7> (accessed on 1 April 2022).
3. Zuo, W.; Li, R.; Zhou, Z.; Li, Y.; Xia, J.; Liu, J. Battery-Supercapacitor Hybrid Devices: Recent Progress and Future Prospects. *Adv. Sci.* **2017**, *4*, 1600539. [CrossRef] [PubMed]
4. Sanguesa, J.A.; Torres-Sanz, V.; Garrido, P.; Martinez, F.J.; Marquez-Barja, J.M. A Review on Electric Vehicles: Technologies and Challenges. *Smart Cities* **2021**, *4*, 372–404. [CrossRef]
5. Li, H.; Ravey, A.; N'Diaye, A.; Djerdir, A. Equivalent consumption minimization strategy for hybrid electric vehicle powered by fuel cell, battery and supercapacitor. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 4401–4406.
6. Li, H.; Ravey, A.; N'Diaye, A.; Djerdir, A. A novel equivalent consumption minimization strategy for hybrid electric vehicle powered by fuel cell, battery and supercapacitor. *J. Power Sources* **2018**, *395*, 262–270. [CrossRef]
7. Zhang, W.; Li, J.; Xu, L.; Ouyang, M. Optimization for a fuel cell/battery/capacity tram with equivalent consumption minimization strategy. *Energy Convers. Manag.* **2017**, *134*, 59–69. [CrossRef]
8. Khan, F.; Oh, M.; Kim, J.H. N-functionalized graphene quantum dots: Charge transporting layer for high-rate and durable Li₄Ti₅O₁₂-based Li-ion battery. *Chem. Eng. J.* **2019**, *369*, 1024–1033. [CrossRef]
9. Zhang, M.; Liang, R.; Or, T.; Deng, Y.P.; Yu, A.; Chen, Z. Recent Progress on High-Performance Cathode Materials for Zinc-Ion Batteries. *Small Struct.* **2021**, *2*, 2000064. [CrossRef]
10. Yao, P.; Yu, H.; Ding, Z.; Liu, Y.; Lu, J.; Lavorgna, M.; Wu, J.; Liu, X. Review on polymer-based composite electrolytes for lithium batteries. *Front. Chem.* **2019**, *7*, 522. [CrossRef]
11. Sun, Y.K. Promising all-solid-state batteries for future electric vehicles. *ACS Energy Lett.* **2020**, *5*, 3221–3223. [CrossRef]
12. Khan, F. N and S co-doped graphene enfolded Ni–Co-layered double hydroxides: An excellent electrode material for high-performance energy storage devices. *RSC Adv.* **2021**, *11*, 33895–33904. [CrossRef]
13. Kasimalla, V.K.; Velisala, V. A review on energy allocation of fuel cell/battery/ultracapacitor for hybrid electric vehicles. *Int. J. Energy Res.* **2018**, *42*, 4263–4283. [CrossRef]
14. Geetha, A.; Subramani, C. A comprehensive review on energy management strategies of hybrid energy storage system for electric vehicles. *Int. J. Energy Res.* **2017**, *41*, 1817–1834. [CrossRef]
15. Sellali, M.; Abdeddaim, S.; Betka, A.; Djerdir, A.; Drid, S.; Tiar, M. Fuzzy-Super twisting control implementation of battery/super capacitor for electric vehicles. *ISA Trans.* **2019**, *95*, 243–253. [CrossRef] [PubMed]
16. Mokrani, Z.; Rekioua, D.; Mebarki, N.; Rekioua, T.; Bacha, S. Proposed energy management strategy in electric vehicle for recovering power excess produced by fuel cells. *Int. J. Hydrogen Energy* **2017**, *42*, 19556–19575. [CrossRef]
17. Majeed, M.A.; Khan, M.G.; Asghar, F. Nonlinear control of hybrid energy storage system for hybrid electric vehicles. *Int. Trans. Electr. Energy Syst.* **2020**, *30*, e12268. [CrossRef]
18. Zhu, L.; Han, J.; Peng, D.; Wang, T.; Tang, T.; Charpentier, J.F. Fuzzy logic based energy management strategy for a fuel cell/battery/ultra-capacitor hybrid ship. In Proceedings of the 2014 First International Conference on Green Energy ICGE 2014, Sfax, Tunisia, 25–27 March 2014; pp. 107–112.
19. Yavasoglu, H.A.; Shen, J.; Shi, C.; Gokasan, M.; Khaligh, A. Power Split Control Strategy for an EV Powertrain with Two Propulsion Machines. *IEEE Trans. Transp. Electrif.* **2015**, *1*, 382–390. [CrossRef]
20. Yavasoglu, H.A.; Tetik, Y.E.; Ozcan, H.G. Neural network-based energy management of multi-source (battery/UC/FC) powered electric vehicle. *Int. J. Energy Res.* **2020**, *44*, 12416–12429. [CrossRef]
21. Zhang, P.; Yan, F.; Du, C. A comprehensive analysis of energy management strategies for hybrid electric vehicles based on bibliometrics. *Renew. Sustain. Energy Rev.* **2015**, *48*, 88–104. [CrossRef]
22. Wu, X.; Hu, X.; Yin, X.; Li, L.; Zeng, Z.; Pickert, V. Convex programming energy management and components sizing of a plug-in fuel cell urban logistics vehicle. *J. Power Sources* **2019**, *423*, 358–366. [CrossRef]
23. Choi, M.E.; Lee, J.S.; Seo, S.W. Real-Time Optimization for Power Management Systems of a Battery/Supercapacitor Hybrid Energy Storage System in Electric Vehicles. *IEEE Trans. Veh. Technol.* **2014**, *63*, 3600–3611. [CrossRef]
24. Li, Y.; He, H.; Peng, J.; Wang, H. Deep reinforcement learning-based energy management for a series hybrid electric vehicle enabled by history cumulative trip information. *IEEE Trans. Veh. Technol.* **2019**, *68*, 7416–7430. [CrossRef]
25. Xu, D.; Cui, Y.; Ye, J.; Cha, S. W.; Li, A.; Zheng, C. A soft actor-critic-based energy management strategy for electric vehicles with hybrid energy storage systems. *J. Power Sources* **2022**, *524*, 231099. [CrossRef]

26. Xu, B.; Rathod, D.; Zhang, D.; Yebi, A.; Zhang, X.; Li, X.; Filipi, Z. Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle. *Appl. Energy* **2020**, *259*, 114200. [CrossRef]
27. Sun, H.; Fu, Z.; Tao, F.; Zhu, L.; Si, P. Data-driven reinforcement-learning-based hierarchical energy management strategy for fuel cell/battery/ultracapacitor hybrid electric vehicles. *J. Power Sources* **2020**, *455*, 227964. [CrossRef]
28. Qi, X.; Luo, Y.; Wu, G.; Boriboonsomsin, K.; Barth, M.J. Deep reinforcement learning-based vehicle energy efficiency autonomous learning system. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; pp. 1228–1233.
29. Hong, S.; Hwang, H.; Kim, D.; Cui, S.; Joe, I. Real Driving Cycle-Based State of Charge Prediction for EV Batteries Using Deep Learning Methods. *Appl. Sci.* **2021**, *11*, 11285. [CrossRef]
30. Burress, T.A.; Coomer, C.L.; Campbell, S.L.; Seiber, L.E.; Marlino, L.D.; Staunton, R.H.; Cunningham, J.P. *Evaluation of the 2007 Toyota Camry Hybrid Synergy Drive System*; Oak Ridge National Laboratory: Oak Ridge, TN, USA, 2008.
31. UQM. “UQM PowerPhase 125”, Datasheet. Available online: https://wiki.neweagle.net/ProductDocumentation/EV_Software_and_Hardware/Traction_Inverters/UQM/PowerPhase_125_DataSheet.pdf (accessed on 30 April 2022)
32. K2 Energy. “High Capacity LFP26650P Power Cell Data”, Datasheet. Available online: <http://liionbms.com/pdf/k2/LFP26650P.pdf> (accessed on 30 April 2022).
33. Maxwell Technologies. “Datasheet: K2 Series Ultracapacitors”, Datasheet. Available online: <https://mp36c.ru/pdf/library/datasheets/SC/K2%20series%20maxwell.pdf> (accessed on 30 April 2022).
34. Rinehart Motion Systems, LLC. “Technical Datasheet of PM100 and PM150 Family Propulsion Inverter”, Datasheet. Available online: http://www.rinehartmotion.com/uploads/5/1/3/0/51309945/pm100-150datasheet_1.pdf (accessed on 9 June 2022).
35. Wiewiora, E. Reward Shaping. In *Encyclopedia of Machine Learning*; Sammut, C., Webb, G.I., Eds.; Springer: Boston, MA, USA, 2011. [CrossRef]
36. Haarnoja, T.; Zhou, A.; Abbeel, P.; Levine, S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. *arXiv* **2018**, arXiv:1801.01290.