# ENERGY MANAGEMENT FOR FUEL CELL HYBRID ELECTRIC VEHICLE

**Lakshmi Priya. G[1*], Anjali J Nair[2]**

*[1*]CSI College of Engineering, Ketti.*
*[2]NSS College of Engineering, Palakkad*

***[*]Corresponding Author:***

## ABSTRACT

*The use of internal combustion engines is being increasingly scrutinized because of their high emission levels. The research and development of electric and hybrid vehicles has been prompted by the demand for cleaner energy technologies. Fuel cell vehicles are gaining attention because they are clean, sustainable, and have a high energy density. Thus, fuel cell hybrid vehicles have the potential to compete with vehicles powered by internal combustion engine in the future, yet there are challenges for fuel cell such as slow dynamics requiring that their operation together should be managed favourably. The main aim of the thesis is to tackle the issue of energy management in fuel cell vehicles. The power train model is the first thing developed for this purpose. Deep deterministic policy gradient (DDPG) is a model-free reinforcement learning algorithm used to achieve efficient energy management. The energy management strategy focuses on running the fuel cell in its high efficiency range while limiting the deviation of state of charge of the lithium-ion battery from a target value. It is found that the DDPG agent trained simply with step power inputs can achieve up to 2.7% less energy consumption compared to commonly used rulebased energy management strategies while maintaining the state of the charge of the battery within a certain interval. Our findings indicate that the DDPG algorithm has a promising potential for use in such applications.*

## I. INTRODUCTION

the industrial revolution, fossil fuels have been the main energy source of vehicles and evolved into mostly gasoline or diesel due to the several advantages it offers, such as long range, high power and energy density, fast replacement, easy storage. The shortage of fossil energy has been caused by the extensive exploitation of these fuels, which only take thousands of years to form. Moreover, its hazardous emissions have caused a negative impact on the environment and human health. As a result, there has been an increase in demand for vehicles that use alternative energy sources recently. To reduce emissions and decrease dependency on fossil fuels, electric motor-powered vehicles are being offered instead of combustion engines. Although there have been recent improvements, there have always been concerns about this vehicle's limited range, long charge duration, and lack of charging infrastructure. However, hybrid electric vehicles that combine multiple sources can resolve these issues. Parallel hybrids, series hybrids, and power-split hybrids are the most popular hybrid electric vehicles (HEVs). The fact that there are multiple configurations of these types of vehicles makes it difficult for classifications to draw a precise line and be conclusive. A hybrid vehicle configuration where an internal combustion engine (ICE) and an electric motor (EM) both partake in the propulsion of the vehicle mechanically is classified under the category of parallel hybrid vehicle. The standalone electric motor is capable of serving the purpose when necessary, so an extra generator is not necessary to achieve this architecture. Running the engine and motor in their efficient ranges, which can solve optimization problems, is the main advantage of such a system. If the ICE is not involved in mechanically propelling the vehicle, it is considered a series hybrid vehicle, unlike parallel hybrid vehicles. The ICE is solely responsible for providing power to charge the battery that the EM utilizes. Due to the fact that the sole purpose of the ICE is to charge the battery in this configuration, it is possible to operate it in its most efficient ranges.purpose of the ICE is to charge the battery in this configuration, it is possible to operate it in its most efficient ranges.

The most promising hybrid vehicle type in terms of achieving zero emissions is FCHEVs. No hazardous end product is generated by the system when chemical reactions take place during energy production. The use of fuel cell technology in transportation is new and could be enhanced. Traditional vehicles powered by ICE, PHEV, HEV, and pure EV have advantages over this vehicle. FCHEV eliminates the range issue of EV due to the similar operation to HEV or conventional vehicles with ICE. Continuous travel requires a hydrogen tank that can be refilled in less than a few minutes. The fuel cell's efficiency can go up to 60% because there is no combustion, unlike ICE. There is no issue with the range of these vehicles. The use of a fuel cell as the sole energy source presents challenges due to the system's structural requirements, which may take some time to start up. This problem can be solved by using it with a battery instead. There are other challenges as well such as the cost, hydrogen supply to customer, energy management strategy, safety and reliability, but in this study we will focus on minimizing the energy consumption of such vehicles [9]. Figure 1.1 demonstrates the structure of an FCHEV. The fuel cell stack, which consists of several fuel cells, uses the hydrogen fuel tank and oxygen in the air to generate electricity along with the battery pack. The DC-DC converter transmits electricity to the EM. A cooling system is necessary because the sources generate heat and produce a side product.

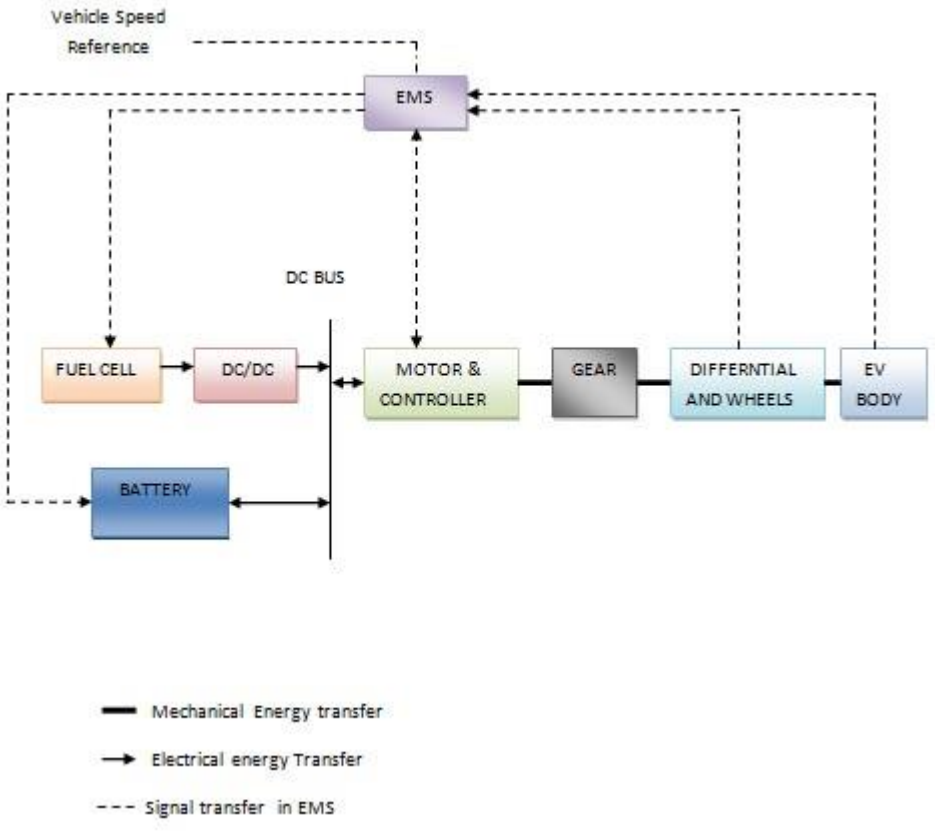


**Figure 1.1 Structure Of Fuel Cell EV**

## II. OBJECTIVES

Propose a novel energy management strategy where the purpose is reducing the energy consumption of a FCHEV model involving the energy sources and DC-DC converters by improving the efficiency of the fuel cell and maintaining the state of charge of the battery within certain limits, by using the reinforcement learning algorithm DDPG

## III. ENERGY MANAGEMENT IN HYBRID VEHICLES

Hybrid vehicles have no single energy source or power converter, so a control method is necessary regardless of whether it's an HEV, PHEV, or FCHEV. Efficiency characteristics and optimal operating points vary between different machines. The objective is to decrease overall energy consumption. Even though HEVs and PHEVs are more common in the market resulting in the fact that energy management strategies (EMS) are extensively researched, there are still studies on EMS in FCHEVs as well. Optimization-based and rule-based EMS can be classified into two main categories. The optimization methods on the other hand can be applied in a way that either a global optimum point is found with the driving cycle data which is known beforehand or a sub optimal point is found with not only the past information but also present and future information. The first method's optimality cannot be updated, but the latter's is dynamic and adaptive.

| FC Type | Operating Temperature and Efficiency | Typical Stack Size | Automotive Applications | Advantages | Disadvantages |
|---|---|---|---|---|---|
| Polymer Electrolyte Membrane (PEM) | <120°C 50-60% | 1 - 100 kW | -Backup power -Portable power -Transportation | -High power density -Low temperature -Quick start-up -Quick load following -Solid electrolyte reduces corrosion | -Sensitive to fuel impurities -Expensive catalysts |
| Solid Oxide (SOFC) | 500-1000 °C 60% | 1 kW-2 MW | -Auxillary power -Electric Utility | -Tolerance to fuel impurities -Fuel flexibility -High efficiency | -Long start-up -Slow dynamic load behaviour -High temperature -Corrosion of components |
| Molten Carbonate (MCFC) | 600-700°C 50% | 0.3 - 3MW | -Electric Utility | -High efficiency -Fuel flexibility | -Low power density -High temperature -Long start-up time |
| Alkaline (AFC) | <100 °C 60% | 1-100 kW | -Military -Space -Backup power -Transportation | -Low cost components -Low temperature -Quick start-up | -Sensitive to CO2 in fuel and air -Electrolyte management -Electrolyte conductivity |
| Phosphoric Acid (PAFC) | 150-200 C 40% | 5-400 kW | -Distributed generation | -Suitable for CHP -Increased tolerance to fuel impurities | -Expensive catalysts Long start-up time -Low power density |

**Figure 3.1 Fuel cell types**

### (a) RULE-BASED STRATEGIES

require a set of conditions to be checked in each time instance. Those rules are derived heuristically and cannot guarantee an optimal operating point. However it is widely used today due to the fact that it is practical, easy to apply and works fast. Based on those rules controller decides how to share the power demand of the vehicle between the energy sources. The strategy is applied to vehicle systems [12] and also in the form of fuzzy logic [13]. The same strategy is applied to a FCHEV as well [14] and the adaptation will be used for comparison in this study.

### (b). REAL TIME OPTIMIZATION (RTO) METHODS

predicts the optimal output within a process and keeps measuring the the real data. By doing so instead of finding a global optimum point, several optimization problems are created to be solved at each time step. The method is developed in order to address the uncertainty of the real-time interaction of the controller with the environment. It is aimed for the controller to respond properly in the case of existence of a disturbance. The method provides a framework for not only past but also present and the future informationto be utilized. The most popular methods applied for RTO are equivalent cost minimization strategy (ECMS) [15] [16] and model predictive control (MPC) [17] [18].

### (c). GLOBAL OPTIMIZATION METHODS

tries to find the optimum point of a given objective function using a set of data most often driving cycles in our case. Mostly a combination of different drive cycles are fed into the model and best set of decisions are made in order to optimize the given function. Dynamic programming (DP) is widely used for that purpose [19] [20] [21] that is a method benefiting from the principal of optimality idea of Bellman equation. Linear programming [22] and convex optimization [23] are also used for that purpose however DP is still the most common approach as it almost ensures that global optimum is found. The only drawback in terms of accuracy stems from the discretization of control input which can be avoided largely for the expense of simulation time. Such methods are obviously excellent for comparing the performance of any other method as it sets the best achievable target for the objective function. Moreover the result of the other optimization or even rule based methods can be updated in order for it to be closer to the best possible outcome [24].Lately as the reinforcement learning algorithms started to become a promising technique and being applied to many control problems, thanks to development of the model-free algorithms that can be applied to any environment defined as an MDP. Q learning and Deep Q learning (DQN) are the most common model-free algorithms in the studies focusing on HEV or PHEV that are similar to FCHEV. In Q learning or also denoted as Q table learning a random action is selected causing a state transition and a reward is obtained. The sum of future and immediate rewards are collected in the Q table consisting of

state and action values. In every step if the current reward is greater than the previous reward, the table containing the value of the reward corresponding to that state and action value is updated.

**IV. FUEL CELL MODEL**

The main power source of the vehicle considered in this study is the polymer electrolyte membrane fuel cell. A fuel cell uses hydrogen and oxygen to generate electricity, heat and water. It is similar to the batteries in some aspects such that it has elements as anode, cathode, electrolyte and separator resulting in a similar architecture. The main difference is that fuel cells are not energy storage devices but energy conversion devices. They can generate electricity as long as the fuel (H2) is flown into the anode. In the anode, particularly the catalyst later, hydrogen is split into H+ ions and electrons through the chemical reactions. Hydrogen ions are allowed to pass through the exchange membrane as electrons are not and instead they move to the outer circuit supplying electricity for the load. On the other side of the fuel cell in the cathode, oxygen flow is performed capturing the electrons and the H+ ions producing water as summarized in figure 2.5. Hydrogen is stored in a tank in a pressurized form whereas oxygen flow is conducted via a compressor which takes the air as the input.

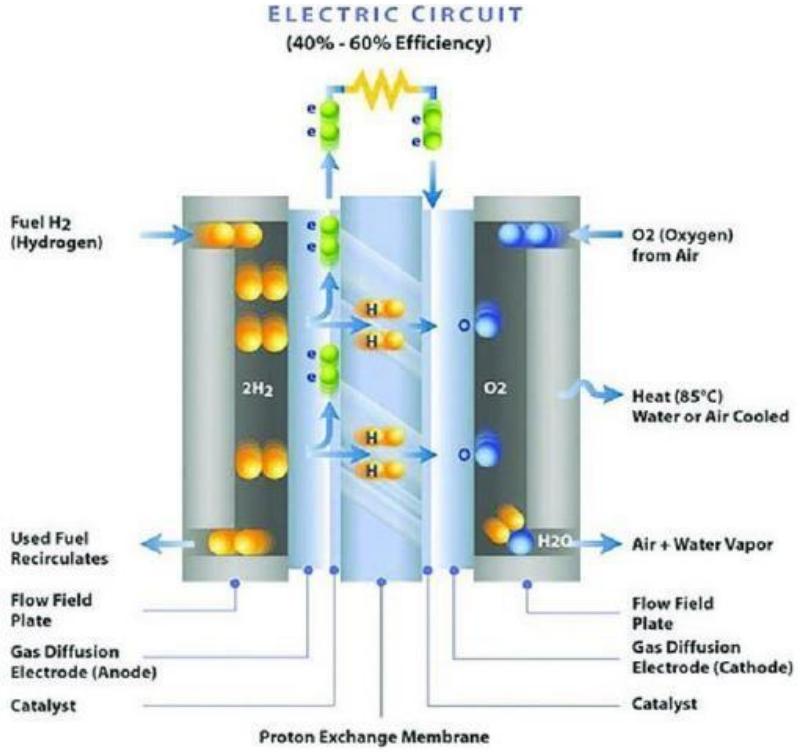


**Figure 4.1. Schematic of a PEM Fuel Cell [6]**

**V. REINFORCEMENT LEARNING**

Reinforcement learning is considered one of the three paradigms in machine learning alongside supervised and unsupervised learning and applied extensively in many areas. It is a method in which an agent learns how to form the relation between the states and the actions based on the reward function. A random action selected starting from the initial time step for a certain state causing a state transition, action interacts with the environment and as a result a reward value is gained for every single time step. The relation is illustrated in figure 3.1. There are several algorithms serving the purpose, using Markov Decision Process (MDP) which contains state, action, reward and the next state ($S_t$ ,$A_t$ ,$R_t$ ,$S_{t+1}$), as formalization of the problem. In order to comprehend the interactions, elements of the reinforcement learning algorithms, states, actions and rewards, must be introduced.

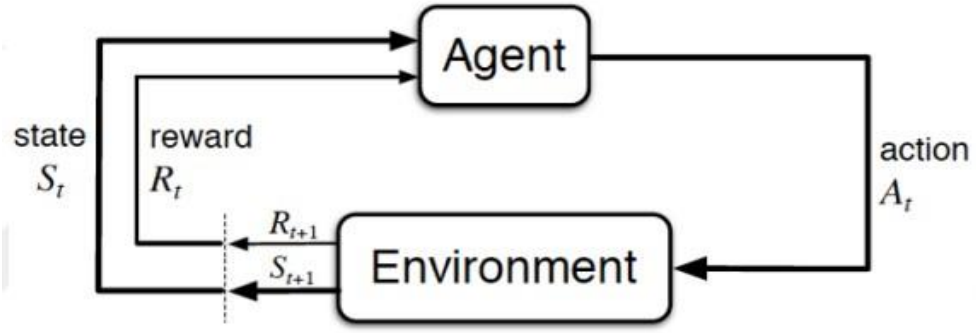


**Figure 5.1: The agent environment interaction in reinforcement learning [7]**

States in reinforcement learning can be considered as an input to the agent consisting of neural networks. It provides the information of the environment after an action interaction. It must contain enough variable for the agent to completely understand the system. If the number of variables in the state increases then it will take extensive amount of time for the training process to be concluded and reward to be maximized. Systems must be investigated thoroughly and minimum number of state variables must be defined because of the curse of dimensionality.

Actions are the outputs of the agent and inputs for the environment. The agent selects those randomly and feeds into the environment then checks the situation via states. As they are the only feature the agent can control, variables in that space must chosen carefully. Compared to state selection it is simpler to choose action variables in any case.

For a state an agent takes an action and it return it gets a reward. Since the main purpose of the reinforcement learning is to maximize reward, it must be somewhatsimilar to the objective function. It is not always easy to select the right reward function. Setting it as a similar function to objective function is a simple yet an inefficient approach. The agent sometimes requires some extra encouragements when it is choosing an action in the right direction. Apart from the function itself that is supposed to produce higher rewards when an action serves its purpose, extra rule based implementations might steer the agent to the right direction in a shorter time. Another factor is the numerical range of reward output. If the difference between two rewards is too high then it is possible for the network to be updated drastically causing divergence. Another issue is that the reward might be deceptive for some episodes. This is particularly a problem of the initialization process. If the range of initial variables of the system is too large, a reward similar to the objective function is likely to fail since it will never be clear which actions are actually good. For instance initialization might start somewhere close to the target and an action even though it is completely inaccurate might take more reward than an action simulated in the system initialized far from the target and that is indeed is the best of all. Furthermore it is possible and even certain that the agent will take some actions that will result in deviation from the goal particularly in the first steps of training. In that case a penalty should be defined in order to discourage the agent to take those actions againAs the simulation progresses the rewards are not accumulated directly, instead future rewards are multiplied by a discount factor to ensure that in a long horizon total reward converges to a value. The value of the discount factor is crucial for both continuous tasks where the problem defined in a time horizon cannot be divided into sub-simulations and episodic tasks where simulation time can be set thus dividing the complete process into sub-groups. In episodic tasks as in our problem until the terminal state is reached, in most case it is the last time step of the simulation or the time when the simulation is stopped as a punishment, target of the value function is updated as in equation 3.1.

$$y_i = R_i + \gamma \max_{A'} Q_t\left(S_i', A' \mid \phi_t\right) \tag{5.1}$$

The target is the sum of the immediate reward and the expected future rewards. As the discount factor approaches to zero target value will always be zero meaning that in every step the next state will be considered as a terminal state and only the immediate reward will be taken into account. On the other hand a discount ratio equal to 1 will cause the future rewards as equally as important compared to the immediate reward.

In the final step the loss function is calculated. As mentioned above a value function was calculated with the immediate and expected rewards. The loss function is the square of the difference between value function target and the current value of the value function as in equation 3.2.

$$L = \frac{1}{M}\sum_{i=1}^{M}\left(y_i - Q\left(S_i, A_i \mid \phi\right)\right)^2 \tag{5.2}$$

It is a metric of how close the value function is to the target or how small the values of the expected rewards are. The update of the mapping occurs at this step, after minimizing the loss function.

In order to apply any model-free algorithm of reinforcement learning the problem must be defined. Unlike model-based algorithms, the model-free algorithms as the name suggests can be applied to any model as long as the problem definition includes the elements to be described are action and state space and reward function. Before introducing those, the objective function should be explained as it will demonstrate itself the reason behind the selection of the elements.

## VI. DEEP Q LEARNING (DQN)

Before explaining DQN method, first the most basic reinforcement learning algorithm called Q-learning must be introduced. In Q learning the Q-value is stored in a Q-table in which the dimensions are state and action. The most common practice is to use the temporal difference method that is integrated into the Bellman equation. The equation is the founding base of the learning algorithm and may be slightly modified in different algorithms. Q value calculation with temporal difference is given:

$$Q^{\text{new}}\left(s_t, a_t\right) \leftarrow Q\left(s_t, a_t\right) + \alpha \cdot \left(r_t + \gamma \cdot \max_a Q\left(s_{t+1}, a\right) - Q\left(s_t, a_t\right)\right) \tag{6.1}$$

The table is updated in each step with the learning rate multiplier denoted as α and ends when the Q-value cannot increase anymore supposedly because the best action sets for each state is found. It is also possible that actions are stuck because of their greediness. The trade-off between exploration and exploitation that is defined by ϵ is the most common problem.

When ϵ that is defined between 0 and 1 increases exploration rate increases as well resulting in more random action selection thus giving priority to the future rewards. When the number is close to 0 the algorithms becomes greedy and approaches to the immediate reward. The most striking downside of the algorithm is that it requires a Q-table whose number of elements is the multiplication of the number of states and the actions. DQN utilizes deep neural network (DNN) instead of a Q table and makes it possible for problems involving with a larger stateaction space to be solved in shorter time or solvable at all. DNN consists of several layers first of which is called the input layer and it ends with the output layer; it is derived from artificial neural networks that only consists of an input, hidden and the output layer. DNN on the other hand benefits from several hidden layers in order to provide the opportunity for more complex correlations to be found. Those layers have nodes and all the nodes in each layer are connected to each other. Depending on the application they the connections may differ however in its basic form it has a feed-forward structure as observed in figure 3.2. In each node there is an activation function whose parameters are updated in order to map the input to the output correctly. The accuracy of mapping or fitting is naturally affected by the number of nodes and the layers. Even though there is not a straightforward guideline to find the numbers that will produce the best results,

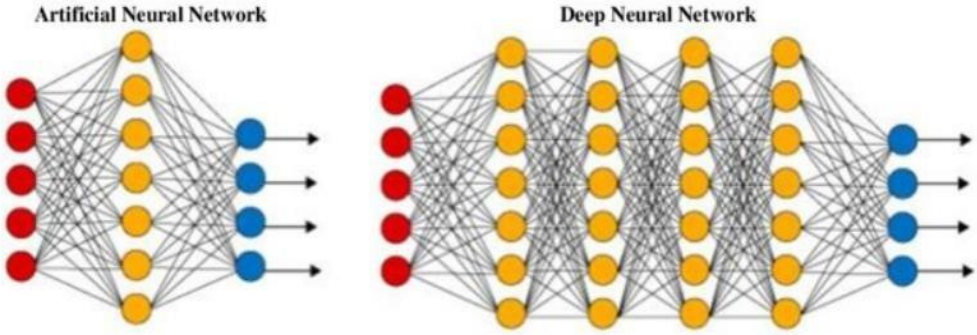


**Figure 6.1 Artificial and Deep Neural Network [8]**

the basic approach is to build the network as simple as possible and try to see if the fitting has acceptable error. When the layer and node number increases it takes a lot of computation time and that might be misleading in terms of convergence. It is possible that in such a case it requires vast amount of time that is not predicted by the user. However it is not simple to define the level of complexity.

Instead of using temporal difference method in the value function, and storing this q value in a table, DQN uses a target value function in which only the immediate and future rewards are summed. The network is then updated with the gradientdescent of the loss function. In addition instead of evaluating every case one by one, experience is utilized as explained in the reinforcement learning section. DQN is applied to our problem and the action to be taken is a form of how much of the demand power will be supplied by the battery. The power flow is controlled by the switches in the converter. $F_{bat}$ is the gain in the converter model that is selected as the action. Instead of choosing $F_{FC}$ and $F_{bat}$ as the action variables it is decided that only one of them will be included. Their sum is constant, one is dependent and the other one is independent variable. The equation below shows the relation between the gains and the battery current.

$$I_{bat} = \frac{F_{bat}}{F_{bat} + F_{FC}} * I_{in} \tag{6.2}$$

As long as sum of $F_{bat}$ and $F_{FC}$is greater than one the system works robustly. It can be concluded from the computational experiments that it is observed that changing this sum improves system response thus it is picked as 4

$$a = \{F_{bat}\}\ where\ F_{FC}\ 4 - F_{bat} \tag{6.3}$$

The range of the action is selected as below after trial and error, as the DQN algorithm requires discrete actions the range is split into 16 steps with a step size of 0.4

$$-2 < F_{bat} < 4 \tag{6.4}$$

Table 6.1 shows the modes of power sharing

Table 3.1: Operation Modes

| Mode | $F_{FC}$ | $F_{bat}$ |
|---|---|---|
| FC charges battery and supplies power | 6 | -2 |
| Only FC supplies power | 4 | 0 |
| FC and battery supplies equal current | 2 | 2 |
| Only Battery supplies power | 0 | 4 |

State variable candidates in the problem are $P_{demand}$,$P_{bat}$,$P_{fc}$,$FC_{eff}$ ,SOC,$SOC - SOC_{desired}$ that are power demand, power supplied by the battery and fuel cell, fuel cell efficiency, state of charge and deviation of the state of charge respectively. Theycan be defined in a different form however those are the main variable candidates. In the training process several combinations are tried and finally reward maximization achieved. Those state variables for DQN are:

$$S = \{SOC - SOC_{desired}, P_{demand}\} \quad (6.5)$$

As the state $P_{demand}$ is indeed the input of the system, there cannot be any limitations. On the other hand the first state SOC $-SOC_{desired}$ is limited as:

$$-SOC_{difference}, Limit < SOC - SOC_{desired} < SOC_{difference}, Limit \quad (6.6)$$

## VII. DEEP DETERMINISTIC POLICY GRADIENT (DDPG)

DDPG algorithm is similar to DQN however differs when it comes to updating the network parameters. DDPG is a member of the actor-critic algorithms though DQN has only one network structure. Actorcritic approach resembles the relation between a child and a mother. When an action is decided and interacts with environment the critic guides the actor in the right direction whereas in DQN there is only one network and it is led by the value function only. DDPG utilizes the actor critic approach and those two different networks are updated with different methods. Critic network is updated in the same way networks are updated in DQN. On the other hand the update of the actor network is conducted by the gradient descent. Similar approach is observed in the way that parameters are updated. However by doing so it is possible to define the action space continuously and decreases the errors caused by the discretizetion. The gradient is calculated as:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a Q\left(s, a \mid \theta^Q\right)\Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu\left(s \mid \theta^\mu\right)\Big|_{s_i} \quad (7.1)$$

The gradient of the critic with respect to action and the actor output with respect to the actor parameters is multiplied in order to find the gradient. DDPG is a more complex algorithm and the training process takes longer time compared to DQN.

However the fact that it does not require discrete action space is a huge advantage. The action is selected according to the current policy and is distorted with a noise function that is decaying throughout the process in order to increase exploration.

The action space defined for DDPG algorithm is the same that of DQN with one difference, they are not discrete. State space on the other hand is slightly different. Again after several training episodes the state variables are chosen as:

$$s = \{SOC, H_{2,eff}\} \quad (7.2)$$

As in DQN state variables are limited in DDPG as well. The limitation for the variables SOC and $H_{2,eff}$ respectively are:

$$SOC_{min} < SOC < SOC_{max} \quad (7.3)$$

$$H_{2,eff\ min} < H_{2,eff} < H_{2,eff\ max} \quad (7.4)$$

Efficiency of the fuel cell is directly related to the power supplied by the fuel cell. As the agent observes the efficiency value without knowing how much power is supplied by the fuel cell, it takes a reward in that state. The idea here is that the agent does not need to know the fuel cell power but only the efficiency curve. So it is not important that if the power is sliding left or to the right as the focus is on the efficiency.

The sign of the power difference is obtained by the other state variable SOC. Reward function is also similar with one little difference and defined as:

$$r = -w_{SOC} * (SOC - SOC_{desired})^2 - w_{H_2} * (H_{2,effmax} - H_{2,eff})^2 \quad (7.4)$$

If the state limits are exceeded then the simulation is stopped and the agent gets a penalty.

## VIII. RESULTS

compares the results of the algorithms implemented in the model that are DDPG, DQN, Rule-based and brute force under different drive cycles. Brute force findings provides a target for the best control actions even though it is limited by the discretization of variables. We present that learning techniques are able to produce better outcome than rule-based method and close to brute force algorithm results.The comparison will be made based on total energy consumption, average fuel cell efficiency and SOC deviation. Under the UDDS cycle the agents trained with DDPG and DQN algorithms are able to keep the SOC level between certain limits and the deviation from the target that is set as % 50 is not large. At the end of the cycle it still has an acceptable value and restarting the cycle from that point will not cause any significant change of SOC behaviour as shown in figure 4.2. In addition efficiency of the fuel cell is high and the system tracks the power demand within a very small margin of error. Since DQN action space is discrete, sudden action changes cause slight overshoots. A similar SOC behaviour is also achieved by the rule based strategy, however figure 8.1. shows that many changes in the SOC are sharper than that of the DDPG-trained controller. Another point is that it is possible for the SOC level to drop slightly from the target level, if it is necessary, a situation that cannot be observed in rule-based EMS.

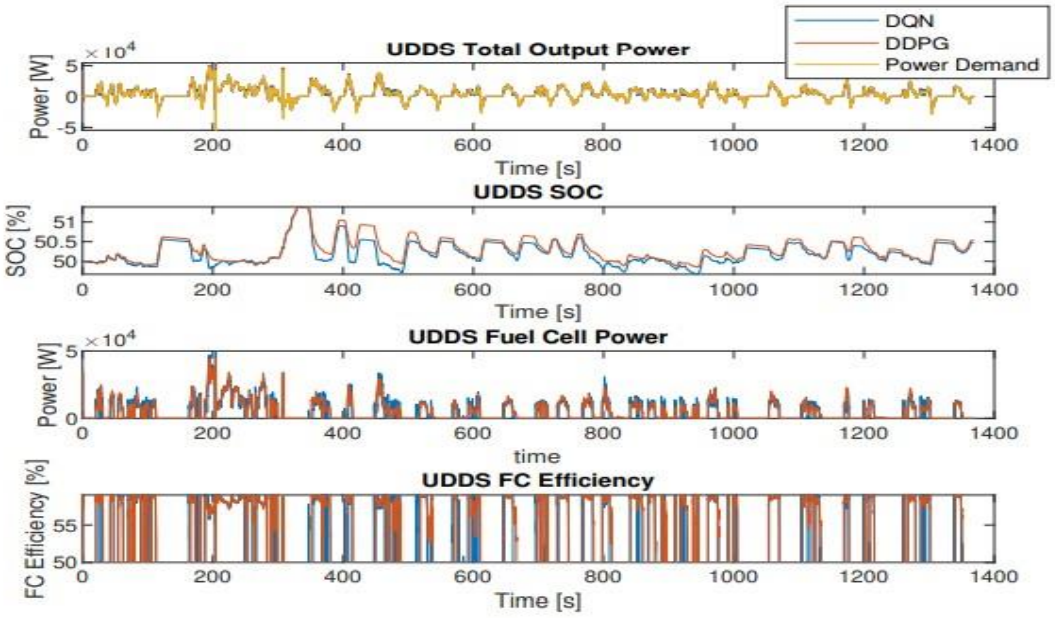


**Figure 8.1 Behavior of the RL-Based EMS trained with DDPG and DQN under the UDDS drive cycle**

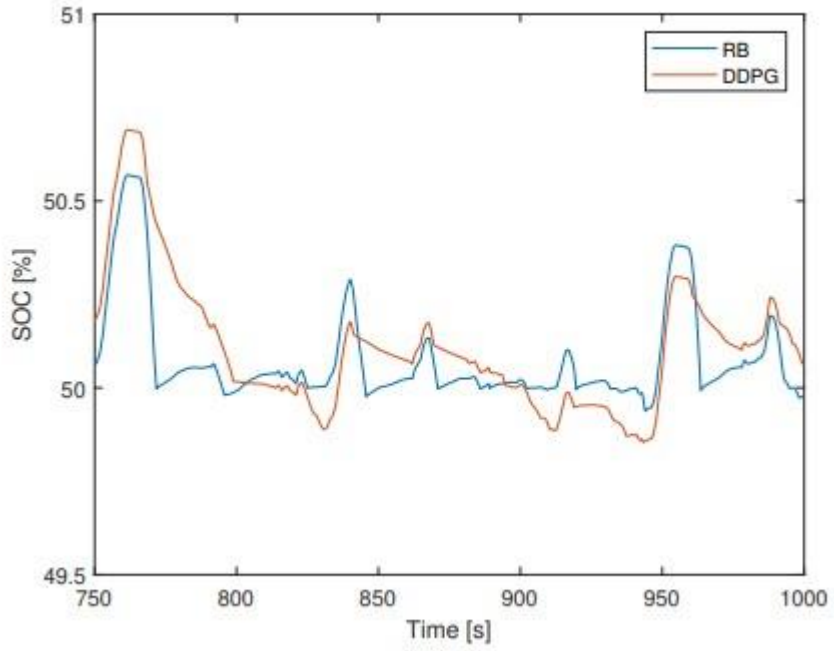


**Figure 8.2 Comparison of SOC behavior of the controller developed with Rule-Based EMS and RLBased EMS trained with DDPG**

## IX. CONCLUSION

The development of a fuel cell vehicle model features a model that simulates a power unit that includes a fuel cell, lithium ion battery, and DC-DC converter. Autonomie software is used to obtain the vehicle load model, which converts input speed to power demands from the converter and all the way up to the electric motor. Voltage outputs ($V_{fc}$ and $V_{bat}$) are produced by both fuel cell and lithium-ion battery models when given input currents. The voltage becomes the input for the DC-DC converters which sends the current signals ($i_{fc}$ and $i_{bat}$) into the energy sources depending on the difference between the source and bus voltage. The energy management system, which produces gains ($F_{bat}$ and $F_{fc}$) and passes them to the voltage current controller, determines how much current to draw from the sources. The controller transmits the switches' duty cycle to the converter according to the gains and voltage differences. The primary goal of this study is to implement reinforcement learning algorithms in the energy management strategy. The conclusion is that the DQN algorithm is the most frequently used model-free reinforcement learning algorithm in energy management systems in HEV, PHEV, and FCHEV, aside from Q-learning. After implementing the algorithm and training the agent with drive cycles, the agent with the highest reward is chosen. Their performance is evaluated under various drive cycles. A similar procedure is followed after implementing another algorithm called DDPG. The continuous action space advantage of this

algorithm makes it highly promising, especially in applications like fuel cell hybrid vehicles. The agent's training will be more time-consuming as a disadvantage. Random step power inputs are substituted for drive cycles during the training process. After the agents have been trained, their performance under drive cycles is measured against energy management strategies that utilize rule-based and optimization-based approaches. Autonomie software was the basis for our rule-based approach, and we picked the optimization-based method as the brute force search algorithm. Total energy consumption, fuel cell efficiency, and battery SOC are the criteria for evaluation. The selection of UDDS, HWFET, and US06 cycles is based on their ability to represent different driver behaviors. It is found that in all of these drive cycles, energy management strategies based on DDPG and DQN are able to consume less energy than the rule based approach while achieving a similar SOC behavior and small deviation in SOC. Their performance is slightly inferior to the BF method.. The DDPG algorithm has the potential to be utilized to find the global optimal, which can be continuously learned in real-time applications.

**REFERENCES**


[1] "Comparison of Fuel Cell Technologies." [Online]. Available: https: //www.energy.gov/eere/fuelcells/comparison-fuel-cell-technologies

[2] "Alternative Fuels Data Center: Fuel Cell Electric Vehicles." [Online]. Available: https://afdc.energy.gov/vehicles/fuel_cell.html

[3] I.-S. Sorlei, N. Bizon, P. Thounthong, M. Varlam, E. Carcadea, M. Culcer, M. Iliescu, and M. Raceanu, "Fuel cell electric vehicles—a brief review of current topologies and energy management strategies," Energies, vol. 14, no. 1, p. 252, 2021.

[4] J. Zhang, L. Zhang, F. Sun, and Z. Wang, "An overview on thermal safety issues of lithium-ion batteries for electric vehicle application," Ieee Access, vol. 6, pp. 23 848–23 863, 2018.

[5] C. Zhang, W. Allafi, Q. Dinh, P. Ascencio, and J. Marco, "Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique," Energy, vol. 142, 10 2017.

[6] W. Nsour, T. Taa'mneh, O. Ayadi, and J. Al Asfar, "Design of stand-alone proton exchange membrane fuel cell hybrid system under amman climate," Journal of Ecological Engineering, vol. 20, no. 9, 2019.

[7] R. S. Sutton and A. G. Barto, Reinforcement learning: An introduction. MIT press, 2018.

[8] B. Mostafa, N. El-Attar, S. Abd-Elhafeez, and W. Awad, "Machine and deep learning approaches in genome: Review article," Alfarama Journal of Basic Applied Sciences, 08 2020.

[9] N. Sulaiman, M. Hannan, A. Mohamed, E. Majlan, and W. Wan Daud, "A review on energy management system for fuel cell hybrid electric vehicle: Issues and challenges," Renewable and Sustainable Energy Reviews, vol. 52, pp. 802–814, 2015. [Online]. Available: https://www.sciencedirect.com/science/ article/pii/S1364032115007790

[10] N. Briguglio, L. Andaloro, M. Ferraro, and V. Antonucci, Fuel Cell Hybrid Electric Vehicles, 09 2011.

[11] Y. Miao, P. Hynan, A. von Jouanne, and A. Yokochi, "Current li-ion battery technologies in electric vehicles and opportunities for advancements," Energies, vol. 12, no. 6, 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/6/1074

[12] X. Li, L. Xu, J. Hua, X. Lin, L. Jianqiu, and M. Ouyang, "Power management strategy for vehicularapplied hybrid fuel cell/battery power system," Journal of Power Sources, vol. 191, pp. 542–549, 06 2009.

[13] N. J. Schouten, M. A. Salman, and N. A. Kheir, "Energy management strategies for parallel hybrid vehicles using fuzzy logic," Control Engineering Practice, vol. 11, no. 2, pp. 171–177, 2003, automotive Systems. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967066102000722

[14] R. K. Ahluwalia, X. Wang, and A. Rousseau, "Fuel economy of hybrid fuel-cell vehicles," Journal of Power Sources, vol. 152, pp. 233–244, 2005. [Online]. Available:https://www.sciencedirect.com/science/article/pii/S0378775305000996

[15] G. Paganelli, T.-M. Guerra, S. Delprat, J.-J. Santin, M. Delhom, and E. Combes, "Simulation and assessment of power control strategies for a parallel hybrid car," Int. J. of Automobile Engineering, vol. 214, pp. 705–717, 07 2000.

[16] C. Musardo, G. Rizzoni, Y. Guezennec, and B. Staccia, "A-ecms: An adaptive algorithm for hybrid electric vehicle energy management," European Journal of Control, vol. 11, no. 4, pp. 509–524, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0947358005710487

[17] Y. Zhou, A. Ravey, and M.-C. Marion-Péra, "Real-time cost-minimization power-allocating strategy via model predictive control for fuel cell hybrid electric vehicles," Energy Conversion and Management, vol. 229, p. 113721, 02 2021.

[18] H. Borhan, A. Vahidi, A. M. Phillips, M. L. Kuang, I. V. Kolmanovsky, and S. Di Cairano, "Mpcbased energy management of a power-split hybrid electric vehicle," IEEE Transactions on Control Systems Technology, vol. 20, no. 3, pp. 593–603, 2012.

[19] C.-C. Lin, H. Peng, J. Grizzle, and J.-M. Kang, "Power management strategy for a parallel hybrid electric truck," IEEE Transactions on Control Systems Technology, vol. 11, no. 6, pp. 839–849, 2003.

[20] R. Wang and S. M. Lukic, "Dynamic programming technique in hybrid electric vehicle optimization," in 2012 IEEE International Electric Vehicle Conference, 2012, pp. 1–8.

[21] W. Zhou, L. Yang, Y. Cai, and T. Ying, "Dynamic programming for new energy vehicles based on their work modes part ii: Fuel cell electric vehicles," Journal of Power Sources, vol. 407, pp. 92–104, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0378775318311571

[22] E. T. Stephen and S. P. Boyd, "Finding ultimate limits of performance for hybrid electric vehicles," SAE Paper, 2000.
[23] X. Hu, L. Johannesson Mårdh, N. Murgovski, and B. Egardt, "Longevityconscious dimensioning and power management of a hybrid energy storage system for a fuel cell hybrid electric bus," Applied Energy, vol. 137, 01 2014.
[24] F. Odeim, "Optimization of fuel cell hybrid vehicles," Ph.D. dissertation, May 2018. [Online]. Available: https://duepublico2.uni-due.de/receive/duepublico_ mods_00046123
[25] W. Li, J. Ye, Y. Cui, N. Kim, S. W. Cha, and C. Zheng, "A speedy reinforcement learning-based energy management strategy for fuel cell hybrid vehicles considering fuel cell system lifetime," International Journal of Precision Engineering and Manufacturing-Green Technology, pp. 1–14. [Online]. Available: https://app.dimensions.ai/details/publication/pub.1140048378
[26] N. P. Reddy, D. Pasdeloup, M. K. Zadeh, and R. Skjetne, "An intelligent power and energy management system for fuel cell/battery hybrid electric vehicle using reinforcement learning," in 2019 IEEE Transportation Electrification Conference and Expo (ITEC), 2019, pp. 1–6.
[27] Y. F. Zhou, L. J. Huang, X. X. Sun, L. H. Li, and J. Lian, "A long-term energy management strategy for fuel cell electric vehicles using reinforcement learning," Fuel Cells, vol. 20, no. 6, pp. 753–761, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/fuce.202000095
[28] L. Guo, Z. Li, and R. Outbib, "Reinforcement learning based energy management for fuel cell hybrid electric vehicles," in IECON 2021 – 47th Annual Conference of the IEEE Industrial Electronics Society, 2021, pp. 1–6.
[29] K. Deng, Y. Liu, D. Hai, H. Peng, L. Löwenstein, S. Pischinger, and K. Hameyer, "Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging," Energy Conversion and Management, vol. 251, p. 115030, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890421012061
[30] P. Wu, J. Partridge, and R. Bucknall, "Cost-effective reinforcement learning energy management for plug-in hybrid fuel cell and battery ships," Applied Energy, vol. 275, p. 115258, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261920307704
[31] P. Wu, J. Partridge, E. Anderlini, Y. Liu, and R. Bucknall, "Nearoptimal energy management for plug-in hybrid fuel cell and battery propulsion using deep reinforcement learning," International Journal of Hydrogen Energy, vol. 46, no. 80, pp. 40 022–40 040, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360319921037745
[32] P. Zhao, Y. Wang, N. Chang, Q. Zhu, and X. Lin, "A deep reinforcement learning framework for optimizing fuel economy of hybrid electric vehicles," in 2018 23rd Asia and South Pacific Design Automation Conference (ASP-DAC), 2018, pp. 196–202.
[33] G. Du, Y. Zou, X. Zhang, T. Liu, J. Wu, and D. He, "Deep reinforcement learning based energy management for a hybrid electric vehicle," Energy, vol. 201, p. 117591, 2020. [Online]. Available: https: //www.sciencedirect.com/science/article/pii/S0360544220306988.
[34] Y. Zou, T. Liu, D. Liu, and F. Sun, "Reinforcement learning-based real-time energy management for a hybrid tracked vehicle," Applied Energy, vol. 171, pp. 372–382, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261916304081
[35] J. Wu, H. He, J. Peng, Y. Li, and Z. Li, "Continuous reinforcement learning of energy management with deep q network for a power split hybrid electric bus," Applied Energy, vol. 222, pp. 799–811, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261918304422
[36] T. Liu, Y. Zou, D. Liu, and F. Sun, "Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle," IEEE Transactions on Industrial Electronics, vol. 62, no. 12, pp. 7837–7846, 2015.
[37] R. Xiong, J. Cao, and Q. Yu, "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle," Applied Energy, vol. 211, pp. 538–548, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261917316707
[38] C. Liu and Y. L. Murphey, "Power management for plug-in hybrid electric vehicles using reinforcement learning with trip information," in 2014 IEEE Transportation Electrification Conference and Expo (ITEC), 2014, pp. 1–6.
[39] H. Shen, Y. Zhang, J. Mao, Z. Yan, and L. Wu, "Energy management of hybrid uav based on reinforcement learning," Electronics, vol. 10, no. 16, p. 1929, 2021.
[40] R. Liessner, C. Schroer, A. M. Dietermann, and B. Bäker, "Deep reinforcement learning for advanced energy management of hybrid electric vehicles." in ICAART (2), 2018, pp. 61–72.
[41] Y. Wu, H. Tan, J. Peng, H. Zhang, and H. He, "Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus," Applied energy, vol. 247, pp. 454–466, 2019.
[42] G. L. Plett, Battery management systems, Volume I: Battery modeling. Artech House, 2015.
[43] J. Pukrushpan, A. Stefanopoulou, and H. Peng, "Modeling and control for pem fuel cell stack system," in Proceedings of the 2002 American Control Conference (IEEE Cat. No.CH37301), vol. 4, 2002, pp. 3117–3122 vol.4.

[44] Y.-X. Wang, K. Ou, and Y.-B. Kim, "Modeling and experimental validation of hybrid proton exchange membrane fuel cell/battery system for power management control," International Journal of Hydrogen Energy, vol. 40, no. 35, pp. 11 713–11 721, 2015.
[45] W. Jiang and B. Fahimi, "Active current sharing and source management in fuel cell–battery hybrid power system," IEEE Transactions on Industrial Electronics, vol. 57, no. 2, pp. 752–761, 2009.
[46] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv preprint arXiv:1312.5602, 2013.
[47] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," arXiv preprint arXiv:1509.02971, 2015.
[48] Argonne National Laboratory, "Autonomie." [Online]. Available: https: //www.autonomie.net
[49] X. Yuan, C. Zhang, G. Hong, X. Huang, and L. Li, "Method for evaluating the real-world driving energy consumptions of electric vehicles," Energy, vol. 141, pp. 1955–1968, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544217319928