HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

Project II

REINFORCEMENT LEARNING FOR MAXIMUM POWER POINT TRACKING OF A PEM FUEL CELL SYSTEM

ĐỖ QUANG THÀNH - Thanh.dq202770@sis.hust.edu.vn
NGUYỄN TUẤN DƯƠNG - Duong.nt202740@sis.hust.edu.vn
NGUYỄN ĐÌNH TOÀN - Toan.nd202775@sis.hust.edu.vn
VÕ KHÔI THÀNH SƠN – Son.vkt202768@sis.hust.edu.vn
Control engineering and Automation

Instructing teacher: Dr. Vu Thi Thuy Nga

Department: Automation
University: Hanoi University of Science and Technology

HÀ NỘI, 2/2024

Acknowledgement

Through this course and research, we have had the chance to learn the basics of a control system. In addition, we are able to cooperate with friends to work on a new and interesting topic which can contribute greatly to society in the future. With this, we want to express our gratitude towards my instructor, Mrs Vu Thi Thuy Nga, as well as our seniors for helping us along the course.

Abstract

Proton exchange membrane fuel cells (PEMFCs) are the most promising fuel cell technology because of their high-power density, low operating temperature, quick startup capability, and low weight. Efficient use of the PEMFC requires keeping it working at an adequate power point and protecting fuel cells from damage problems. Through this course, we learn how to extract the maximum power from the PEMFC system and protect it from membrane damage by stabilizing the hydrogen and oxygen partial pressure. This work proposes a novel approach for controlling fuel cell systems using reinforcement learning (RL) techniques for Maximum Power Point Tracking (MPPT). The RLMPPT control method presented in this study utilizes an MDP model to optimize the fuel cell's operation without requiring prior knowledge of its characteristics. Simulation results demonstrate the effectiveness and versatility of the RLMPPT method across various operating conditions, including changes in temperature, water membrane content, and electrical load. By leveraging RL techniques, this study contributes to advancing MPPT strategies for fuel cell systems, offering a promising avenue for enhancing their energy conversion efficiency and adaptability in real-world applications.

CONTENTS

FIGURE LIST

TABLE LIST

I. Introduction

Fuel cells are one of the environmentally friendly energy sources that generate electricity through an electrochemical reaction. At the moment, there are six main fuel cell types classified according to the electrolyte and fuel used: phosphoric acid fuel cell (PAFC), molten carbonate fuel cell (MCFC), alkaline fuel cell (AFC), solid oxide fuel cell (SOFC), proton exchange membrane fuel cell (PEMFC), and direct methanol fuel cell (DMFC). Among the mentioned types of fuel cell, PEMFC stands out to be the most popular type for mobile and portable applications, due to some advantages such as: heat and water management, a high power density, the reaction of electrode kinetics, alternative catalysts, a low weight, and a low operating temperature. However, PEMFC presents some disadvantages: very high sensitivity to impurities of hydrogen, expensive catalyst and membrane, a gas diffusion layer and flow field layers, degradation, and production difficulties of the membrane electrode assembly.

PEMFC consists of a polymer electrolyte membrane placed in the middle of two electrodes called anode and cathode. Hydrogen fuel is fed through the anode and an oxidant (air or pure oxygen) is pumped into the cathode. Hydrogen molecules are split into electrons and hydrogen protons at the anode catalyst. Hydrogen protons migrate toward the cathode through the membrane and react with the returning electrons and oxygen to produce water and heat. Free electrons at the anode will flow to the cathode through an external load and provide electricity.

PEMFC system control has to take into account problems related to harvesting electrical energy from the PEMFC stack. Fuel cells have a nonlinear voltage–current characteristic, and the power has several local maximum power points in the I–P characteristic under various operating conditions. Therefore, an MPPT algorithm must be established to improve and optimize the PEMFC system efficiency. The problem of fuel starvation as a result of sudden load variations can lead to serious membrane damage in the fuel cell. This problem can be avoided by controlling the inlet flow rates of hydrogen and oxygen to stabilize the partial pressures and protect the fuel cell from damage.

In this research, however, we are considering the source tank in ideal condition, with the pressure at a constant level.

To optimize the fuel cell application performance, several MPPT techniques have been proposed. Among these techniques, Reinforcement Learning involves utilizing feedback from the environment to develop behavioral strategies, making it a potent data-driven approach for tackling intricate control challenges. Recent research has applied Reinforcement Learning technology to address Maximum Power Point Tracking (MPPT) issues in emerging energy systems like wind energy conversion systems, photovoltaic arrays, and hybrid electric vehicles.

The objective of this work is to apply Reinforcement Learning algorithm to fuel cell system to realize MPPT control, by introducing a Markov Decision Process (MDP) model, to optimize the performance of the fuel cell under changing environmental factors. More precisely, it introduces a versatile RLMPPT control approach aimed at monitoring the MPP utilizing the fuel cell current and output voltage of the fuel cell system under standard test conditions. These parameters can be roughly estimated by referencing the fuel cell module's datasheet and analyzing its electrical configuration.

Throughout this work, we managed to show the efficiency of the proposed RLMPPT control method under fluctuating temperature and membrane water content. In addition, it is illustrated through simulations that the controller's functionality remains unaffected by the electrical characteristics of the PV source. Finally, the RLMPPT control method is capable of adapting to dynamic electrical loads.

II. PEM fuel cell system modeling

2.1. PEM fuel cell

The FC output voltage can be described as follows:

$$V_{cell} = E_{Nernst} - \eta_{act} - \eta_{ohmic} - \eta_{con} \tag{1}$$

E_{Nernst} is the reversible open-circuit voltage, it is described by the Nernst equation as:

$$E_{Nernst} = 1.229 - 8.5 \times 10^{-4}(T - 298.15) + 4.308 \times 10^{-5}\left(lnP_{O_2} + lnP_{H_2}\right) \tag{2}$$

where P_{H_2} is the hydrogen partial pressure (atm), P_{O_2} is the is oxygen partial pressure (atm) and T is the absolute temperature (K).

η_{act} is the activation voltage drop, it is given in the Tafel equation as:

$$\eta_{act} = \xi_1 + \xi_2 T + \xi_3 T lnC_{O_2} + \xi_4 T lnI_{FC} \tag{3}$$

where I_{FC} is the fuel cell current (A), and ξ_i (i=1-4) are parametric coefficients for each cell model. C_{O_2} represents the concentration of dissolved oxygen in the interface of the cathode catalyst which can be calculated as:

$$C_{O_2} = \frac{P_{O_2}}{(5.08 \times 10^6)e^{\frac{-498}{T}}} \tag{4}$$

η_{ohmic} is the overall ohmic voltage drop, it can be expressed as:

$$\eta_{ohmic} = I_{FC}R_m \tag{5}$$

where R_m is the ohmic resistance and given by:

$$R_m = \frac{r_m t_m}{A} \tag{6}$$

where A is the cell active area (cm2), *tm* is the membrane thickness (cm). r_m is the membrane resistivity (Ωcm) to proton conductivity and can be calculated as:

$$r_m = \frac{181.6[1 + 0.03\left(\frac{I_{FC}}{A}\right) + 0.0062(T/303)^2(\frac{I_{FC}}{A})^{2.5}]}{[\lambda_m - 0.634 - 3(\frac{I_{FC}}{A})]e^{4.18(T-303/T)}} \tag{7}$$

where λ_m represent the membrane water content and it is a function of the average water activity a_m:

$$\lambda_m = \begin{cases} 0.043 + 17.81a_m - 39.85{a_m}^2 + 36{a_m}^3 \ if \ 0 < a_m < 1 \\ 14 + 14(a_m - 1) \ if \ 1 < a_m < 3 \end{cases}$$

(8)

The average water activity is function of the cathode water vapor partial pressure $P_{v,ca}$,the anode water vapor partial pressure $P_{v,an}$ and the saturation pressure of water P_{sat}. It can be expressed as:

$$a_m = \frac{1}{2}(a_{an} + a_{ca}) = \frac{1}{2}\frac{P_{v,an} + P_{v,ca}}{P_{sat}}$$

(9)

P_{sat} can be obtained using the following empirical expression:

$log_{10}P_{sat} = -2.1794 + 0.02953T - 9.1813 \times 10^{-5}T^2 + 1.4454 \times 10^{-7}T^3$

(10)

η_{con} is the concentration voltage drop, it is expressed as:

$$\eta_{con} = -\frac{RT}{nF}\ln\left(1 - \frac{I_{FC}}{i_L A}\right)$$

(11)

where F is the Faraday's constant, n is the number of electrons participating in the reaction, i_L is the limiting current and R is the universal gas constant.

The output voltage of a fuel cell stack constitutes by N_{FC} fuel cells connected in series is given by:

$$V_{st} = N_{FC}V_{cell}$$

(12)

2.2. System modeling

The PEM fuel cell system adopted in this study is shown in Fig. 1.

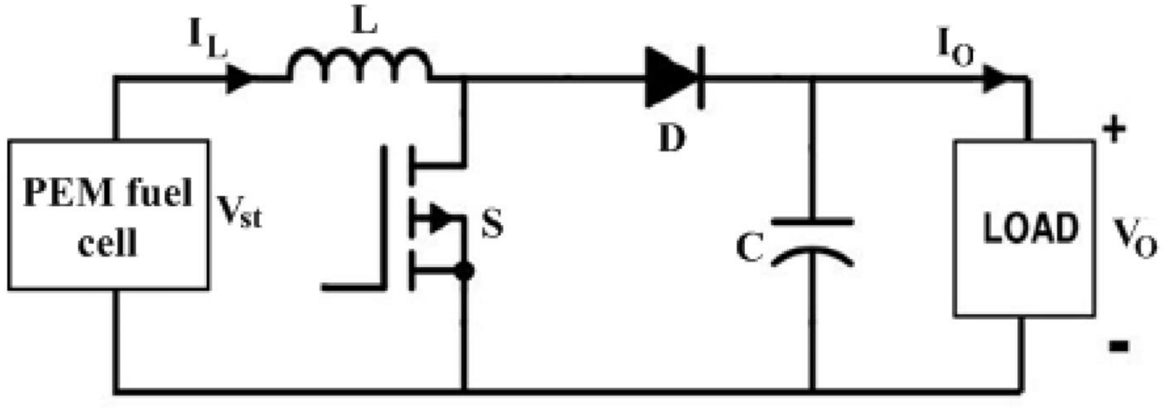


Figure 1. The proposed system configuration

It is constituted by an FC stack, a DC/DC boost converter, and a resistive load. The boost converter is used to increase the system's efficiency by controlling the fuel cell system's operation point through adjusting the duty cycle of the converter. The specification details of the PEM fuel cell system is given in *Table 1*.

Table 1. Specification details of the PEM fuel cell system

Parameter	Value
$\boldsymbol{N_{FC}}$	24
A	232 cm^2
ξ_1	0.944
ξ_2	-0.00354
ξ_3	-7.8×10^{-8}
ξ_4	1.96×10^{-4}
n	2
$\boldsymbol{i_L}$	2 Acm^2
T	335 K
C	7000×10^{-6} F
L	29×10^{-3} H
R	8314.47 j$(kmol\ K)^{-1}$
F	96484600 $Ckmol^{-1}$

The dynamic equation of the system can be expressed as follows:

$$\begin{cases} \dot{I}_L = \dfrac{V_O}{L}(u-1) + \dfrac{V_{st}}{L} \\ \dot{V}_O = \dfrac{I_L}{C}(1-u) - \dfrac{V_O}{RC} \end{cases} \tag{13}$$

where I_L and V_o are the inductor current and the voltage at the output terminals of the boost converter, u is the duty ratio. It is assumed that I_L is equal to I_{FC}.

Eq. (13) can be written as:

$$\dot{x} = F(x,t) + G(x,t)u(t) \tag{14}$$

$$where \begin{cases} x = [I_L \ \ V_o]^T \\ F(x,t) = [\dfrac{V_{st} - V_o}{L} \ \ \dfrac{I_L}{C} - \dfrac{V_O}{RC}]^T = [F_1 \ \ F_2]^T \\ G(x,t) = [\dfrac{V_O}{L} \ \ -\dfrac{I_L}{C}]^T = [G_1 \ \ G_2]^T \end{cases} \tag{15}$$

The fuel cell output power depends on the load and the operating conditions like air pressure, oxygen partial pressure, cell temperature, and membrane water content. Using

MATLAB simulation with various values of temperature and membrane water content (Resistive load, oxygen partial pressure, and hydrogen partial pressure are regulated respectively to 50 Ω, 2 atm, and 2 atm.), we obtain the power-current characteristics of the fuel cell system in fig. 3:

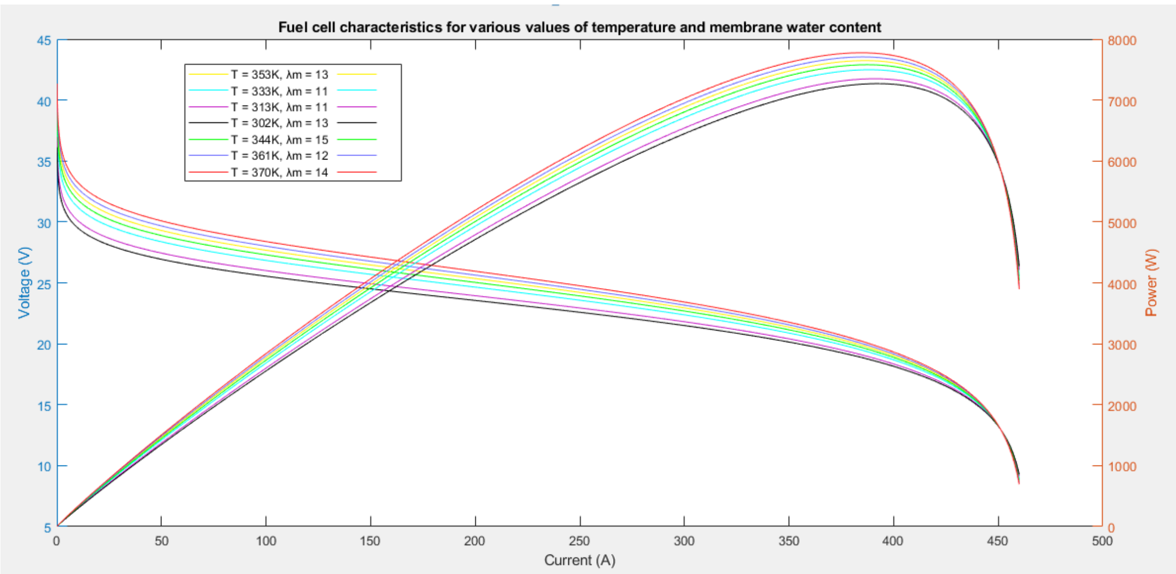


Figure 2. Fuel cell characteristics for various values of temperature and membrane water content

These curves show the nonlinear characteristic of the fuel cell system, and the power has several local maximum power points (MPP) in the P–I characteristic under variation of cell temperature and membrane water content. Thereby MPP tracking should be used to track its changes.

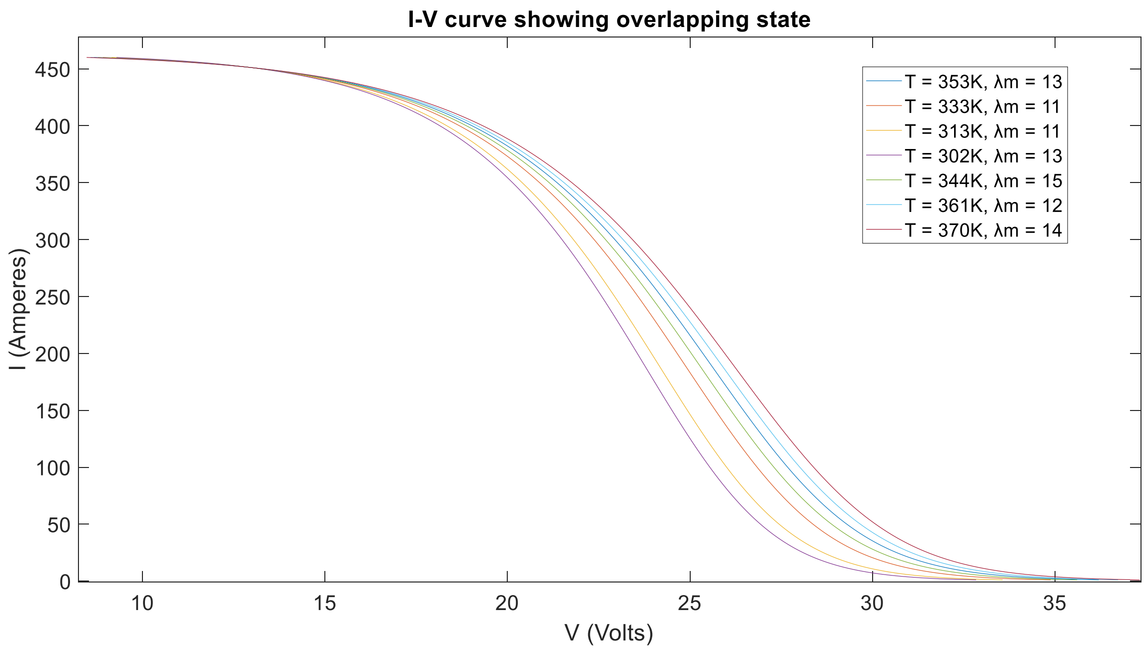
2.3. MPPT control problem

The fuel cell operation point refers to the power generated by the fuel cell system, determined by the current (I_{fc}) and voltage (V_{fc}) it produces at any given time. This creates a current against voltage graph (I-V) curve for constant environmental conditions. The Maximum Power Point (MPP) is a specific point on this curve where the FC system generates the most power.

When a load is connected to the fuel cell system, the operating point and generated power are influenced by the resistance of the load. If the load's resistance matches the ratio of voltage to current at the MPP, i.e. $R_{LMPP} = V_{MPP}/I_{MPP}$, the operating point aligns with the MPP, and MPP tracking is unnecessary. However, if a different load with a different resistive load is connected, the operating point will differ to the MPP. In such cases, the fuel cell system doesn't produce its maximum power, and MPP tracking becomes necessary.

Achieving Maximum Power Point Tracking (MPPT) is essential for the efficient operation of fuel cell sources because, in most applications, the resistance of the load doesn't match the value that corresponds to the Maximum Power Point (MPP) under different environmental conditions. Additionally, the electrical load often varies dynamically, further complicating the task. In cases even when MPPT task has been performed, any change in the load will shift the operation point away from the MPP, Therefore, the MPP must be tracked continuously. Additionally, the environmental conditions are not constant, causing the I-V curve of the fuel cell to change accordingly.

The slope of the load line is defined by the resistance value of the load and is equal to Slope = $1/R_L$. The operation point for each I-V curve is the point that the load line intersects each curve. The load line represents the relationship between the current and voltage of a resistive load. The value of the resistance that corresponds to the MPP at this condition does not match the MPP at other conditions, therefore the Slope of the load line has to change in order to move the operation point to the MPP. This can be achieved by connecting a DC/DC converter between the fuel cell source and the load.



The DC/DC boost converter has the ability to move the operation point of the fuel cell source by changing the slope of the resistive load line:

$S = \frac{1}{R_L(1-D)^2}$, where D is the duty cycle of the converter.

III. Reinforcement Learning-Based Maximum Power Point Tracking (RLMPPT) Control Method

This research introduces a novel approach utilizing reinforcement learning to address the maximum power point tracking problem in fuel cell systems. The RL methodology allows problem-solving without prior knowledge of the source behavior or predefined dynamics. The RL algorithm is designed to learn the system behavior or optimal system configuration based on the response of the source. Implementation of the reinforcement learning approach on the fuel cell operation necessitates the definition of a Markov Decision Process (MDP) model representing the source's behavior.

The MDP is formally represented as a tuple (S, A, T, R), where S is a finite set of states describing the operating point of the fuel cell, A is a finite set of actions applicable to the converter to alter the state of the fuel cell, T is a transition function defined as T: S x A x S → [0; 1], and R is a reward function defined as R: S x A x S → $\mathbb{R}$. The transition function, T, indicates the transition probability from state s ∈ S to a new state s' ∈ S if action a ∈ A is applied, and R represents the received reward from the same action and state transition. An MDP forms a sequential model considering a sequence of state transitions based on implemented actions. The objective is to extract a policy that "solves" the MDP and identifies an optimal policy maximizing payoff. The following sections define an MDP model for the fuel cell MPPT control problem and introduce a reinforcement learning algorithm.

3.1. State space:

A state describes the system’s condition in any given instance and is crucial for the performance of any MDP solving algorithm such as reinforcement learning. When a state space is defined, a number of issues need to be taken into consideration. A state should be descriptive enough and include all required information to describe a system’s condition and allow decision making. However, excess information can lead to a very large state space that can be intractable. On the other hand insufficient information may obstruct the system’s ability to discriminate between states, leading to inadequate decision-making abilities, oscillating between states and to non-optimal policies.

Additionally, an MDP needs to hold the Markov property which is also required for RL algorithms in order to converge to an optimal policy. That means that the transition probability from a state depends on information of this state and the applied action only, and not on historic data.

3.2. Actions:

The action list A comprises a finite set of actions applicable to a fuel cell to alter the system's operation. In the context of an MPPT control problem, an action is defined as a change in the DC/DC converter duty cycle, affecting the produced power. The duty cycle, ranging from 0 to 1, has different optimal values for various operating conditions and sources. While a continuous action list would provide high accuracy, it would lead to computational challenges. To maintain computational efficiency, this work defines a discrete finite action list A according to specified rules:

1) The actions need to include positive and negative changes.
2) The smallest change needs to provide enough resolution to achieve maximum power.
3) An action of zero change needs to be included in order to avoid oscillations between states.

3.3. Return and reward:

For every applied action the system reacts and makes a state transition, generating a response that is monitored as a return from the "environment". The system's return is then structured as a reward in order to correlate it to actions. For the MPPT control problem we propose to use the following function:

$$reward = \begin{cases} \omega_p \dfrac{\Delta P}{\Delta t} & , \dfrac{\Delta P}{\Delta t} < 0 \\ \omega_n \dfrac{\Delta P}{\Delta t} & , \dfrac{\Delta P}{\Delta t} \geq 0 \end{cases}$$

Where, $\frac{\Delta P}{\Delta t}$ is the discrete change of the power P in one time cycle. The reward is built to be non-symmetrical in order to clearly separate between positive and negative states, and weights ω_p and ω_n are introduced for that reason as well as to enhance the positive effect of an action. Therefore, weights must differ ($\omega_p \neq \omega_n$). The proposed reward follows the change of the produced power to indicate a positive or negative effect of actions and is always proportional to the change. It aims to reflect the operational quality of the PEMFC source receiving positive, negative and zero values in cases where the fuel cell power increases, reduces, or remains unchanged, respectively. When a fuel cell is stabilized at a maximum power, the reward becomes zero as the power change is neither positive nor negative. In that way the algorithm will avoid oscillations between states once it has converged.

Additionally, the proposed reward is independent of any particular type of PEMFC Source and the MPP. That allows seamless application of the proposed method to different systems without knowledge of the system characteristics. The reward will also work for different operating conditions while the MPP changes.

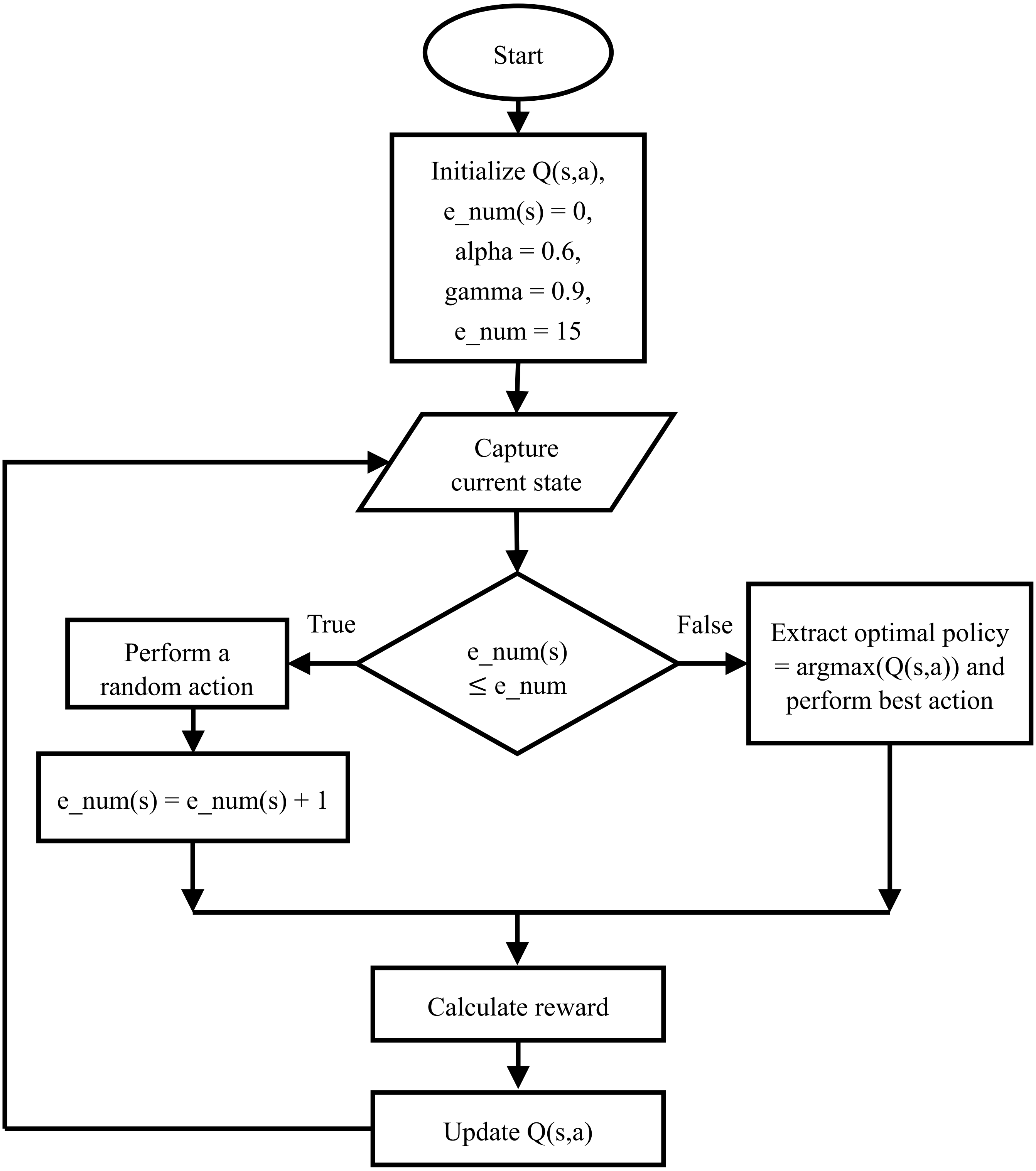
3.4. Reinforcement learning algorithm

Reinforcement learning algorithms aim to learn the behavior of a system or a policy that will optimize the system's performance by interacting with its environment. The learned system is modeled as an MDP and RL algorithms have been linked to MDP model solving where the system's model and transition probabilities may not be known. Such algorithms aim to learn the system's transition probabilities and then plan an optimal policy, or learn the optimal policy directly according to its behavior (returns). Learning a model requires large amount of data and lot of testing which is not always possible in real systems. Learning a policy can be a lot more data efficient and fast while it can also provide results online.

The MPPT control problem is a deterministic problem since the transitions will be the same ($\mathrm{T}: \mathrm{S} \times \mathrm{A} \rightarrow S$) for every state action combination under the same environmental conditions. Additionally, the problem is time invariant and the same power is produced under the same operating conditions. That means a deterministic policy can be learned and the implemented exploration strategy does not have to be constant or periodic in order to keep the model up to date. It is therefore proposed to randomly explore the state-action space before the policy exploitation. The algorithm explores for a certain number of rounds based on the action list size. The number of exploration rounds is described as $e_{num} = a_{num} * m$, where e_{num} is the number of defined actions and m is a multiplier defined heuristically, that should be large enough for the algorithm to have enough state-action visits. However, it should not be too large, to keep the method computationally efficient.

Once the whole state action is explored the algorithm can be said to converge to an optimal policy. For the RLMPPT control problem it is chosen to randomly explore all possible actions for a predetermined number of times. After that, the controller greedily chooses the action with the best state action-value. Additionally, an RL algorithm can be on-policy or off-policy, depending whether the policy is used while it is learned or not. For the MPPT control problem and on-policy algorithm is used for fast results that can adapt while the operating conditions and the MPP changes. The RLMPPT controller starts by exploring the different state actions until m number of exploration rounds per states are completed. In other words, when the counter of random actions per state $e_{num}(s)$ becomes equal to e_{num}. This is not done at once but every time an unexplored or partially explored state occurs, a random action is chosen. State parameters are presented in the previous section and the reward is computed. A Q-learning update rule is used to calculate the state action values, with a being the learning rate and γ the discount factor. Big learning rates will allow faster convergence, while however might cause oscillations between non-optimal values. Therefore, a small a is chosen, that will allow convergence. The discount factor will indicate the significance of future reward and therefore a close to 1 value should be chosen. Finally, a policy is exported based on the optimal state action combination.

Q-learning flow chart:



IV. RLMPPT verification and evaluation

To assess the effectiveness of the proposed RLMPPT control method, a series of simulations were conducted under varied operating conditions. The objective was to evaluate the efficiency of the RLMPPT control method in terms of maximizing power output, as well as its ability to seamlessly adapt to different scenarios. The power circuit and controller were simulated using Matlab. The controller generates changes in the duty cycle between consecutive time steps, which are then converted into a control signal that drives the switch of the boost converter. In the following, first the implementation of the proposed RLMPPT control method is presented, followed by simulations and evaluation results.

4.1. RLMPPT control method implementation

To run the simulations, the RLMPPT control method presented in Section 3 is implemented. The V_{fc} state variable is discretized in 90 values and the same discretization is implemented for state variable I_{fc}. To be more precise, the V_{fc} state variable is discretized between 0 and 18 V with a discretization value of 0.2 V and state variable I_{fc} is discretized between 0 and 450 A with discretization value of 5 A. The above settings form a state space of 8100 states. In addition, the action list is composed of 7 actions and defined as A = {-0.0002; -0.02; -0.6; 0; 0.6; 0.02; 0.0002}. This results to a state action space of 56700 state actions. The action list is following the action definition rules presented in Section 3. Finally, the reward is calculated using the presented equation above for $w_p = 1$ and $w_n = 4$.

For the following set of simulations, the multiplier m determining the number of times each action is performed in a certain state is equal to 25 and enum becomes equal to 15. Thus, the exploration "revisits" each action twenty-five times (m = 25) in every state regardless the time of occurrence.

4.2. Simulations and results

Below, we outline specific scenarios aimed at validating the implementation of the RLMPPT control method and assessing its efficiency and effectiveness across various operating conditions

4.2.1 Fuel cell power test:

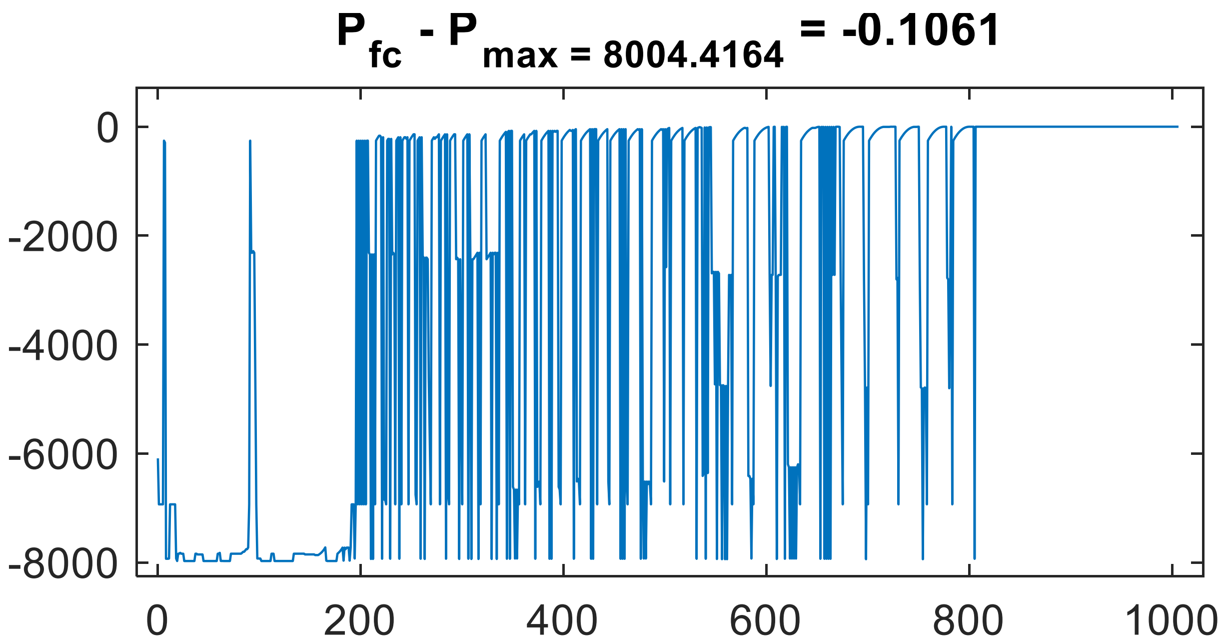


Figure 3. Test condition T = 400, lambda = 12

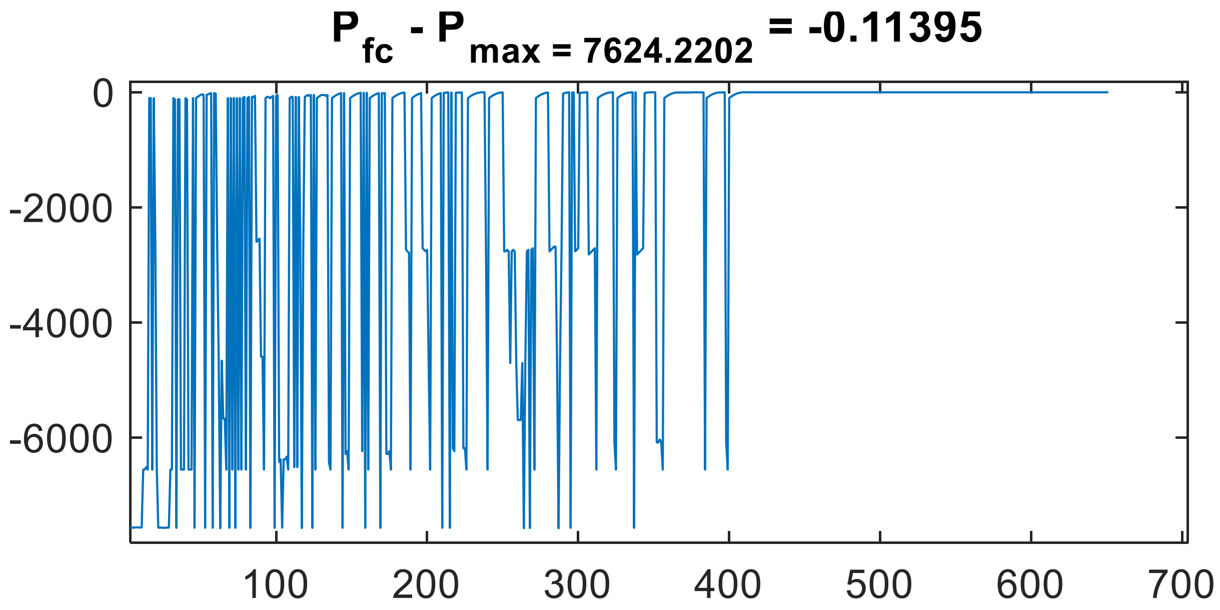


Figure 4. Test condition T = 350, lambda = 12

4.2.2 Temperature input signal:

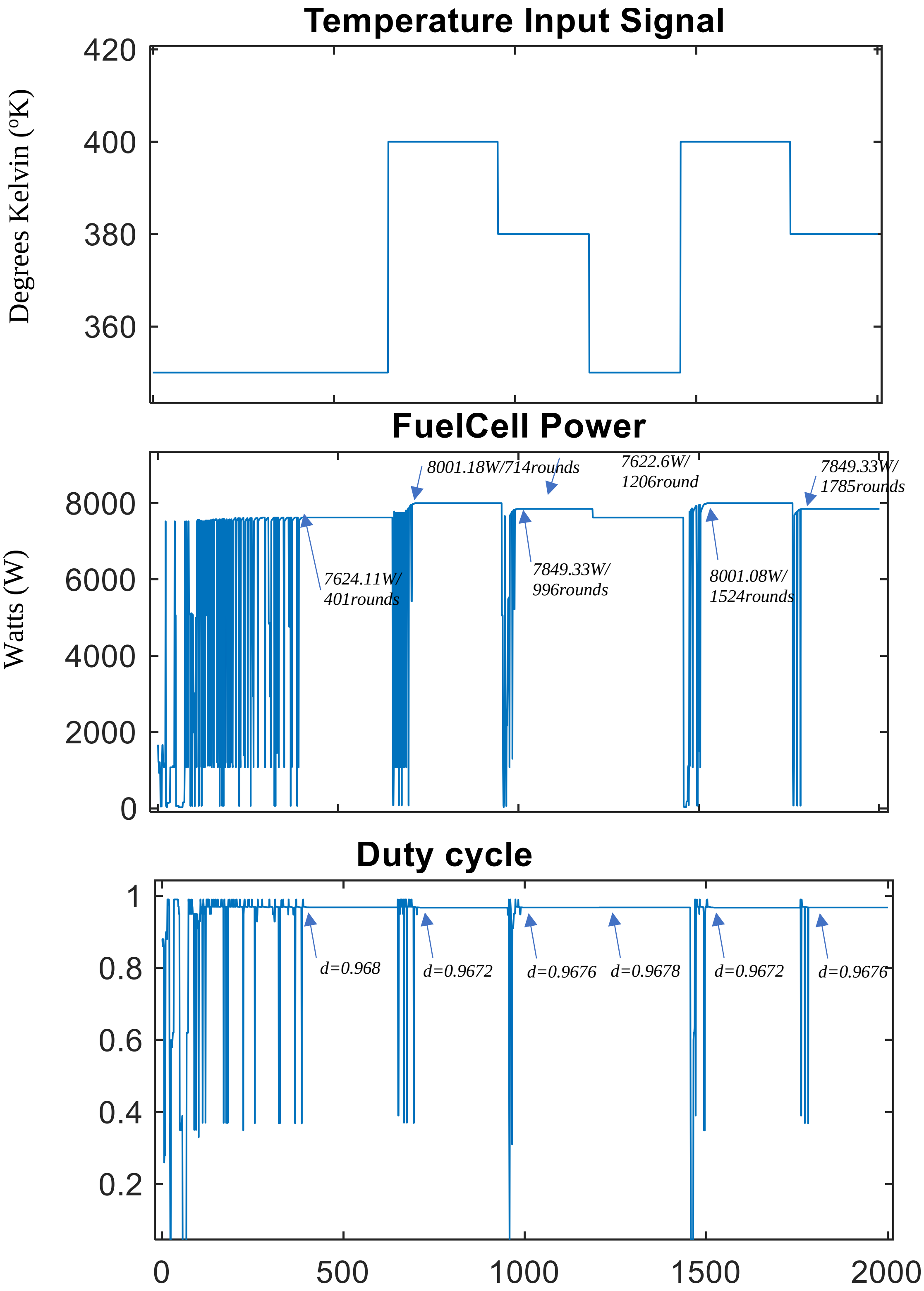


Figure 5. FC performance during varying temperature a)Inpout signal b) FC produced power c) Duty cycle

4.2.3 Water-membrane input signal:

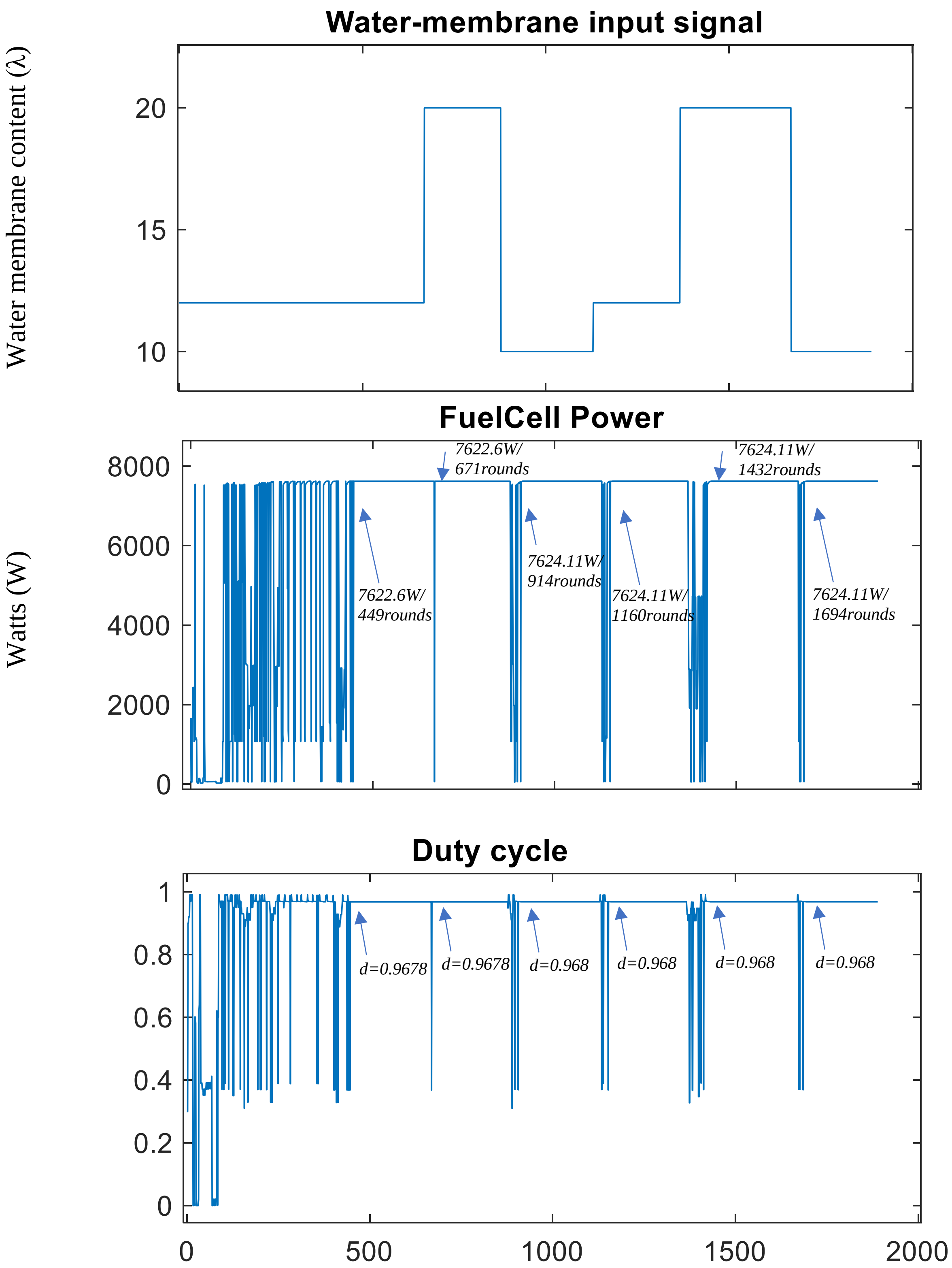


Figure 6. FC performance during varying water-membrane a)Inpout signal b) FC produced power c) Duty cycle

4.2.4 State parameters I_{fc} and V_{fc} close to MPP

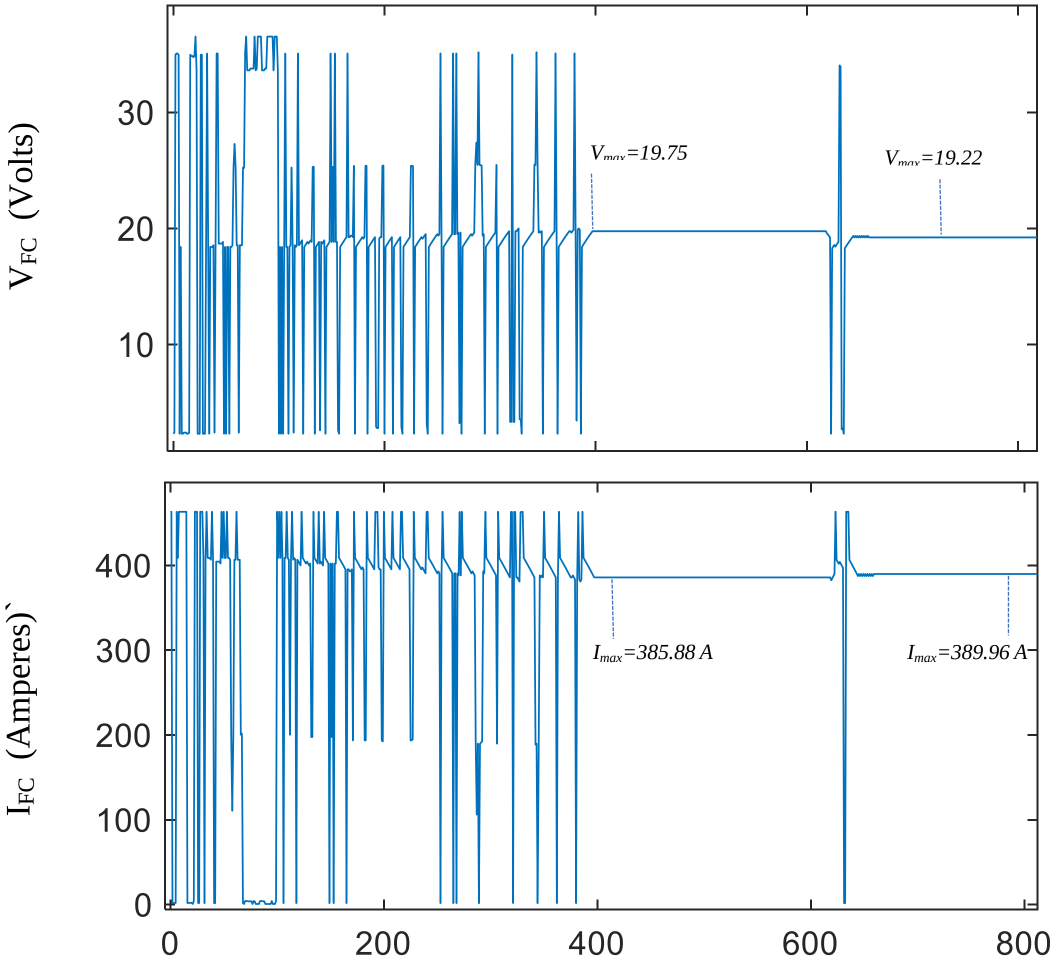


Figure 7. I_{fc} and V_{fc} with λ=15, T= 350°K ➔ λ=15, T= 333°K

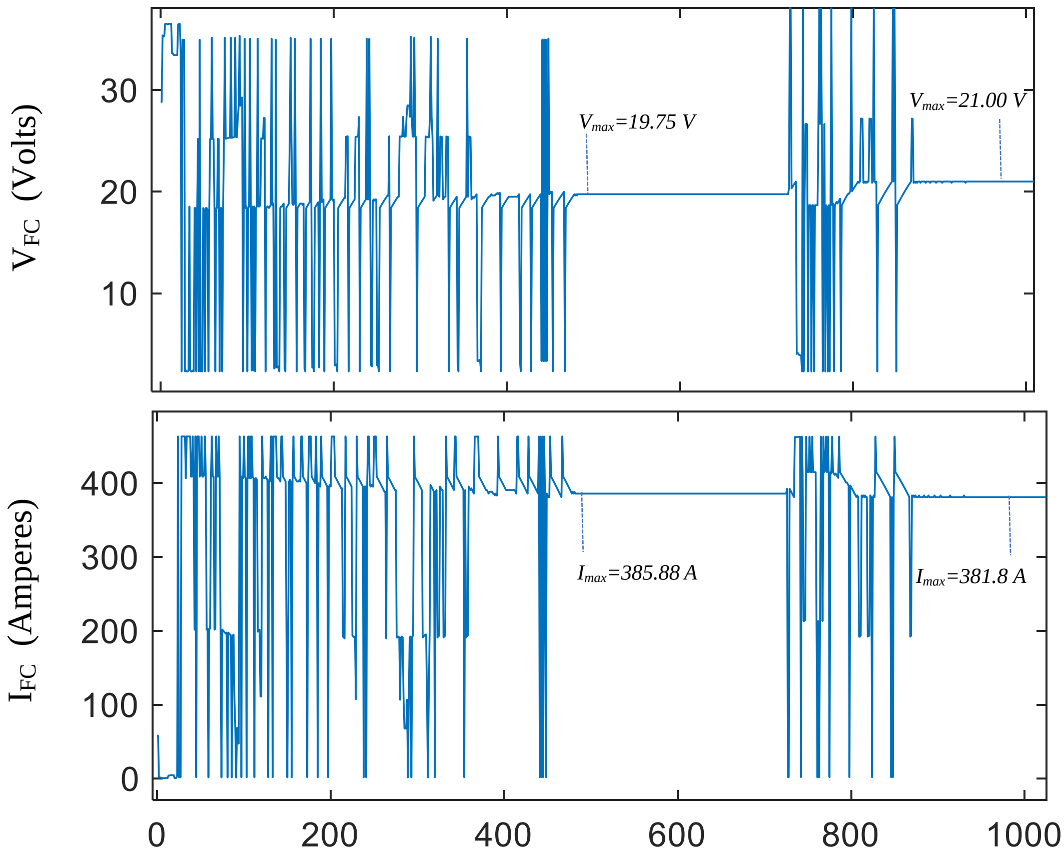


Figure 8 I_{fc} and V_{fc} with λ=15, T= 350°K ➔ λ=15, T= 400°K

4.2.5 Varying load and environment conditions

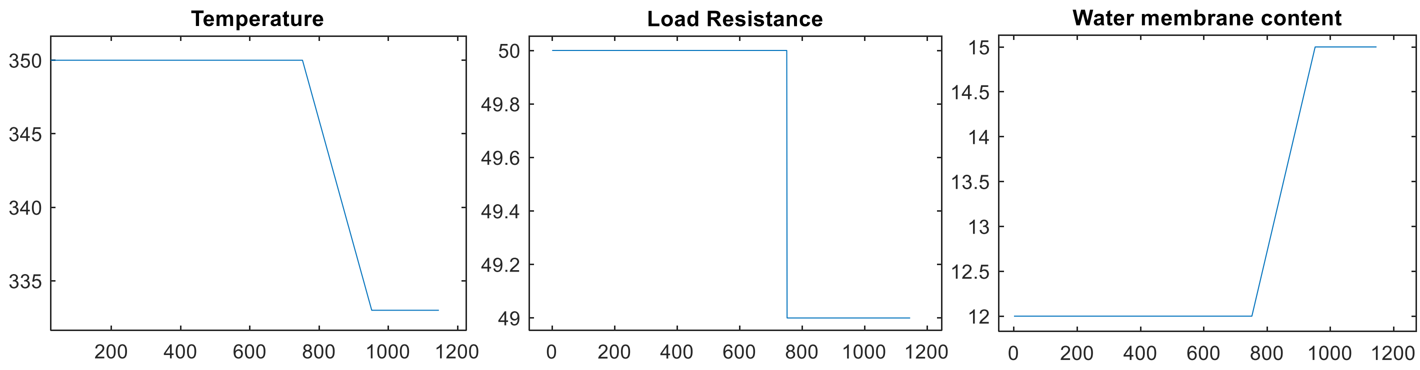


Figure 9. a) FC Temperature b) Load resistance c) FC Water membrane content

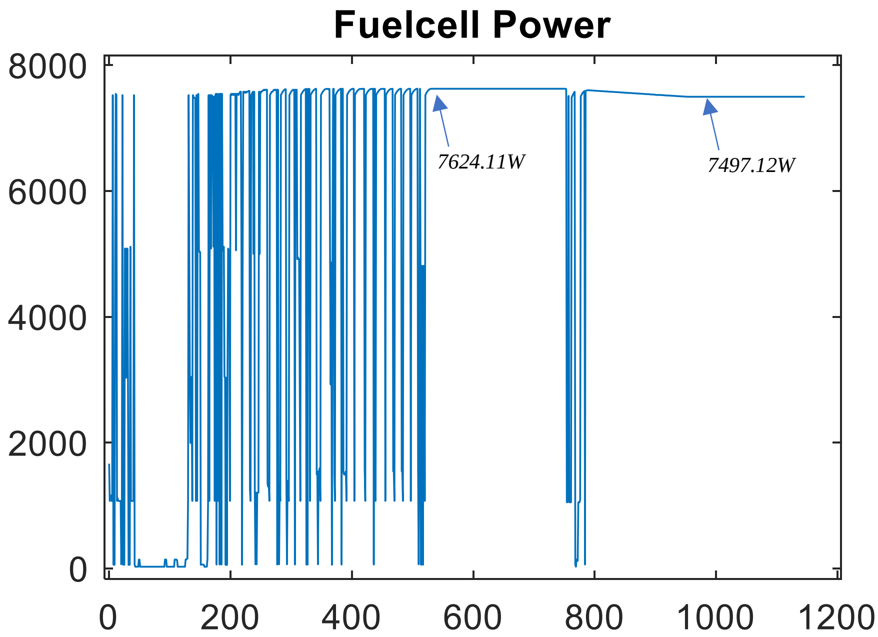


Figure 10 d) FuelCell power produced

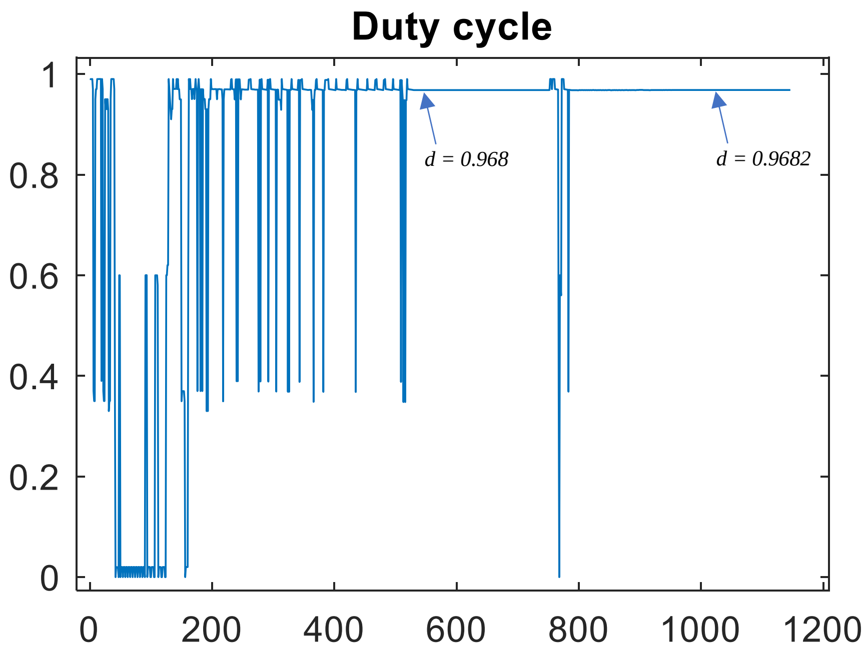


Figure 11 e) Duty cycle

V. Conclusion

Following our comprehensive research efforts, we have achieved milestones in the development of a fuel cell model capable of accurately simulating its performance across a wide range of temperatures and membrane water content levels. This study introduces an MDP model to address the MPPT control challenge in fuel cell sources using reinforcement learning. The proposed RLMPPT control method relies on two distinct state parameters capable of identifying whether an operating point is close to the Maximum Power Point (MPP). The algorithm developed for this MDP can operate optimally across diverse operating conditions without the need for additional configuration. Through implementation and simulation across various scenarios, the efficiency and reliability of the RLMPPT control method are assessed. Simulation results demonstrate very good performance under varying conditions including temperature, membrane water content, and electrical load.

This study has shown promising outcomes in developing a universal MPPT control approach that is not reliant on specific fuel cell characteristics. Nonetheless, there remains ample room for future investigation in this area. Subsequent research could target state space reduction to identify unvisited or redundant states. Additionally, more focus should be placed on optimizing the reinforcement learning algorithm. Conducting a comparative analysis between various RL algorithms would be helpful in identifying the most effective method for addressing the MPPT control challenge. Enhancing the exploration strategy could potentially enhance algorithm performance in terms of efficiency and time.

VI. Future works

Moving forward, our focus will be on refining and optimizing the adaptive sliding mode control mechanism, ensuring it can adapt seamlessly to dynamic changes in operating conditions, thereby enhancing the fuel cell's stability and efficiency. Furthermore, to align the simulation results with real-life scenarios, we aim to integrate a pressure control mechanism into the fuel cell system. By accounting for the influence of pressure on the cell's performance, we can obtain more accurate and reliable simulations, enabling us to make informed decisions during the design and operational stages.

Moreover, our vision extends beyond the realm of simulations, as we aspire to bring our research to practical fruition. In real-life conditions, fuel cell technologies play a vital role in sustainable energy solutions. Therefore, we will be dedicated to translating our theoretical advancements into practical applications, contributing to the wider adoption of fuel cell technologies, and fostering a more sustainable future.

REFERENCE

[1] Ahmed Souissi, Energy Reports, Adaptive sliding mode control of a PEM fuel cell system based on the super twisting algorithm, 2021.

[2] Na, W.K., Gou, B., Feedback linearization based nonlinear control for PEM fuel cells. IEEE Trans. Energy Convers, 2008.