Modeling and Implementation of a Proximal Policy Optimization Algorithm for Non-Inverting Buck-Boost Converter Control

Ramanantsihoarana Harisoa Nathalie[1]; Rastefano Elisée[2]

[1;2]École Doctorale Sciences Et Techniques De l'Ingénierie et De l'Innovation EDSTII,
EAD SDE Antananarivo, Madagascar

Publication Date: 2025/07/15

Abstract: This paper presents a novel approach to controlling non-inverting buck-boost converters using Proximal Policy Optimization (PPO) algorithm for renewable energy applications, particularly photovoltaic systems. Traditional PID controllers face significant limitations when dealing with the complex nonlinear dynamics, external disturbances, and varying operating conditions inherent in renewable energy systems. The proposed PPO-based control strategy addresses these challenges by providing adaptive and intelligent control capabilities. Through comprehensive simulation and experimental validation, we demonstrate that the PPO algorithm successfully learned optimal control policies within 10,000 episodes, maintain excellent voltage regulation under various operating conditions. The results confirm the effectiveness of the proposed approach in maintaining stable output voltage regulation under varying load conditions, input voltage fluctuations, and temperature variations.

Keywords: *Buck-Boost Converter, Proximal Policy Optimization, Deep Reinforcement Learning, Power Electronics Control, Renewable Energy Systems.*

How to Cite: Ramanantsihoarana Harisoa Nathalie; Rastefano Elisée (2025) Modeling and Implementation of a Proximal Policy Optimization Algorithm for Non-Inverting Buck-Boost Converter Control. *International Journal of Innovative Science and Research Technology,* 10(7), 701-710.
https://doi.org/10.38124/ijisrt/25jul480

I. INTRODUCTION

Power conversion systems in renewable energy applications, particularly photovoltaic systems, face increasing challenges in terms of stability and optimal performance. Non-inverting buck-boost converters, widely used in these applications for their ability to handle significant input voltage variations, exhibit complex and highly nonlinear dynamics that make their control particularly challenging [1,2]. Traditional control approaches based on PID regulators, while effective under nominal conditions, quickly show their limitations when faced with external disturbances, load variations, and changing environmental conditions, notably temperature fluctuations and solar irradiation variations characteristic of photovoltaic systems [3,4]. In this context of increasing complexity, research has oriented toward the application of Reinforcement Learning (RL) techniques for power converter control. The Proximal Policy Optimization (PPO) algorithm, developed as an improvement over traditional policy gradient methods, presents significant advantages in terms of learning stability and robustness against parametric variations [5,6]. Unlike classical control approaches that require precise system modeling, reinforcement learning-based methods can autonomously adapt to converter dynamics by directly exploiting measured data [7,8]. Recent work in the field has demonstrated the effectiveness of PPO algorithms for DC-DC converter control, with superior performance in terms of transient stability and recovery time compared to traditional proportional-integral (PI) controllers [9,10]. However, the specific application to non-inverting buck-boost converters in photovoltaic systems presents particular challenges related to dual operation (buck and boost modes), rapid variations in lighting conditions, and energy efficiency requirements [11,12]

The emergence of artificial intelligence in power system control opens new perspectives for overcoming the limitations of conventional methods, particularly in the context of renewable energies where adaptability and robustness are critical requirements for ensuring optimal performance of the conversion system [15,16]. Comparative analysis reveals that the PPO algorithm presents significant advantages for controlling non inverting buck-boost converters. Deep reinforcement learning-based controllers offer a promising alternative for improving the dynamic behavior and efficiency of power electronic converters, particularly in uncertain environments where conventional controllers based on small-signal models show their limitations [17]. PPO distinguishes

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

itself through its real-time learning capability without requiring precise environment modeling, which is essential for photovoltaic systems facing unpredictable variations [18]. Compared to other reinforcement learning algorithms, PPO is relatively easy to implement and requires fewer computational resources, making it particularly suitable for real-time control applications [19]. The PPO algorithm also presents excellent robustness against variations in operating conditions, using qualitative feedback that indicates whether the action taken is correct or not, unlike supervised approaches that require training data with precise outputs [20].

II. METHODOLOGY

A. Buck-Boost Converter Modeling

Fig.1 presents the circuit of a non-inverting buck boost converter.

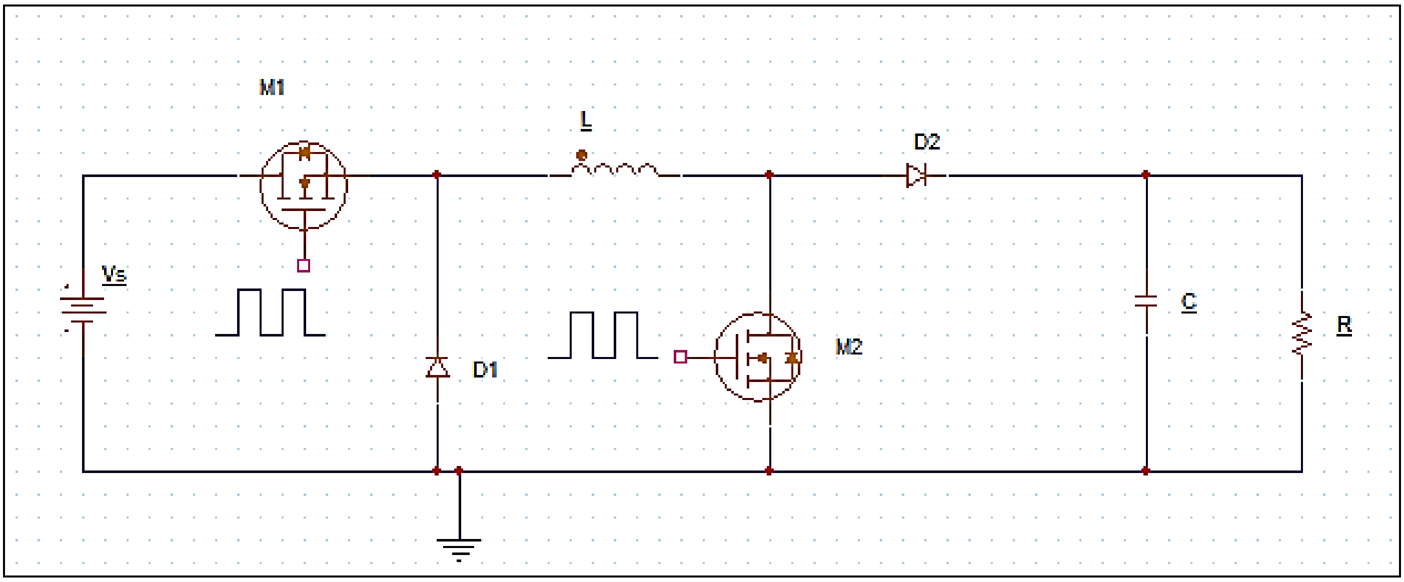

Fig 1 Non Inverting Buck-Boost Converter

Given that the non-inverting buck-boost converter combines the conversion principles of both Buck and Boost, the adopted method consists of separately analyzing each of these two operating modes. We first establish the state representations specific to buck and boost configurations, considering their respective topologies and characteristic dynamic equations. The general state representation is expressed in matrix form as (1).

$$\begin{cases} \dot{X} = Ax + Bu \\ Y = Cx + Du \end{cases} \tag{1}$$

Where X represents the state vector, U the input (here the supply voltage), Y the output (here the output voltage), and A, B, C, D are the characteristic matrices of the system.

➤ *Buck Mode*

During transistor conduction, the inductance stores energy and the load charges progressively. The associated state equation is given in (2).

$$V_{in} = L\frac{di_L}{dt} + V_C \,, \frac{V_C}{R} - C\frac{dv_C}{dt} - i_L = 0 \text{ avec } V_C = V_{out} \tag{2}$$

This leads to the state representation during the conduction phase as in (3).

$$\begin{cases} \dot{X} = \begin{bmatrix} \frac{di_L}{dt} \\ \frac{dv_C}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \\ Y = \begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} \end{cases} \tag{3}$$

When the transistor is in off state, the stored energy is transferred to the load via the diode. The average state representation is expressed as in (4).

$$\begin{cases} \dot{X} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} \frac{D}{L} \\ 0 \end{bmatrix} V_{in} \\ Y = \begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} \end{cases} \tag{4}$$

The transfer function for buck mode is given in (5).

$$G(s) = \frac{D}{LCs^2 + \frac{L}{R}s + 1} \tag{5}$$

➤ *Boost Mode*

For boost configuration, during transistor conduction (ON), the inductance stores energy. The state representation during conduction is given in (6).

$$\begin{cases} \dot{X} = \begin{bmatrix} \frac{di_L}{dt} \\ \frac{dv_C}{dt} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{RC} \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \\ Y = \begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} \end{cases} \tag{6}$$

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

The average state representation for boost mode is expressed in (7).

$$\begin{cases} \dot{X} = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} V_{in} \\ Y = \begin{bmatrix} 0 & 1 \end{bmatrix} \times \begin{bmatrix} i_L \\ V_c \end{bmatrix} \end{cases} \tag{7}$$

The transfer function for boost mode is given in (8).

$$G(s) = \frac{(1-D)}{LCs^2 + \frac{L}{R}s + (1-D)^2} \tag{8}$$

B. Mathematical Model for Buck Boost Converter Control

In the context of power converters, reinforcement learning can be formalized by a Markov Decision Process (MDP) defined by (9).

$$\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma) \tag{9}$$

Where:

S is the state space where S_t = [Vout(t), I_L(t), load(t), T(t)] , A the action space where A_t = D(t) with D(t) the PWM duty cycle, $P(S_{\{t+1\}}|S_t, A_t)$ is the dynamics from converter differential equations, $R(S_t, A_t)$ the reward function and $\gamma \in [0,1]$ the discount factor

C. Implementation of the Algorithm PPO

Proximal Policy Optimization belongs to the family of policy gradient methods in reinforcement learning. Unlike value-based methods, PPO directly optimizes the policy function that maps states to actions. The objective of PPO is to learn a control policy that maximizes the performance of the buck-boost converter under dynamic conditions. PPO interacts with the environment, collects trajectories of states, actions, and rewards, and updates the action policy prudently to avoid overly abrupt changes, hence the term "proximal." The PPO architecture consists of:

- Policy Network: a function, generally a neural network, that chooses the action to take according to the current system state
- Critic Network: a neural network that estimates the value of a state or advantage of an action
- Actor-Critic Architecture: combines the advantages of learning a policy by the actor and estimating a value function by the critic.

Fig. 2 presents the structure of our PPO agent within the environment.

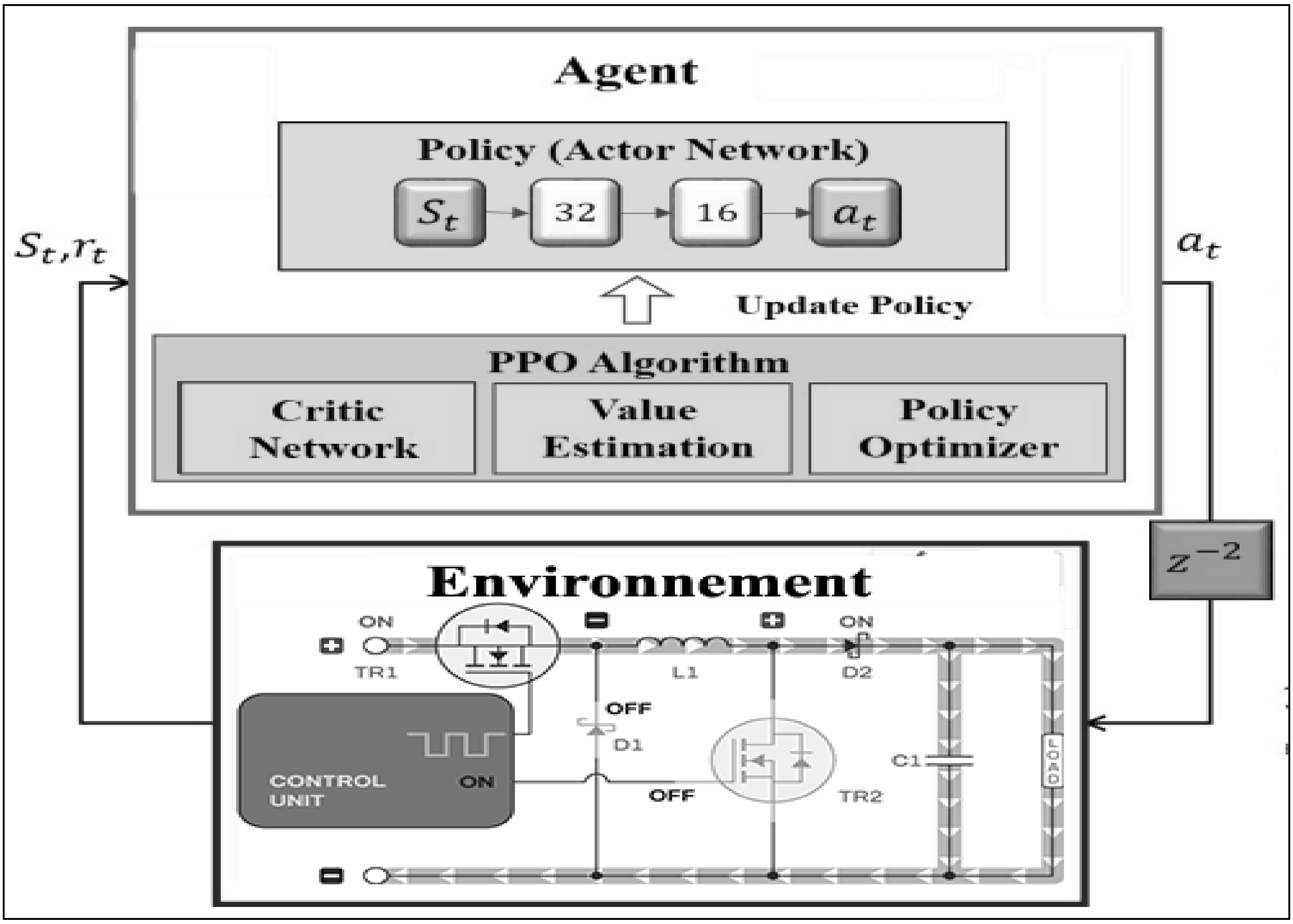

Fig 2 Structure of the PPO Agent within the Environment

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

➢ *Environment Definition*

In our case, the environment is the dynamic model of the buck-boost converter, considering:

- The dynamics of inductance current I_L
- The dynamics of output voltage V_{out}
- Possible variations of load R_{Load} and temperature T_j at the MOSFET level.

The system parameters are assigned in the Table 1.

Table 1 Parameters of the Non Inverting Buck Boost Converter

Parameters	Value
L (inductance)	220 µH
C (output capacitance)	1000 µF
Vsource_value range (voltage input value range)	[5 V, 24 V]
V_{ref} (voltage reference)	12 V

Table 2 Hyperparameters of the PPO Agent

Hyperparameters	Value	Description
learning_rate	$1e^{-4}$	Low learning rate for stable updates
n_steps	1024	Number of steps for good compromise between speed and accuracy
batch_size	64	Standard size for efficient learning .
gamma	0.98	Gives more importance to future rewards .
gae_lambda	0.92	Reduces GAE variance .
clip_range	0.2	Stabilizes policy updates.
ent_coef	0.005	Encourages exploration over exploitation

➢ *PPO Agent Configuration*

The PPO hyperparameters, listed in Table 2, are configured for optimal performance.

➢ *Signal Processing Interface*

A signal processing interface serves as communication between the buck-boost environment and the PPO agent. This interface collects data from the buck-boost converter, processes these data to resolve numerical problems, determines system states, evaluates taken actions and finally guides the agent to take optimal decisions.

Fig. 3 presents the signal processing subsystem that interfaced the buck-boost converter model with the PPO-based reinforcement learning agent.

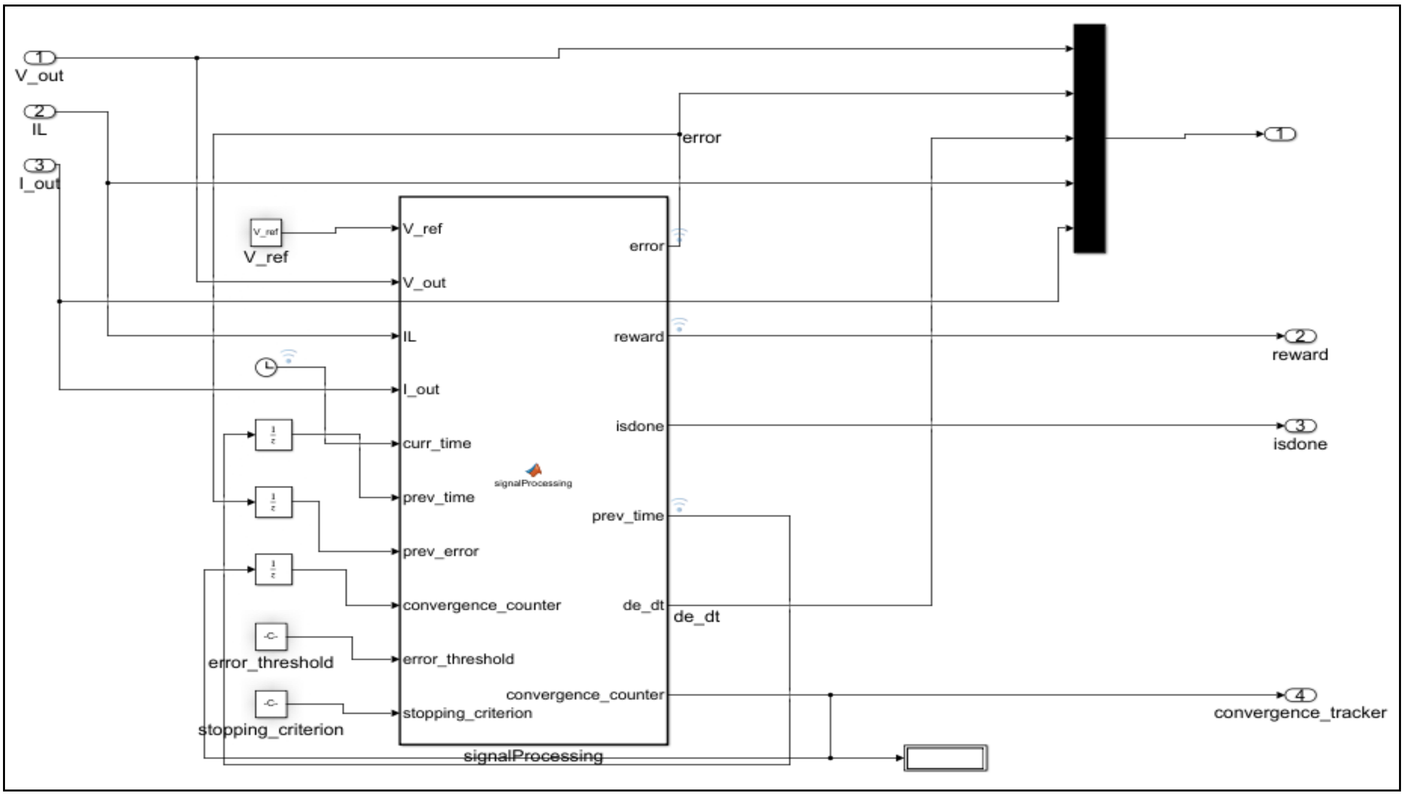

Fig 3 Signal Processing Interface between the Converter and the PPO

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

III. RESULTS AND DISCUSSION

A. Simulation Environment Setup

The PPO algorithm was implemented using Python with the following key libraries: *Gymnasium* for creating the custom buck-boost converter environment, *Stable-Baselines3* for PPO algorithm implementation and *TensorFlow* for neural network backend support.

The simulation environment accurately models the converter dynamics, including: the switching behavior and losses, parasitic resistances and capacitances, the temperature effects on component parameters, load variations and disturbances.

B. Training Phase and Evaluation of the Policy

At this stage, it is essential to consider the time required for the PPO agent to learn how to interact with the buck-boost converter environment. To ensure good convergence and stability of learning, the number of episodes has been set at 10,000. This allows the agent to explore the state space sufficiently and gradually improve its policy. Fig. 4 shows the curves of changes in total awards, actor losses, entropy, and critic losses in the 2000-episode interval.

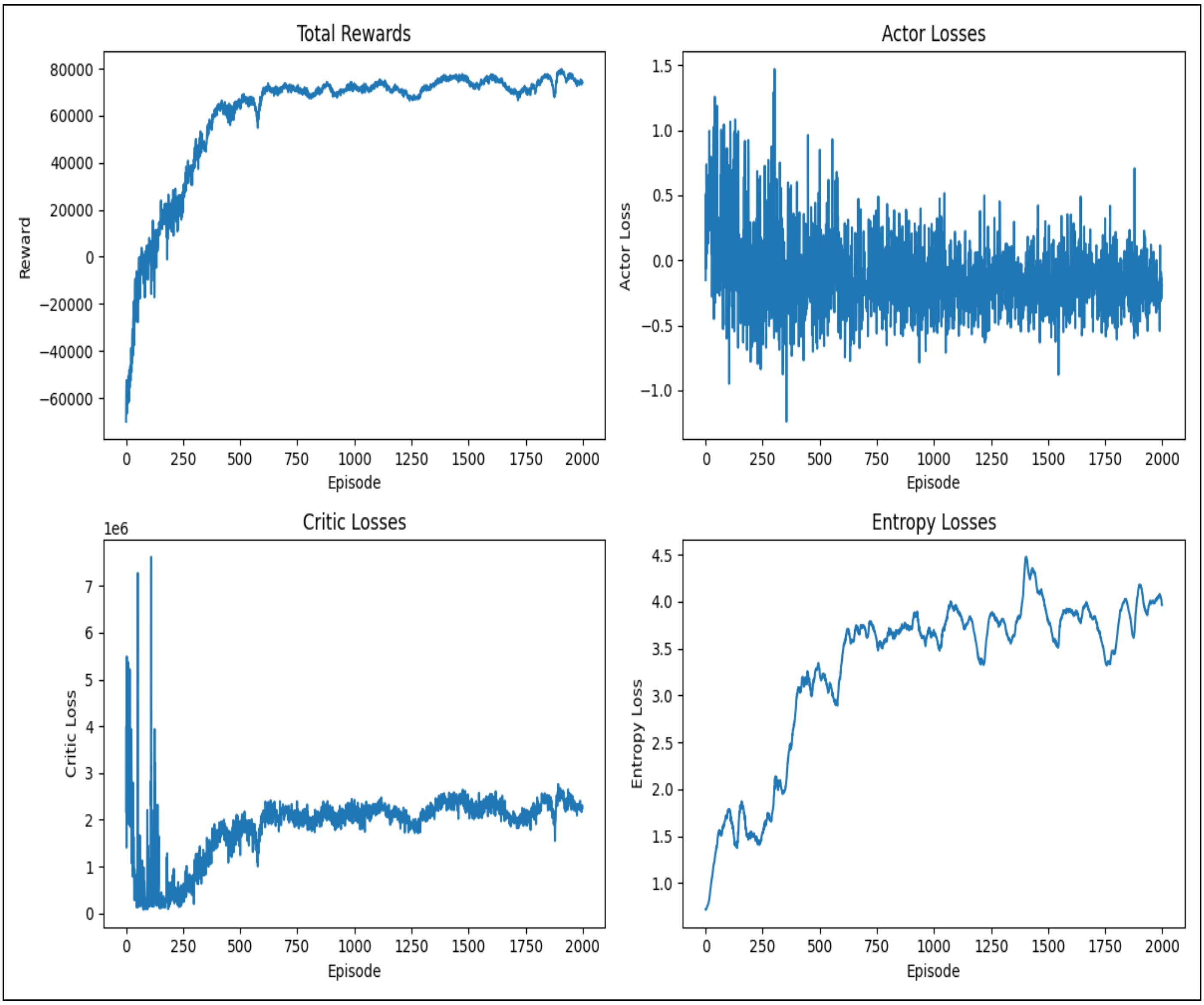

Fig 4 Evolutions of Total Rewards, Actor Losses, Critic Losses and Entropy Losses

Fig.4 shows that the total rewards increase rapidly and then stabilize, indicating a continuous improvement in the agent's performance. The actor's losses are decreasing overall with moderate fluctuations, reflecting a gradual adjustment of policy. Critical losses are high and then stabilize, showing a better estimate of value. The loss of entropy remains moderate, guaranteeing a good balance between exploration and exploitation.

C. Tests Phases

➢ *Case 1: Input Voltage Variation*

Fig.5 shows the waveform of the output voltage when the input voltage was varied while maintaining a constant resistive load of 100 Ω. The output voltage remained stable around the reference value of 12 V, with minimal overshoot and quick settling time.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

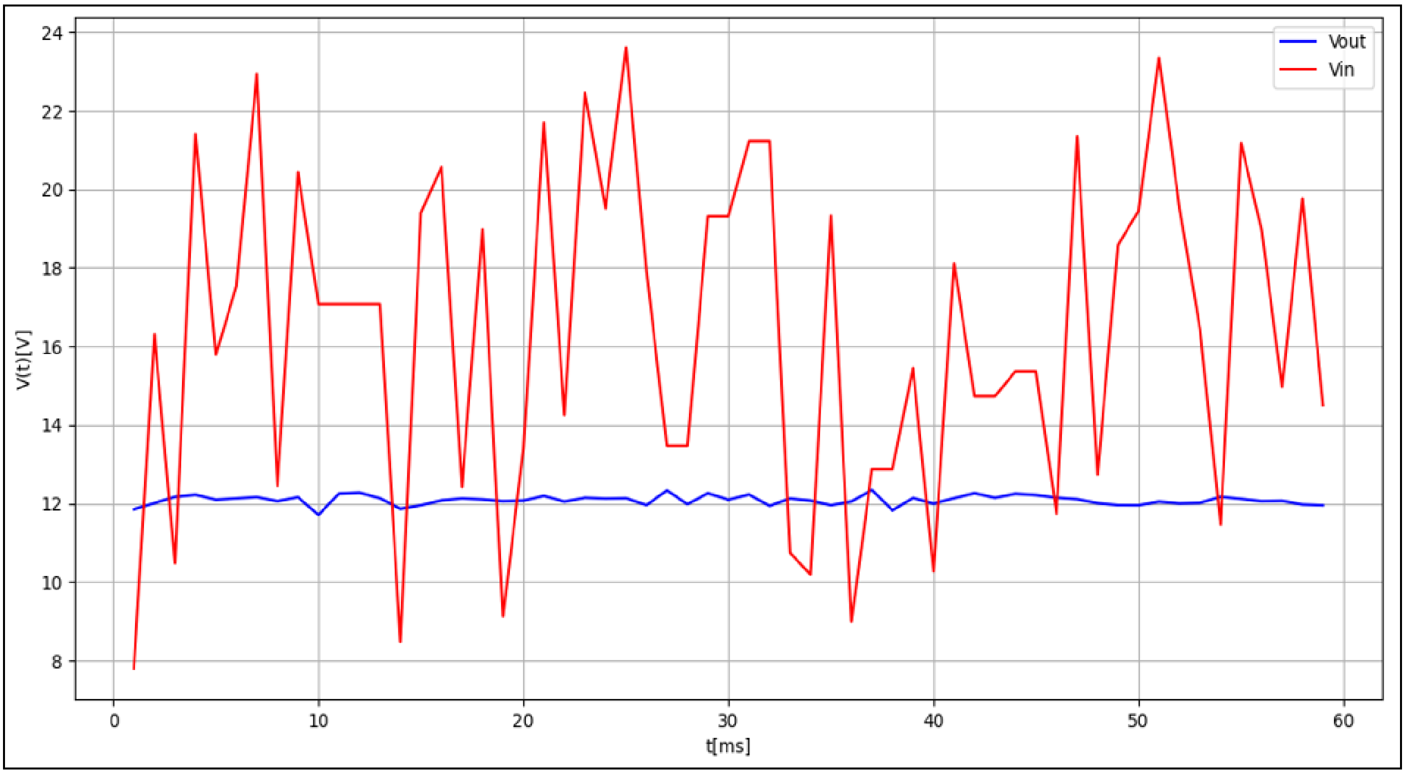

Fig 5 Wave form of the Output Voltage for a Variation of Input Voltage

Fig.6 shows a brief voltage drop of Vout observed around 23 ms, which was attributed to load adaptation. Despite this, our PPO agent, by varying the duty cycle, tries to stabilize the output voltage with respect to the reference voltage, $V_{ref.}$=12 V in a minimum time.

➢ *Case 2: Extreme Voltage Conditions*

Here, we set V_{in} to 5 V (minimum voltage) and then 24 V (maximum voltage) in order to observe the ability of the controller to lower and increase the output voltage. Fig. 7 and Fig.8 show the output and input voltage for the value of V_{in}=5 V and 24 V, respectively.

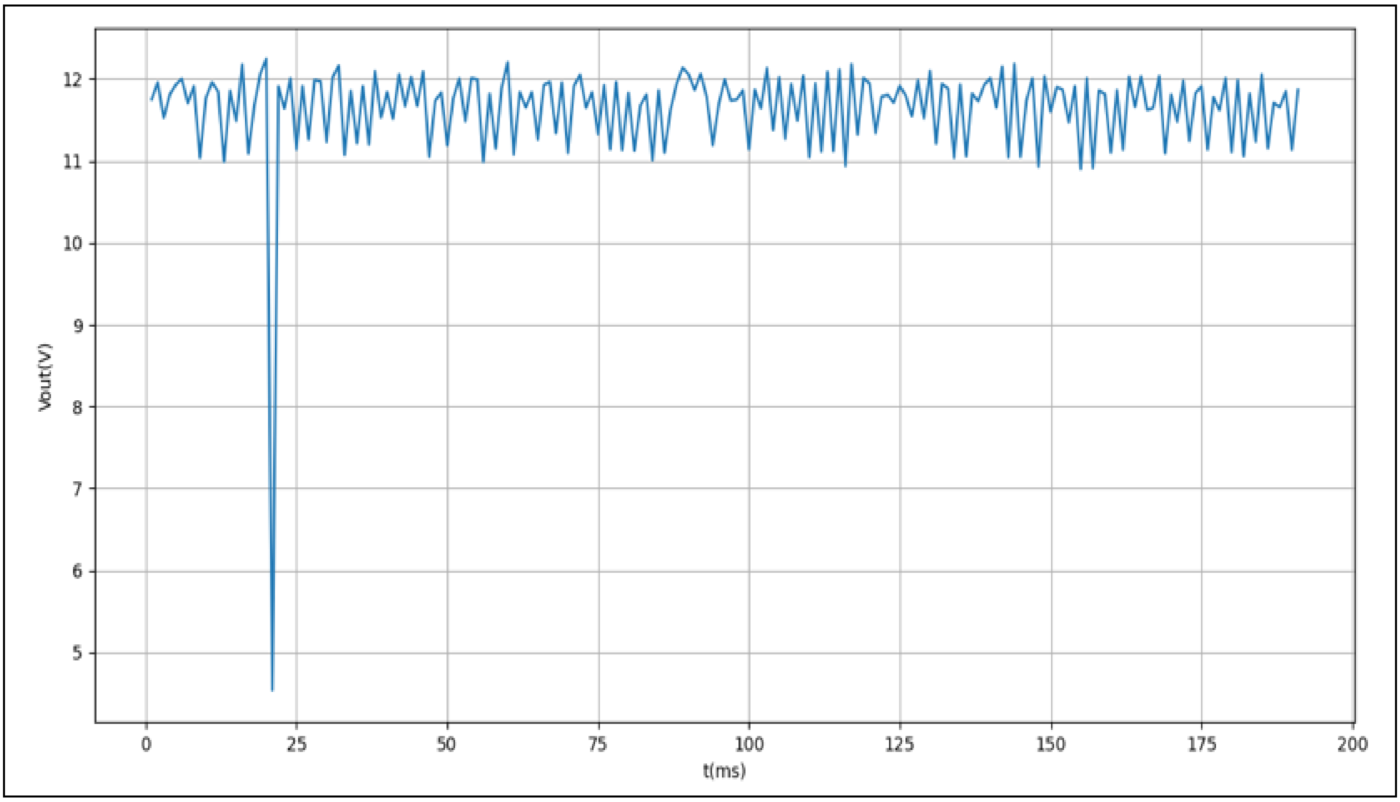

Fig 6 Wave form of the Output Voltage after a Change in Input Current

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

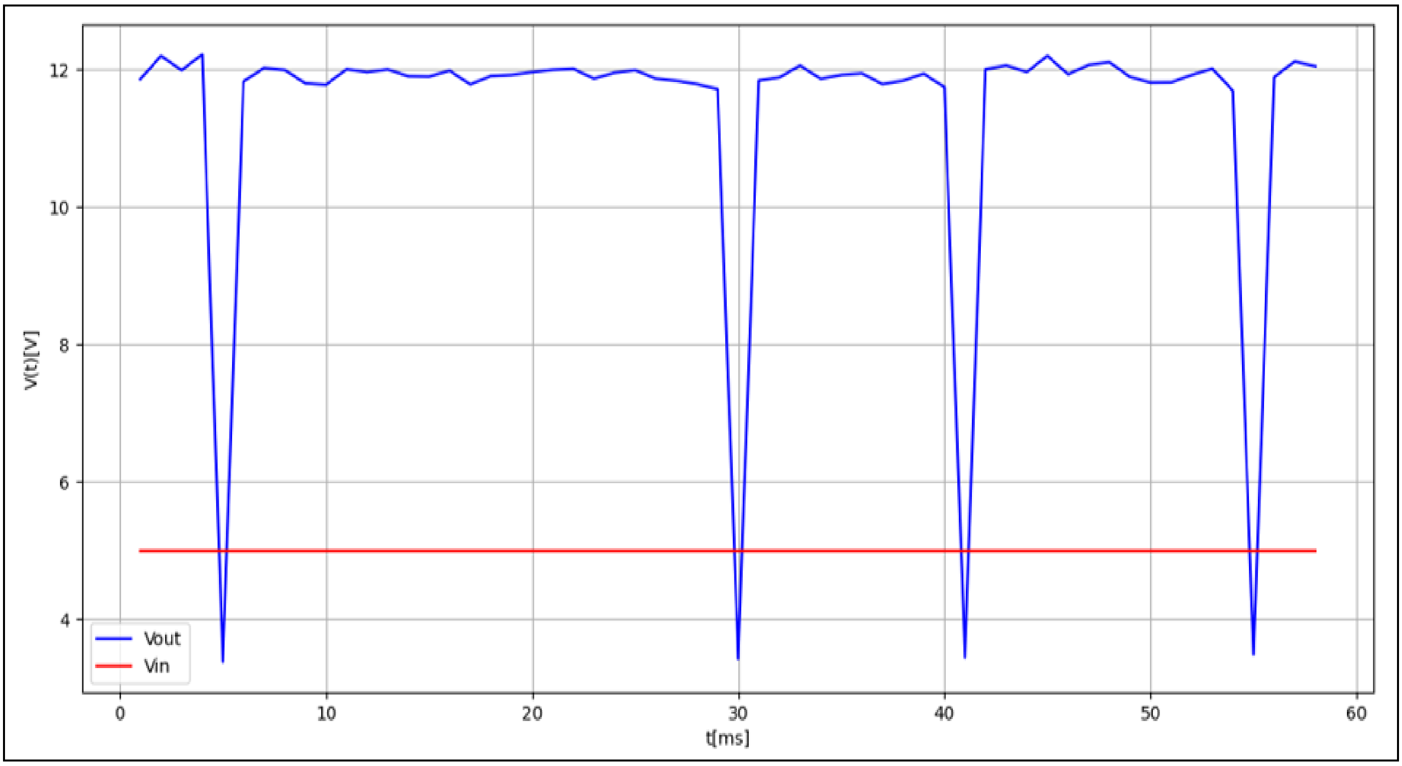

Fig 7 Wave form of the Output Voltage for V_{in}=5 V

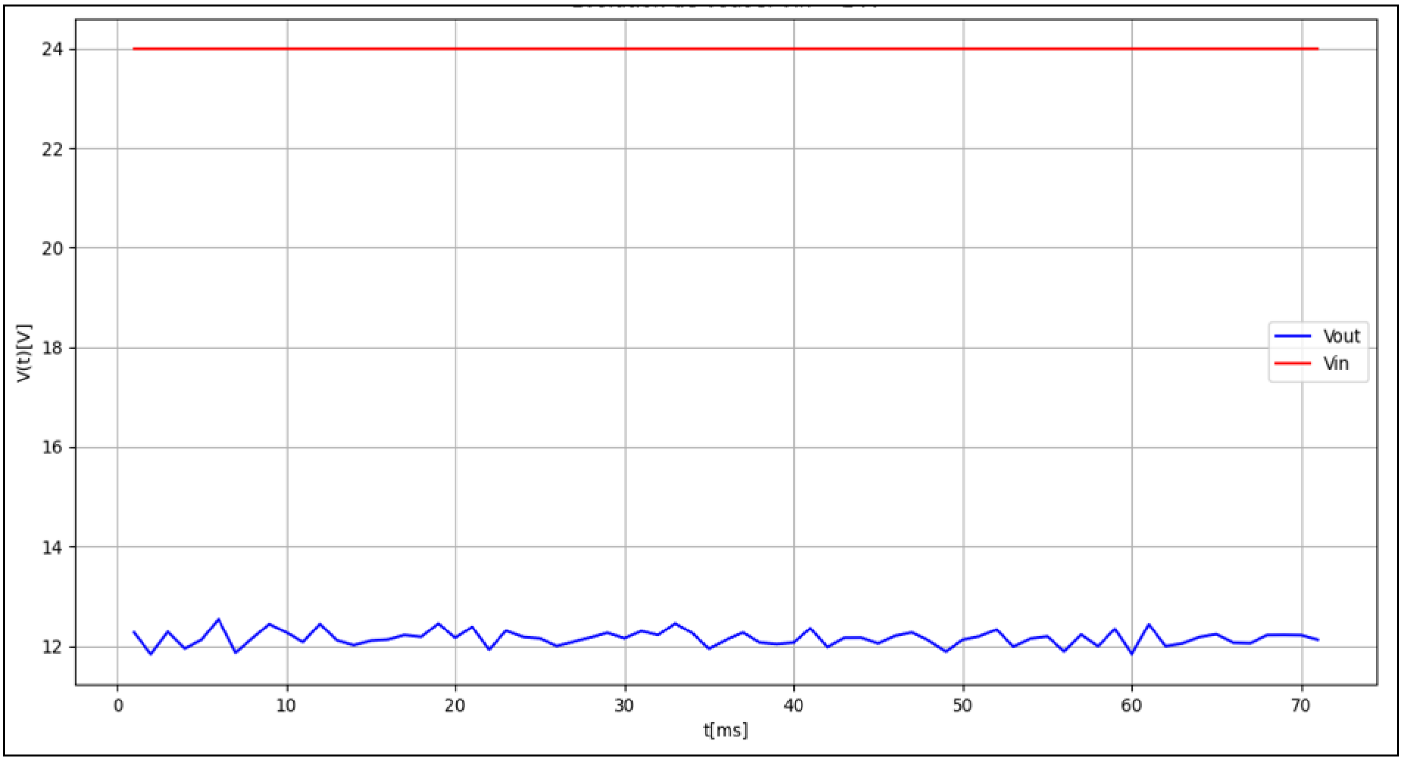

Fig 8 Wave form of the Output Voltage for V_{in}=12 V

Testing with minimum input voltage (5 V) and maximum input voltage (24 V) revealed the controller's ability to operate across the full range. For a minimum input (5 V), we observed faster stabilization with brief initial disturbance due to limited training in extreme boost mode. And for a maximum input (24 V), we observed slightly longer stabilization time but excellent steady-state performance

➢ *Case 3: Temperature Variation*

We vary the temperature between -55 and +175 °C, and we set V_{in} to 14.5 V (average input voltage) and we vary the value of the load. Fig.9 shows how the agent acts in the face of a temperature change. Here, we use a scale for temperature: 2.5 represents 20 °C.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

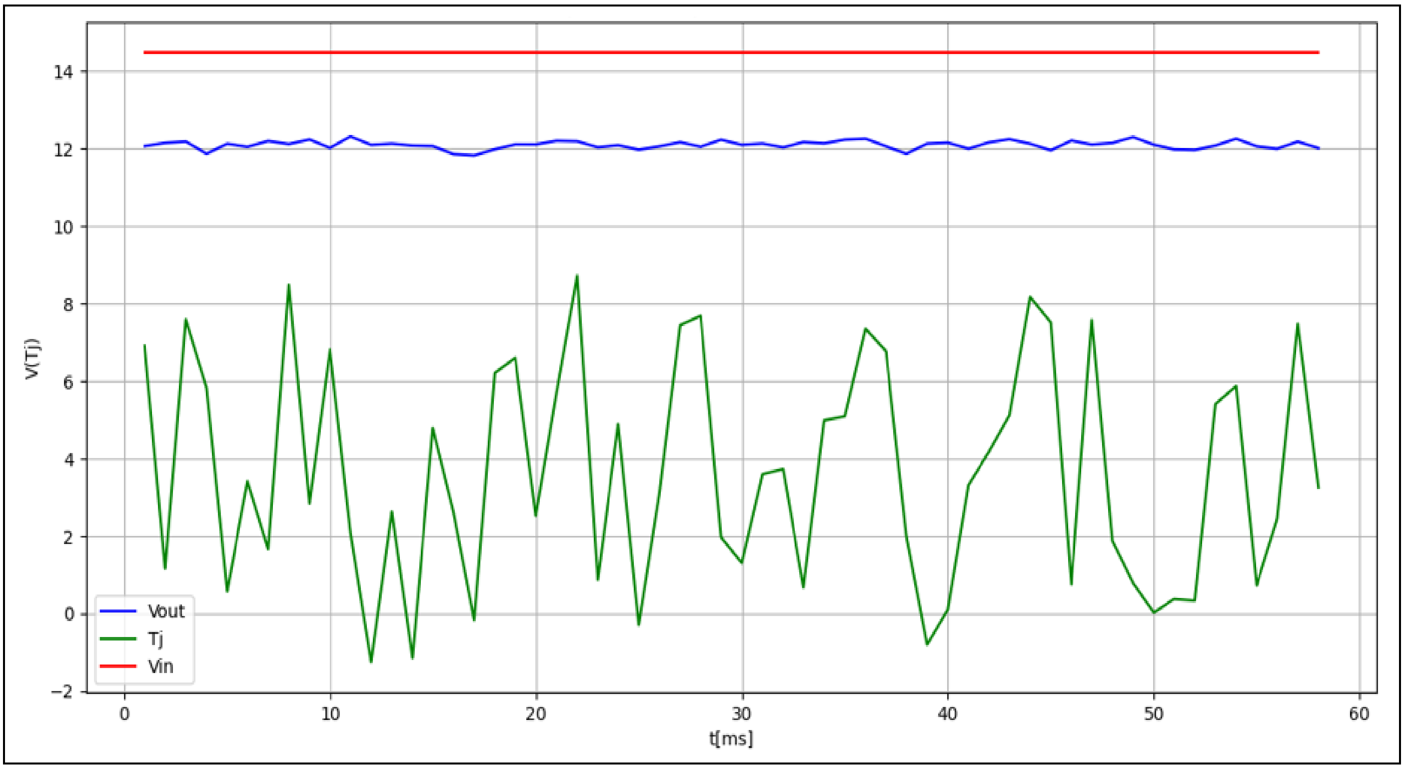

Fig 9 Wave form of the Output Voltage for a Variation of the Temperature T_j and the Load

Temperature variations from -55 °C to +175 °C (scaled representation) showed minimal impact on system performance. The controller maintained stable output regulation while adapting the switching frequency to manage MOSFET temperature and prevent thermal damage.

➢ *Case 4: Dynamic load variations*

Fig.10 shows the waveform of the output voltage and current as per the variation in the current of the inductor that reflects the current and voltage control of the smart controller.

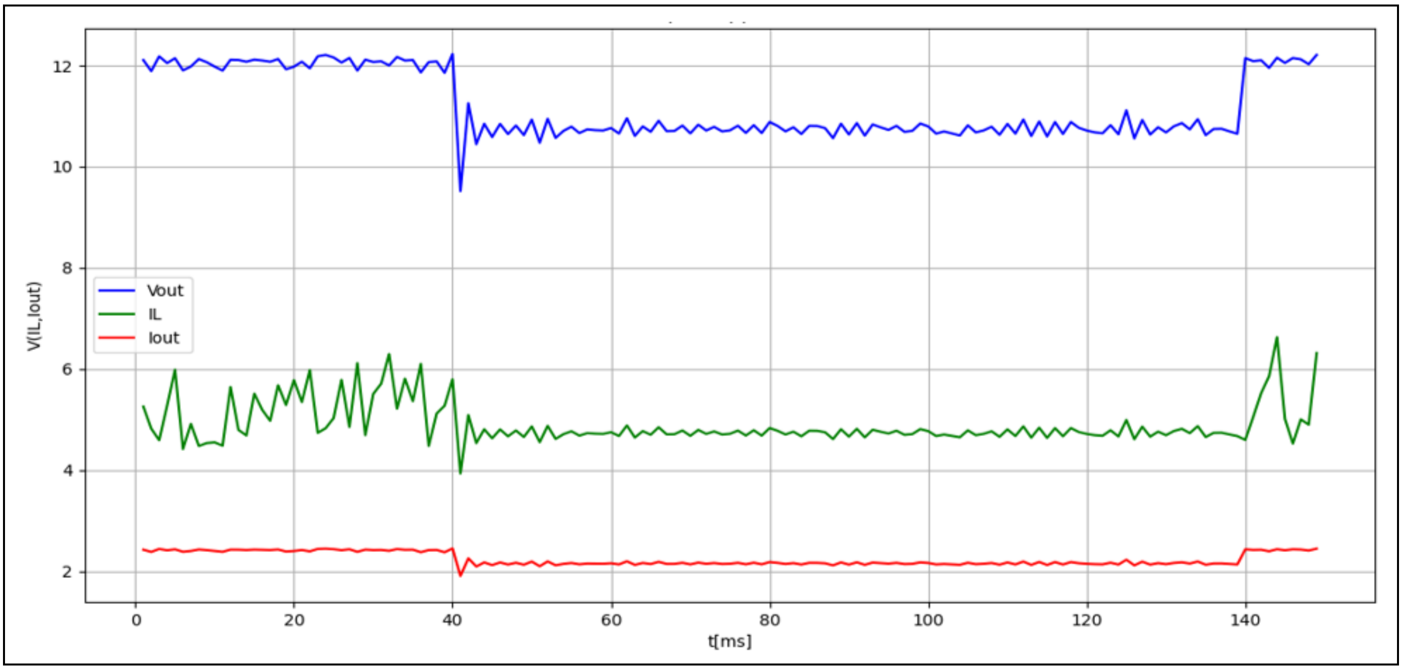

Fig 10 Wave form of the Output Voltage and Current for a Variation of I_L

Under varying load conditions, the controller demonstrated excellent current and voltage regulation. The system maintained output voltage within functional limits while adapting inductor current and output current to ensure overall stability.

D. Experimental Validation

The practical integration of the PPO model on the Raspberry Pi aims to transfer the decision capabilities learned in simulation to a real embedded system, ensuring direct interaction with the buck-boost converter. In this architecture,

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

the Raspberry Pi plays a central role: it hosts the PPO algorithm, reads the voltage and current sensors, and directly generates the control signals.

After training the PPO model in a simulated environment (via Gymnasium and Stable-Baselines3), the model is exported and then transferred to the Raspberry Pi in .zip format. The steps are:

- Loading the model with model = PPO.load ("ppo_model.zip").
- Reset the same runtime environment as the drive (voltage range, current, etc.).
- Adaptation of observation/action functions for the physical environment.
- Interface with the physical converter.

The Raspberry Pi interacts directly with the converter through:

- Sensors connected in GPIO to measure the output voltage (V_{out}), the output current (I_{out}) and possibly the temperature.
- PWM pins generated by the Raspberry Pi itself (via pigpio or RPi.GPIO) to adjust the control duty cycle.
- A control update frequency set between 100 Hz and 500 Hz, depending on the dynamic requirements of the converter.

The experimental validation using dual DC motors as load demonstrated the practical viability of the approach:

- Output Voltage Stability: 11.9 V (±0.1 V) under varying load conditions
- Current Range: 4.79 A to 5.07 A with smooth regulation
- Power Delivery: 57.3 W to 59.4 W with minimal oscillation
- Response Time: Sub-10ms control loop execution.

E. Discussion

The results demonstrate that PPO-based control provides a viable and superior alternative to traditional control methods for non inverting buck-boost converters . Results shows :

- Convergence: The PPO algorithm successfully learned optimal control policies within 10,000 episodes
- Stability: Excellent voltage regulation under various operating conditions.
- Efficiency: Significant improvements in energy efficiency and power quality.
- Practicality: Successful implementation on low-cost embedded hardware.

The approach addresses the fundamental limitations of traditional controllers while providing practical benefits for renewable energy systems. The ability to maintain stable operation under varying environmental conditions makes it particularly suitable for photovoltaic applications where input conditions change frequently.

IV. CONCLUSION

This paper has presented a comprehensive investigation of PPO-based control for non-inverting buck boost converters in renewable energy applications. The research demonstrates that deep reinforcement learning techniques can effectively address the limitations of traditional control methods while providing superior performance in terms of efficiency, stability, and adaptability. Key achievements of this work include: a comprehensive modeling of buck-boost converter dynamics suitable for reinforcement learning, successful adaptation of PPO algorithm for power electronics control, extensive testing under various operating conditions demonstrating superior performance, and the practical deployment on Raspberry Pi platform with real-time performance. Experimental confirmation of 97 % efficiency and enhanced power quality This research opens new possibilities for intelligent control in power electronics, particularly for renewable energy applications where adaptability and robustness are crucial. The successful implementation of PPO-based control on low-cost embedded hardware demonstrates the practical viability of AI-driven control strategies in power electronics, paving the way for more intelligent and efficient renewable energy systems.

REFERENCES

[1]. Mahafzah, K. A., et al. "Enhanced photovoltaic panel diagnostics through AI integration with experimental DC to DC Buck Boost converter implementation." Scientific Reports, vol. 14, 2024.
[2]. Recent developments of multiport DC/DC converter topologies, control strategies, and applications: A comparative review and analysis. ScienceDirect, 2023.
[3]. Sun, Y., et al. "Optimizing intelligent startup strategy of power system using PPO algorithm." Intelligent Decision Technologies, 2024.
[4]. Deep learning based buck-boost converter for PV modules. Heliyon, vol. 10, no. 5, 2024.
[5]. Proximal policy optimization-based reinforcement learning approach for DC-DC boost converter control: A comparative evaluation against traditional control techniques. PMC, 2024.
[6]. A Deep Reinforcement Learning Approach to DC-DC Power Electronic Converter Control with Practical Considerations. Energies, vol. 17, no. 14, 2024.
[7]. Reinforcement Learning-Based Control of a Power Electronic Converter. Mathematics, vol. 12, no. 5, 2024.
[8]. Learning-Based Model Predictive Control of DC-DC Buck Converters in DC Microgrids: A Multi-Agent Deep Reinforcement Learning Approach. Energies, vol. 15, no. 15, 2022.
[9]. Model predictive control of DC/DC boost converter with reinforcement learning. PMC, 2022.
[10]. A Novel Nonlinear Deep Reinforcement Learning Controller for DC–DC Power Buck Converters. IEEE Transactions, 2020.
[11]. Deep learning based buck-boost converter for PV modules. PMC, 2024.

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/25jul480

[12]. Enhanced photovoltaic panel diagnostics through AI integration with experimental DC to DC Buck Boost converter implementation. Scientific Reports, 2024.
[13]. A review of machine learning applications in power system protection and emergency control: opportunities, challenges, and future directions. Frontiers in Smart Grids, 2024.
[14]. Overview of Control Algorithm Verification Methods in Power Electronics Systems. Energies, vol. 14, no. 14, 2021.
[15]. A Deep Reinforcement Learning Approach to DC-DC Power Electronic Converter Control with Practical Considerations. Energies, vol. 17, no. 14, 2024.
[16]. A review of machine learning applications in power system protection and emergency control: opportunities, challenges, and future directions. Frontiers in Smart Grids, 2024.
[17]. Machine learning for modern power distribution systems: Progress and perspectives. Journal of Renewable and Sustainable Energy, vol. 15, no. 3, 2023.
[18]. Proximal Policy Optimization — Spinning Up documentation. OpenAI, 2024.
[19]. Proximal policy optimization - Wikipedia. Accessed 2024.
[20]. Machine learning for modern power distribution systems: Progress and perspectives. Journal of Renewable and Sustainable Energy, 2023.