Eksplorium *p*-ISSN 0854-1418
Volume 46 No. 1, May 2025: 1102–1129 *e*-ISSN 2503-426X

Single-Phase Grid-Connected Photovoltaic Systems using aDeep Reinforcement based MPPT Algorithm with Grey-Wolf Optimization under Partial Shading Condition

[1]Kandukuri Pradeep, Dr. K. Vinay Kumar[2]*

[1]Research Scholar, EEE department, Chaitanya Deemed university, Hyderabad

Pradeepkv56@gmail.com

[2]Associate Professor, EEE department, Chaitanya Deemed university, Hyderabad

kvinaykr1@gmail.com

Article Received: 24 Feb 2025, Revised: 27 April 2025, Accepted: 05 May 2025

Abstract: A detailed MATLAB model of a residential PV system that is linked to the grid is shown in this article. A state-of-the-art ILPAO approach was used to fine-tune the model. It has a unique conversion topology with two steps. An optimized current controller, a scanning-enabled Deep Reinforcement Learning based Maximum Power Point Tracking algorithm (SEDRL-MPPT), and a high-gain voltage regulator are all part of the suggested control scheme that attempts to solve the issue of multiple local maxima in the PV power-voltage characteristics caused by partial shading.The simulations showed that the stabilized current controller improvements significantly improved the responsiveness and stability of the DC connection's transient voltage, and oscillations and overshoot are reduced by the boost voltage regulator gains while operating under dynamic irradiance circumstances. If the irradiation is not uniform, SEDRL-MPPT will maximize energy extraction. Comparative analysis against a baseline control configuration confirms that the proposed method's effectiveness in improving system stability, dynamic performance, and power quality under realistic environmental disturbances. These findings validate the proposed approach as a robust solution for efficient and reliable residential PV-grid integration.

Keywords: Grid-connected photovoltaic system, two-stage converter, improved lightning attachment procedure optimization, scanning-enabled Deep Reinforcement Learning based Maximum Power Point Tracking

1 INTRODUCTION

Rising energy use shows no signs of slowing down [1]. The quick development of renewable energy sources is necessary if we are to reduce our dependence on fossil fuels and our negative effects on the environment. The energy business is dominated by solar electricity since it is the world's second-most-used renewable energy source [2]. Concerns about greenhouse gas emissions have increased, while the price of photovoltaic (PV) systems has been steadily declining, leading to their widespread adoption, particularly in sunny locations.

The production processes of photovoltaic (PV) modules, converter power electronics, and power point tracking (MPPT) To enhance the system throughput, optimisation of the controllers is necessary. When a DC/DC converter or inverter is used in conjunction with the MPPT algorithm, the MPP can reliably achieve the target, regardless of weather or solar radiation fluctuations. These review studies have given a thorough overview of all MPPT techniques used throughout the years [2, 3, 4]. They show how several groups may be created using attributes like memory, reaction speed, robustness, and sensor needs. Because of their accessibility and ease of use, traditional MPPT methods have enjoyed widespread adoption [5]. Aside from P&O, Incremental Conductance (IC) is the dominant method in this area.

Furthermore, Karami [6] included several tried-and-true methods, including the more conventional ones, such as OCC, OV, SC, and CC. When the sun's beams are constant, conventional approaches usually work, says Mohapatra [7]. Their energy conversion capability is diminished while working in partial shadow (PSC) since they are functioning at a local maximum power point (MPP). One big negative is this. With a large duty cycle step size, it can oscillate about the MPP; with a small one, the monitoring period is extended. In order to address the P&O approach's issues with sluggish tracking speed, poor convergence, and excessive oscillation, Ahmed [8] suggested allowing more freedom with scale of steps. With the MPP still some distance away, the controller has a lot of freedom to select a significant increment. Reducing oscillation at the MPP is why the step size is intentionally kept short. For further detail on these altered methods, see the Works Cited [2,3,4,5]. Rezk provided MSPT control with a variety of soft computing methods, including neuro-fuzzy (ANFIS) [11,12], ANN [10], and FLC [9]. Among the various evolutionary algorithms, the genetic algorithm (GA) finds use in cuckoo search (CS), many bee colony algorithms (BCAs), bat-inspired optimisation (BATs), bio-inspired memetic salp swarm algorithms (BATs), and many more. Jiang works as [19] claims that these methods, which use evolutionary algorithms and soft computing techniques, might address global nonlinear problems. One more option is to stick to partial structural constraints (PSCs) and yet use the global maximum partial probability (MPP). However, they have two major flaws. Inadequate convergence randomness can increase computing time, making it difficult for those without extensive knowledge of PV systems or access to expensive CPUs to find viable solutions. Particle swarm optimisation (PSO) is one prominent way for managing MPPT now, according to Rezk et al. [4]. Combining it with other algorithms can lead to novel approaches to MPPT control problems, as seen in Suryavanshi's PSO with P&O [21] and Garg's PSO with GA [22].

There has been a rise in interest and application of reinforcement learning (RL) in recent years[23,24] because it can learn from past data involving interactions with the environment, as opposed to more conventional approaches that need sophisticated mathematical models of the control system. The optimal solution to the MPPT control issue may lie with RL, as it surpasses meta-heuristics in both domains (Kofinas et al., 25). Despite the paucity of prior efforts, Q-learning is now the recommended method. Using Q-learning and MPPT control, Wei built a wind energy system that can adjust its speed in [26]. For tidal energy conversion systems in particular, the authors have created a maximum power point tracking (MPPT) controller [27]. Research into solar energy conversion systems has also focused on maximum power point tracking (MPPT) management [25,28,29]. These systems have a small action and state space, which is a drawback. In contrast to the four states established by Hsu et al. [28] and Youssef [29], Kofinas et al. [25] established a state action space with four thousand state actions by linking eight hundred states with five actions. This indicates that systems with large state and action spaces have longer calculation durations. A new methodology was suggested by Pohan and Lai [30] that integrated P&O and Q-learning techniques. An ideal duty cycle may be found using a Q-learning controller by monitoring the temperature and radiation-split control zones. A smaller step size is achieved by adjusting these duty cycles using the P&O controller. Using a Q network and a Q table, respectively, Chou [31] has

developed two RL-based MPPT algorithms. However, PSCs are under-represented in those studies. Instead of using a single trained agent, techniques [32,33] use a multitude of agents to tackle the MPPT control issue. Researchers looked into the possibility of using transfer reinforcement learning algorithms to keep tabs on the world's biggest PowerPoint presentation [33]. We have developed a new approach for regulating PV systems in scenarios with partial shadowing using memetic reinforcement learning [32]. The main problem with these methods is that they have always only considered a limited, separate state and action space.

One novel approach to machine learning that has recently gained traction is deep reinforcement learning (DRL). Optimal control is a problem that it could solve [34,35,36]. Atari and Go are among the games used in the study to prove that the DRL approach is effective [37]. A control problem with a large state space can be efficiently solved using DRL. One of the several advantages of using DRL in contexts is that it is applicable when the state space and the action space are same. Several fields have found real-world applications for DRL, including robots [35,38], computer vision [38], healthcare [40], smart grid [41], game design [37], and natural language processing [39]. In regards to the electrical grid, Zhang [42] provided a brief overview of DRL. Similarly, deep reinforcement learning is used to approximation functions in wind energy conversion systems for maximum power point tracking (MPPT) [43,44], although this time a neural network is utilised rather than a Q-value table.
Maximum Power Point Tracking (MPPT) algorithms still aren't prepared for extreme partial shade conditions, no matter how far they come. The outcome is a power-voltage characteristic curve with several local maxima. Traditional MPPT methods and control systems are inadequate when attempting to ascertain the GMPP. The efficiency of energy extraction decreases because of this. Although the grid and the environment have high expectations for the performance of Deep Reinforcement Learning (DRL) algorithms, no one has yet investigated their use for maximum power point tracking (MPPT) regulation in residential grid-connected photovoltaic (PV) systems.

This research highlights a significant limitation in the current state of MPPT controllers that are able to effectively adjust to changing operating circumstances, shading, and non-uniform irradiance, all while maintaining system stability and excellent power quality. As a result, this study offers a novel MPPT control strategy that combines an ILAPO-optimized control framework with DRL-enhanced algorithms to overcome these limitations. In residential PV energy conversion systems, Achieving optimal power harvesting, enhanced dynamic responsiveness, and dependable functioning in the face of real-world environmental disturbances is the goal. With an emphasis on maximum power point tracking (MPPT) under partial shadowing circumstances (PSC), this eight-part research intends to provide a comprehensive evaluation of grid-connected residential PV systems. The primary In the first part, "Introduction," we learn about the significance of the study and about the latest innovations in PV system optimisation. An extensive introduction to photovoltaic systems is given in Section 2, "Modelling of PV Module under PSC," which also delves into the mathematical workings of modules and the intricate consequences of uneven illumination.In Section 3.1, the basics of Deep Reinforcement Learning (DRL) are covered, and in Section

3.2, the ILAPO approach is introduced. Provided in Section 3: Background is the essential theoretical context. Section 4: Methodology details the main methodological contributions and how they were put into practice. It takes a look at the following sections: (4.1) The Definition of States, Actions, and Rewards in SEDRL-MPPT, (4.2) The ILAPO-based Controller Parameter Optimisation Using ITAE Objective, (4.3) The Integration and Workflow of the proposed system, and (4.4) The Simulation Setup for Grid-Connected Residential PV System with Two-Stage Conversion. In Section 5: Results, we show quantitative and qualitative results that result from the system behaviours. Section 7: Conclusion provides a concise summary of important results and future study directions, while Section 6: Discussion presents a critical analytical discourse on these empirical findings. In Section 8: References, all relevant intellectual contributions are cited with great care.

2 MODELLING OF PV MODULE UNDER PSC

2.1 Mathematical Model of PV Module

By forming a p-n junction in a very thin layer of semiconductor materials, photovoltaic (PV) solar cells are able to transform the energy from the sun into electricity [30]. For PV system modelling, a trustworthy solar cell model is crucial [6]. A single-diode model is simpler and more practical, regardless of its accuracy. Based on a single diode, the electrical circuit used in the research is reminiscent to a solar cell [28]. For example, a perfect cell's output current can be [3,28,45]:

$$K = K_{ph} - K_d - K_{sh} \tag{1}$$

Where K_{sh} is the current through the parallel resistance, which is determined by

$$K_{sh} = \frac{V + KR_s}{R_p} \tag{2}$$

Where R_p and should be set to its maximum value to represent the parallel resistance, along with R_s represents the series resistance that results from all the impediments that the current must traverse, and it must be maintained at a minimum. A flow of electricity generated by light, or K_{ph}, is also directly related to how bright the light is

$$K_{ph} = \left[K_{sc} + C_K \left(T_c - T_r \right) \right] \times \frac{Q}{Q_{STC}} \tag{3}$$

Where C_K thinks about the temperature coefficient of the cell's short-circuit current and K_{sc} this is the current across a short circuit when tested under STC (T = 25, = 1000 W/m2). Q is the relative irradiation, T_r is the temperature used as a benchmark, and T_c represents the temperature at which the cell is actively working. A current via a diode, or K_d, is provided by

$$K_d = K_0\left[\exp\left(\frac{eV_d}{AqT_c}\right)-1\right] \quad (4)$$

Where A is the diode's ideal factor, q = 1.38 × 10−23 is Boltzmann's constant, and e = 1.6 × 10−19 is the electronic charge. V_d is the voltage of the corresponding diode, and K_0 is the diode's reverse saturation current. They are determined by

$$V_d = V + KR_s \quad (5)$$

Typically, PV cells are connected in series to form PV modules. The current through a photovoltaic module may be easily calculated using a simple mathematical model that takes into account both temperature and sun irradiation:

$$K_{pv} = K_{ph} - K_O\left[\exp\left(\frac{e\left(V+KR_s\right)}{AqT_cN_s}\right)-1\right] \quad (6)$$

Where the number of series resistance cells is represented by N_s.

Environmental conditions have a significant impact on a PV module's properties, as the equation above illustrates. The ACS-335-M PV module from American Choice Solar is utilized in this research to simulate a photovoltaic system. Table 1 provides an illustration of its specifications.

Data Sheet for the ACS-335-M Solar Photovoltaic Module by American Choice Solar

Specifications	Value
Maximum Power (W)	334.905
Voltage at MPP (V)	41.5
Current at MPP (A)	8.07
Open circuit voltage, Voc (V)	49.9
Short circuit current, Isc (A)	9
Temperature coefficient of Voc (%/ ◦C)	−0.36
Temperature coefficient of Isc (%/ ◦C)	0.09

2.2 Partial Shading System Effect

Arrays of photovoltaic (PV) solar modules connected in series or parallel provide the necessary voltage and current. When two modules are linked in series under PSC, the PV curve will display two peaks. Solar panels are no differen; a series connection of five can only generate five peaks. The methods outlined here are applicable to many different types of PV systems. Putting three photovoltaic modules in sequence allows you to mimic the variation between a global and a local maximum power point, which can simplify things. The illustration shows how to prevent PV modules from overheating under partial shadow

circumstances (PSCs) by using a blocking diode and bypass diodes [2,3]. When many things cast shadows on multiple PV modules—be it buildings, poles, or even bird droppings—this phenomenon is known as partial shading. Rather of being an asset, it becomes a burden in this situation. As time passes, hot spot phenomena can cause shaded PV modules to suffer damage [14,46,47]. A parallel connection of a bypass diode prevents the PV system and its modules from overheating. The bypass diode has the ability to reverse bias if it is constantly exposed to direct sunlight. If a photovoltaic module uses a diode rather than the module itself to transmit sunlight, a forward bias will result. The power curve, on the other hand, will show several peaks with local and global maxima when a partial shadow and bypass diode are present. You may maximise the efficiency of your PV array and potentially reduce power loss by 70% by operating the system when it is operating at its peak efficiency [2].

2.3 PV System Introduction

Changes in temperature and solar energy may greatly affect PV solar cells due to their nonlinear properties. Based on the data, we can see that temperature has a negative correlation with PV output power, and that the two variables are directly proportional.

Results from previous research show that knowing the optimal terminal voltage of the array is crucial for PV panels to operate at their maximum power point (MPP) under specific weather conditions [47, 48]. A constant MPP can only be achieved with efficient MPPT control [7]. Additionally, PSCs cause a PV panel's P-V curve to have several peaks. An intelligent MPPT controller may be able to resolve the issues with conventional MPPT methods. The DC-DC converter is a key part of the MPPT method. By adjusting the pulse width modulation (PWM) signal's duty cycle D, a DC-DC converter may regulate the voltage output of a photovoltaic array. By analysing this signal, maximum power point tracking (MPPT) controllers determine the optimal voltage for energy generation. Visit to learn about the operating duration of a DC-DC boost converter [30].

$$D = 1 - \frac{V_{in}}{V_{out}} \tag{7}$$

3 BACKGROUND

3.1 Basic Concept of DRL

A more advanced kind of reinforcement learning (RL), DRL is briefly described here. One kind of unsupervised machine learning, reactive learning (RL) relies on the idea that agents are neutral when faced with outside stimuli [49]. The use of reinforcement learning to address issues with sequential decision-making is gaining popularity among computer science researchers [24,36,50]. By repeatedly trying different strategies in a particular situation, reinforcement learning (RL) can find the policies or behavioural approaches that maximise the total projected discounted rewards [51]. The standard components of an RL model are an agent, a setting, tasks, states, and incentives [23]. The agent then goes on to activate the RL algorithm, making use of its environment as a stand-in for the object it is interacting with. In response to changes in its environment, the agent applies its knowledge to the new state. Following that, it incorporates two future states along with environmental incentives. Once

the agent receives the reward, it can evaluate its most recent activity. An episode ends when the setting changes, and a new one begins when it is prepared. This loop will keep going until all of the conditions are met [23].

To determine the optimal action, some approaches apply the value function $Y\rho(y)$Vπ(s), which represents the probability of the agent reaching a certain state [51]. It is expected that this will happen if government policy is followed. How the action-value function works Equation $Q\sigma(y,a)$Qπ(s,a), which forms the basis for several additional processes, denotes the anticipated following the policy M, returning to the current state of doing this activity. The $Y\sigma(y)$Vπ(s) and $Q\sigma(,a)$Qπ(s,a) functions may be generated using the following formula [23,42,51]:

$$V^{\pi}\left(s_t\right)=E\left[R_t\left|s_t=s\right.\right]=E\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}\left|s_t=s\right.\right] \tag{8}$$

$$Q^{\pi}\left(s_t,a_t\right)=E\left[R_t\left|s_t=s,a_t=a\right.\right]=E\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}\left|s_t=s,a_t=a\right.\right] \tag{9}$$

An off-policy, model-free RL approach known as Q-Learning has been gaining traction across several sectors. It is possible to repeatedly show the $Q\rho(y,a)$ function using the Bellman equation in Q-Learning [23,51]:

$$Q^{\pi}\left(s_t,a_t\right)=E\left[r_{t+1}+\gamma Q^{\pi}\left(s_{t+1},a_{t+1}\right)\left|s_t,a_t\right.\right] \tag{10}$$

An optimal strategy $\pi*$ achieves the largest cumulative reward over an extended period of time. Currently, [23] provides the finest value function and action-value function.

$$\pi^{*}=\arg\max_{\pi}V^{\pi}\left(s\right) \tag{11}$$

$$V^{*}\left(s\right)=\max_{\pi}V^{\pi}\left(s\right) \tag{12}$$

$$Q^{*}\left(s,a\right)=\max_{\pi}Q^{\pi}\left(s,a\right) \tag{13}$$

Among the many current AI hotspots, deep reinforcement learning (DRL) stands out. Autonomous learning is made possible by engaging with a particular environment. A number of domains have profited substantially from DRL's use of RL and DL. Computer games, artificial intelligence, natural language processing, and company and financial management are all part of this category. Using look-up tables to store and index data is a big problem with reinforcement learning. For practical issues involving expansive state-and-action areas, this could prove troublesome. So, value functions or policy functions can be approximated by neural networks [37,51]. So, it is possible to assign Q values to states or state-action pairings using neural networks.

The two types of solution strategies shown in Figure 1 are model-based and model-free. Learnt or known models are used in model-based DRL. One major advantage of the model-based approach is how few instances it need for learning. However, the computational

complexity increases significantly when the model is unexpectedly challenging to train. Dealing with RL that does not rely on models, nevertheless, will yield better results. It works with minimal computing power and doesn't need an accurate description of the environment. A model-free decision-relationship mapping that is based on values or policies is available. The objective of value-based approaches is iteratively enhance the value function until convergence is achieved. Presented here are the objective function and updating mechanism [36,42]:

$$J(\theta)=E\left[\left(r_{t+1}+\gamma\max_{a}Q\left(s_{t+1},a_{t+1}|\theta\right)-Q\left(s_{t},a_{t}|\theta\right)\right)^{2}\right] \quad (14)$$

$$\theta_{t+1}=\theta_{t}+\alpha\left(\left(r_{t+1}+\gamma\max_{a}Q\left(s_{t+1},a_{t+1}|\theta\right)-Q\left(s_{t},a_{t}|\theta\right)\right)\right)\nabla_{\theta}Q\left(s_{t},a_{t}|\theta\right) \quad (15)$$

where α is learning rate, and θ is the weights of the neural network.

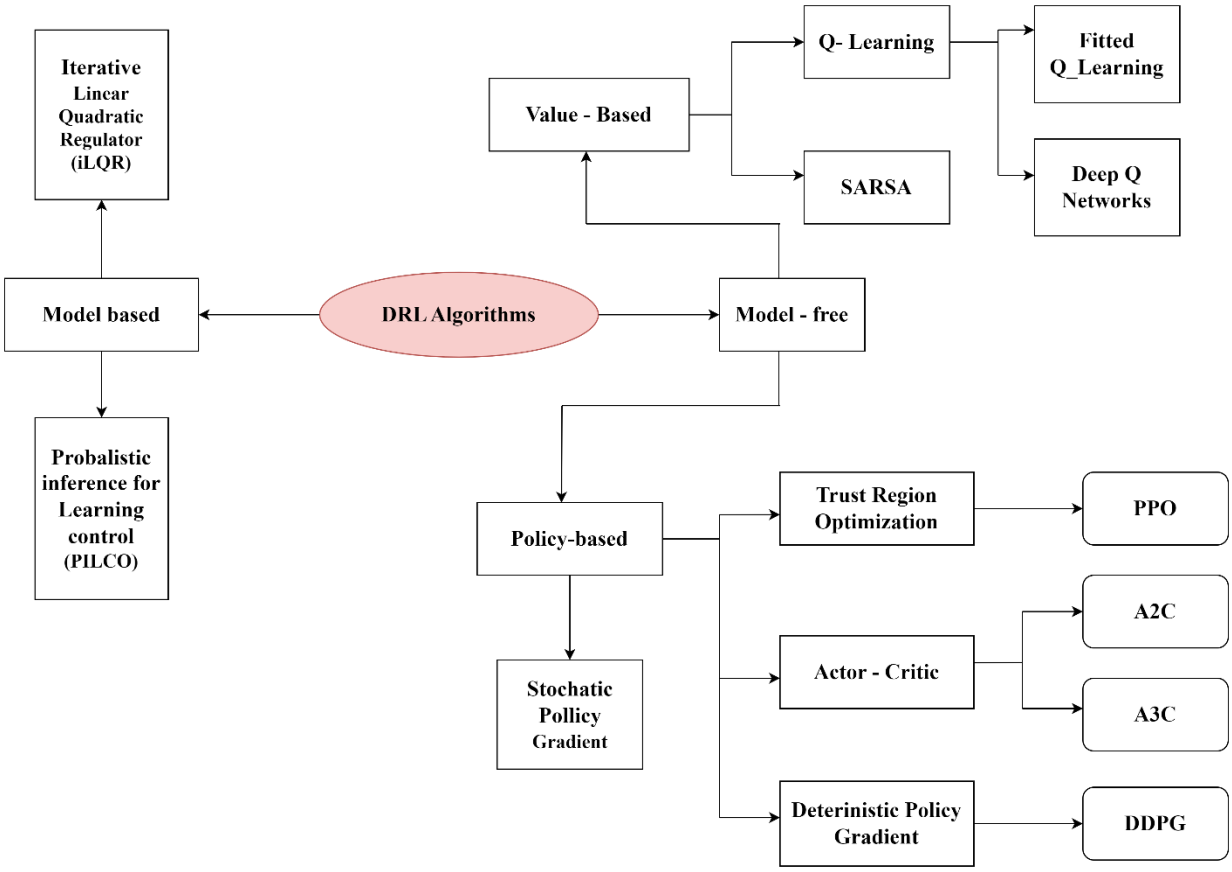

Figure 1. Introduction of deep reinforcement learning (DRL) algorithms.

Algorithms based on policies optimise the policy at each time step and use the current policy to determine the value maximise interest until policy convergence is reached. First things first, as you can see in the image below, we update the weight matrix and get the objective function's gradient [36,42]:

$$\nabla_{\theta}J(\theta)=E\left[\sum_{t=0}^{T}\nabla_{\theta}\log\pi_{\theta}\left(a_{t}|s_{t}\right)\sum_{t=0}^{T}r\left(s_{t},a_{t}\right)\right] \quad (16)$$

$$\theta \leftarrow \theta + \alpha \nabla_{\theta} J(\theta)$$

3.2 Improved Lightning Attachment Procedure Optimization (ILAPO)

As an acronym for "Improved Lightning Attachment Procedure Optimisation," ILAPO describes a particular algorithm. The occurrence of lightning and thunder (Nematollahi et al.)laid the groundwork for it. A new optimisation method called LAPO is based on the idea that clouds may store a great deal of electrical energy, just like lightning. Lightning is produced by an accumulation of electrical charges on the form cloud, which leads to a rise in electrical power and the occurrence of lightning strikes at various locations (Refai et al.). The phenomena of LAPO optimization is comprised of many phases.

Clouds are transferred to the earth in the first stage. The next step is to make a flash. Composing lightning and thunder is the final step. Steps in the LAPO optimisation process include starting a population, fading branches, moving leaders up and down, and improving execution(EL Sayed et al.). One possible outcome of the initialisation of the population is the occurrence of discharges; this is because each dot acts as a member of the population. The Downward Motion leader uses a random selection procedure to choose the members to serve as test dots. After the Leader of Downward Motion has created an ideal and substandard dot, the procedure known as the Leader of Upward Motion is applied. The branch's fading may be likened to coordinating lightning by looking at a dot made by a leader travelling upwards. Every production instance will undergo this process again(Kamel and Youssef), (Tan et al.). Figure 2 illustrates the flowchart that delineates the LAPO optimization approach.

To increase the searching capabilities and prevent the traditional LAPO from stagnating, the improved lightning attachment process optimisation (ILAPO) is suggested. LAPO is an effective method for solving a wide range of optimisation problems. But it's still similar to other meta-algorithms. In certain cases, it could experience local optima and stagnation. Improving the regular LAPO's exploration and exploitation capabilities is what the proposed ILAPO is all about (Ebeed et al.). By randomly moving the test locations using Levy Flight, the following changes are made to the exploration component of the first iteration process (Viswanathan et al.):

$$X_{s\_new}^{i} = X_{s}^{i} + \propto \oplus Levy(\beta) \tag{17}$$

where $\propto$ is a step size parameter that may be derived as follows:

$$\alpha \oplus Levy(\beta) \sim 0.01 \frac{u}{|v|^{1/\beta}} \left(X_{s}^{i} - X_{best}\right) \tag{18}$$

where u and v can be found from (3) and (4) as follow

$$u \sim N(0, \phi_{u}^{2}), \quad v \sim N(0, \phi_{v}^{2}) \tag{19}$$

$$\phi_{u} = \left[\frac{\Gamma(1+\beta) \times \sin(\pi \times \beta/2)}{\Gamma[(1+\beta)/2] \times \beta}\right]^{1/\beta}, \quad \phi_{v} = 1 \tag{20}$$

where Γrepresents the standard beta function. Points surrounding the best solution are determined using a logarithmic spiral function are updated in a spiral pass during the last iteration of the procedure to enhance the exploitation phase:

$$X_{s\_new}^{i} = \left|X_{best} - X_{s}^{i}\right| e^{bt} \cos(2\pi t) + X_{best} \quad (21)$$

where the logarithmic spiral form is defined by a constant b.

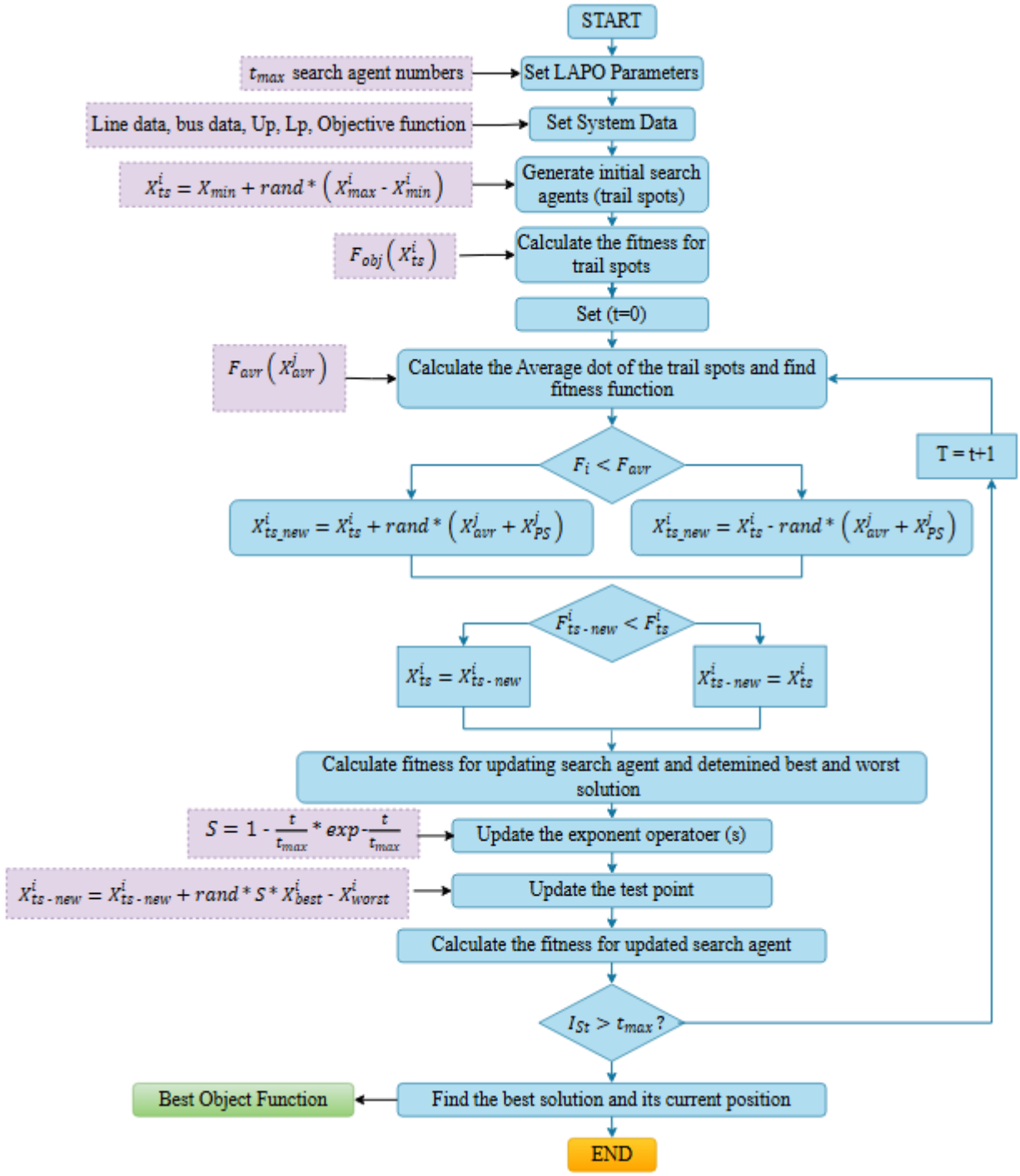

Figure 2 Flowchart of LAPO Optimization

4 METHODOLOGY

This study builds a residential PV system's adaptive MPPT controller using Deep Reinforcement Learning (DRL). Unlike static operating points or pre-defined mathematical models used in maximum power point tracking (MPPT), dynamic response learning (DRL)

allows the controller to develop an optimal control strategy through never-ending dialogue with nature. The DRL agent monitors the system, adjusts the control input, and receives feedback on incentives to improve its policy. Our data-driven method outperforms state-of-the-art MPPT algorithms in partially shadowed situations due to the several local maxima on the PV power-voltage curve.

4.1 Definition of States, Actions, and Rewards in SEDRL-MPPT

The suggested control strategy revolves around the SEDRL-MPPT algorithm, which combines a scanning mechanism with a DRL agent to provide reliable tracking of the peak power point on a global scale.

States (S): The state vector is constructed from real-time measurable electrical parameters that reflect the system's operating condition and environment. Specifically:

- PV array voltage V
- PV array currents I
- Voltage variation $\Delta V=V(t)-V(t-1)$
- Current variation $\Delta I=I(t)-I(t-1)$

These four parameters form a comprehensive state representation that allows the DRL agent to discern trends in power output and detect changes in irradiance or shading patterns.

Actions (A): The agent may adjust the duty cycle of the DC-DC converter with respect to the PV array. This is the area where they may take action. Steps consist of:

- A precise increment in the duty cycle
- Limiting the duty cycle to a predetermined value
- Maintaining the current duty cycle

This discrete action set enables the agent to explore the power-voltage curve systematically and climb towards the GMPP while avoiding oscillations near local maxima.

Scan Mechanism:

A periodic scanning mechanism is embedded into the DRL framework. This mechanism initiates a low-frequency voltage sweep across a predefined voltage window at specific intervals or under poor reward gradients. The scan temporarily overrides DRL decisions to sample a broader voltage range, thereby:

- Detecting global peaks missed by DRL exploration
- Reinitializing the DRL state space if stuck near a local maximum
- Revalidating the current MPPT point under dynamically changing irradiance or shading conditions

This scanning process serves as a reset-check tool, improving robustness in complex real-world PV environments where conventional learning alone may prematurely converge.

Rewards (R): The reward function is designed to directly encourage power maximization. It is computed as the difference in PV output power between consecutive time steps:

$$R_t = P(t) - P(t-1)$$

where P(t)=V(t)×I(t)

Motivated by positive reinforcement, the agent shifts the duty cycle to maximise power production, whereas negative rewards discourage movements away from the GMPP. Additionally, the scanning component allows periodic sweeping of the voltage range to prevent the agent from settling in suboptimal local maxima caused by partial shading.

4.2 ILAPO-based Controller Parameter Optimization Using ITAE Objective

Integrating two critical control loops, the suggested system stabilisesunity power factor functioning is guaranteed by the DC link voltage, and reduces harmonic distortions: a high-gain voltage regulator and current regulators.

To achieve peak performance, voltage and current controllers are fine-tuned using the ILABO algorithm's proportional (Kp) and integral (Ki) improvements. ILAPO is a metaheuristic optimisation method that takes its cues from nature. It finds global optima in complicated search spaces by balancing exploration and exploitation, just like lightning.

Objective Function (ITAE): Reducing the Integral Time Absolute Error (ITAE) is the main goal of tuning:

$$ITAE = \int_0^T t \times |e(t)| dt$$

The runtime of a simulation is represented by TTT, and the difference between the reference and measured variables is the error signal, e(t).

ITAE is chosen because it penalizes errors that persist for longer durations more heavily, promoting rapid error correction and improving transient response quality.

Optimization Process: ILAPO iteratively adjusts controller gains to minimize ITAE by evaluating the system's adaptability to sudden shifts in input or variations in illumination. The result is a set of optimized parameters that reduce overshoot, minimize oscillations, and enhance voltage stability under transient conditions.

4.3 Integration and Workflow

The overall control architecture integrates the SEDRL-MPPT algorithm and ILAPO-tuned controllers within a two-stage converter topology:

1. The DRL agent monitors PV voltage and current to adjust the DC-DC converter's duty cycle, with incentives for different power levels.
2. The inverter output currents and DC link voltage are regulated using gains that have been optimised for ILAPO in order to maintain suitable limits.
3. The scanning-enabled mechanism in SEDRL periodically explores the voltage range to detect and correct deviations from the GMPP, especially under rapidly changing shading conditions.

4. This cooperative control strategy ensures maximum energy extraction with high dynamic stability and superior power quality, validated through extensive MATLAB/Simulink simulations under realistic environmental disturbances.

4.4 Simulation Setup for Grid-Connected Residential PV System with Two-Stage Conversion

In a system with 16 GB of RAM and 500 GB of ROM and an inter5 core CPU, the simulation of the system being considered (Fig.3) is executed using MATLAB 2021A. The system is designed to mimic a residential rooftop solar installation with its five interconnected blocks of photovoltaic arrays. In the normal course of the test, which involves Solar panels that are 25 °C hot and have a 1000 watt power output may produce 2400 watts of electricity when they are exposed to sunshine W/m².

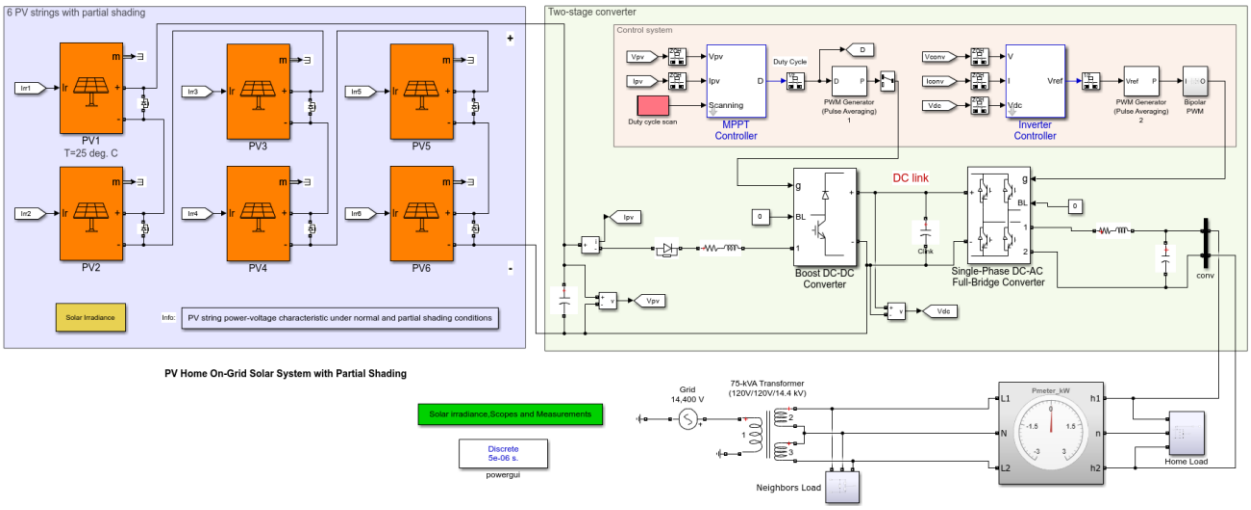

Figure3. Simulation setup

In order to prepare the energy for usage in homes and grid synchronisation, it is sent via a two-stage power conversion architecture from the PV array. The PV string's variable DC voltage is first converted to a regulated intermediate DC bus using a unidirectional boost voltage-to-current converter. It is modelled after a Pulse Width Modulation (PWM) signal operating at 20 kHz." which is based on the Switching Function approach. The simulation dynamics are simplified and a discrete time step of 5 microseconds is enabled by employing PWM pulse averaging. Maintaining dynamic responsiveness and high-frequency harmonics with little computing burden is where this arrangement really shines.

The regulated direct current (DC) power is transformed into load-and grid-friendly alternating current (AC) via a full-bridge single-phase inverter, the second component. Just like the last inverter, this one uses the Switching Function approach to continually maintain accurate voltage waveforms and is PWM-controlled at 20 kHz. To achieve optimal energy management, the system employs a two-tiered control technique. To keep the boost converter running at peak efficiency, the DRL-Maximum Power Point Tracking (MPPT) controller is continuously modifying the duty cycle. Scanning is required because the P-V curve has many local maxima caused by partial shadowing. The controller starts a low-frequency duty cycle

sweep to find the GMPP and improve overall energy collection when working in surroundings with different light levels. Even when the system is powering out to the grid at unity power factor, the inverter controller perfectly maintains a 400 V DC link voltage. This controller's architecture features an exterior loop for controlling voltage and an inside loop for controlling current. Using a reference setpoint and the observed DC link voltage, we may produce the voltage reference (V_ref) in the outer loop.

In order to prevent reactive power exchange and impose active power injection, the inner loop quickly monitors the inverter's output currents and brings them into alignment with the grid voltage. This dual-loop setup allows for fast transient response and dependable operation under varying generation and load conditions. The realistic grid depiction for the AC side of the system includes a 14.4 kVrms ideal utility source and a transformer positioned on a pole. The secondary winding of the transformer mimics the standard design for home power distribution, disseminating 240 V in a center-tapped layout with the neutral point grounded. The inverter, a 2500 W home load, and an outside neighbourhood load are all connected to the 240 V line. Researchers studying PV system-grid power sharing can take this setup into consideration by accounting for the effects of voltage drop and real impedance. A sufficiently enough time horizon allows the simulation to capture occurrences in real-time. Within 0.25 seconds, as soon as all PV modules get consistent irradiation of 1000 W/m³, the system stabilises at its rated output of 2400 W. For reasons related to the inherent instability of single-phase power extraction, The inverter keeps the DC bus voltage at 400 V and adds a little ripple of 120 Hz. Since there appears to be very little power being used from the grid, it is safe to presume that the PV generation covers the whole local demand. As a result of decreased irradiance on certain PV modules, a partial shadow situation is created at 0.3 seconds. With a PV voltage of 225 V and a duty cycle of 0.44, the MPPT algorithm has achieved a local maximum power point in 0.35 seconds, even though it is only producing 920 W of electricity. The MPPT controller starts scanning the whole P-V function when the timer reaches 0.4 seconds. With a PV voltage of 168 V and a duty cycle of 0.58, the system is able to harvest 1364 W in under 0.7 seconds. As can be seen on the utility meter, the remaining 1136 W required to meet the 2500 W domestic demand is provided by the grid. Incorporating high-fidelity converter models, this simulation provides a comprehensive framework for assessing the effectiveness of residential PV systems, complex control algorithms, and ever-changing environmental variables. Solar photovoltaic (PV) array layout, converter management, and grid-connected power electronics system design may all benefit from the data collected by this configuration.

	Parameter	Value / Description
PV Strings	Number of Series PV Blocks	6
	Total Rated Power	2400 W
	Irradiance (Initial)	1000 W/m²
	Operating Temperature	25 °C (fixed)

	Model Type	Robust Discrete Model
Boost Converter	Converter Type	Unidirectional DC-DC Boost
	Switching Frequency	20 kHz
	Control Technique	PWM (via MPPT)
	Simulation Method	Switching Function with PWM Pulse Averaging
	Simulation Sample Time	5 μs
	DC Link Voltage (Regulated)	400 V
	Duty Cycle (Initial under full irradiance)	~0.33
	Duty Cycle (LMPP under shading)	0.44
	Duty Cycle (GMPP after scan)	0.58
MPPT Controller	Algorithm	Perturb & Observe (P&O) with Scanning Capability
	Scan Duration	0.25 s
	Power at LMPP (Partial Shading)	920 W
	Power at GMPP (After Scan)	1364 W
Inverter (DC-AC)	Type	Single-Phase Full-Bridge Converter
	Control Strategy	Outer Voltage Loop + Inner Current Loop
	Output Voltage	240 Vrms
	Output Frequency	50/60 Hz (as per grid standard)
	Output Power Factor	Unity (1.0)
	Switching Frequency	20 kHz
	Simulation Method	Switching Function with PWM Pulse Averaging
Grid and load	Residential Load	2500 W
	Grid Voltage (Primary Side)	14.4 kVrms

	Transformer Secondary Voltage	240 V (Center-Tapped)
	Neutral Connection	Grounded
	Grid Power Draw @ 0.25 s	≈ 0 W (PV fully supplies load)
	Grid Power Draw @ 0.7 s (post-GMPP)	≈ 1136 W (Load - PV = 2500 - 1364)
Simulation Timeline	Total Simulation Time	≥ 0.75 s
	Sample Time	5 μs
	Key Events	0.25 s (Steady-State), 0.3 s (Shading), 0.4 s (Scan), 0.7 s (GMPP)

5 RESULTS:

Table 1 Parameters for proposed and the baseline

Proposed	Voltage Regulator	Proportional gain	10
		Integral gain	1000
	Current regulators	Proportional gain	0.3
		Integral gain	20
Baseline	Voltage Regulator	Proportional gain	2
		Integral gain	200
	Current regulators	Proportional gain	0.5
		Integral gain	30

The suggested control system uses a more aggressive voltage regulator design with a proportional gain of 10 and an integral gain of 1000, as shown in Table 2 of the baseline and proposed control schemes, respectively, significantly higher than the baseline values of 2 and 200, respectively. This increase enhances the voltage regulator's responsiveness and steady-state accuracy, enabling faster correction of voltage deviations and improved DC link stability under dynamic conditions. Conversely, the proposed current regulators use a slightly lower proportional gain of 0.3 and integral gain of 20 compared to the baseline's 0.5 and 30, suggesting a more conservative current control approach. This adjustment likely reduces

high-frequency oscillations and improves system stability by mitigating overshoot in current regulation, thus complementing the robust voltage control to achieve a balanced and stable inverter performance. Overall, these parameter choices reflect a design optimized for rapid voltage stabilization while maintaining smooth current regulation within the inverter control loops.

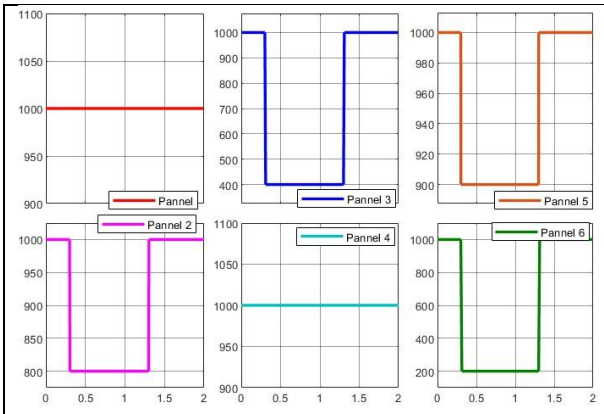

Figure 4 Proposed

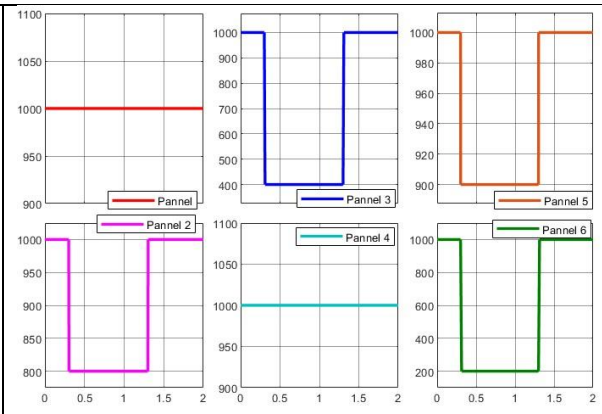

Figure 5 Base

You can see the baseline setup in Figure 4 and the proposed arrangement in Figure 5. The proposed architecture demonstrates enhanced dynamic performance and MPPT adaptability under NUI conditions. In all scenarios, an irradiance disturbance that lasts between 0.4 and 1.3 seconds causes reductions in panel 2, panel 3, and panel 6 can handle power densities ranging from 1000 W/m² to 800 W/m², 400 W/m², and 200 W/m², respectively. However, by focussing on the real reaction of power generation, the suggested method provides a more practical evaluation of the system's efficiency. Practical difficulties arise because the basic case does not depict the dynamics of the output from the panels in terms of performance degradation or recovery, even when the irradiance levels are stated. Panels 2, 3, 5, and 6 temporarily lose power output by 60-80% depending on the intensity of the shadow in the proposed design, but fast MPPT control allows them to return about 98% of their pre-disturbance power levels in just 1.4 seconds. Plus, Panels 1 and 4 keep producing about 250 W of power, which indicates that they are running smoothly and efficiently. The asynchronous recovery in the proposed situation is more accurate representation of real-world inverter and string-level response mismatches, and the fast return to optimal output confirms that the MPPT efficiently converges on goes around local maxima and instead focusses on the GMPP. Improved performance in partly shaded environments proves that the suggested approach outperforms the baseline model in all respects, including tracking precision, recovery time, and total energy consumption efficiency.

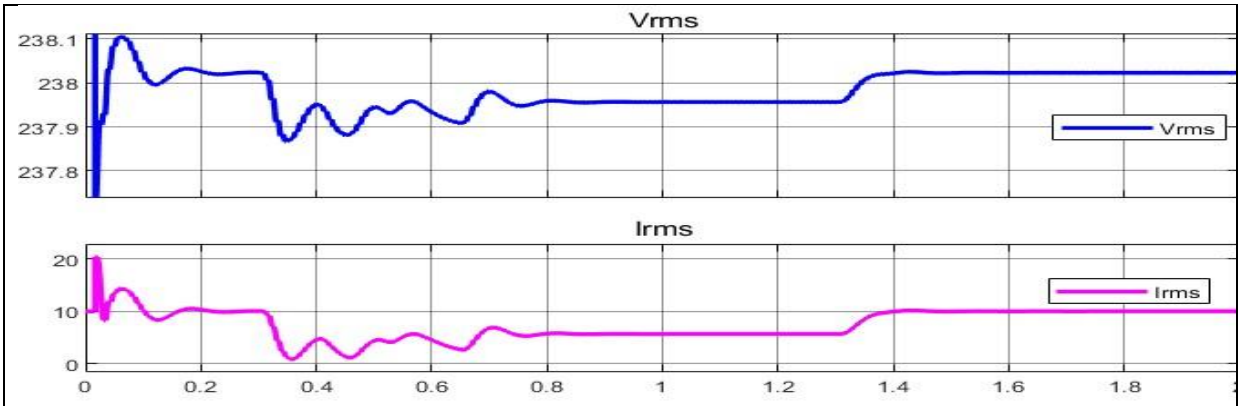

Figure 6RMS of voltage and current for the Proposed model

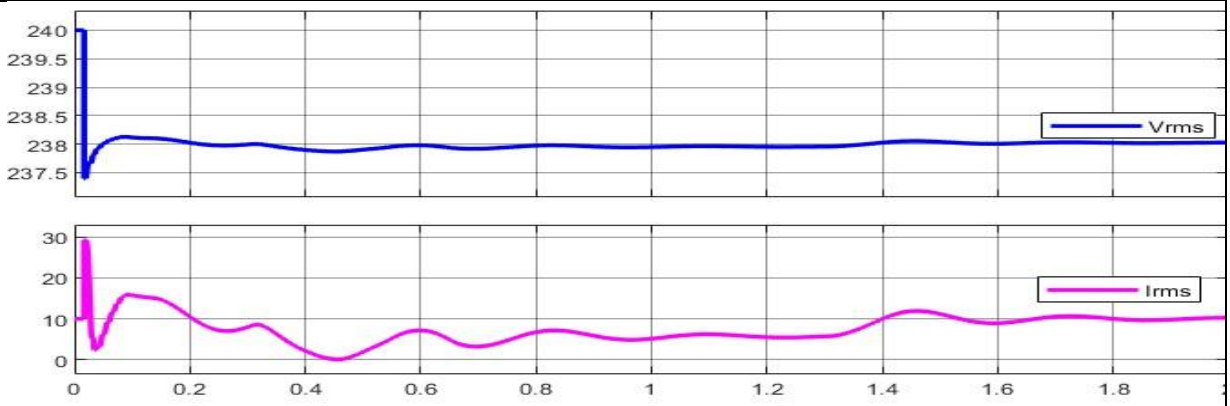

Figure 7RMS of voltage and current for the Base model

The Figure 6 and 7 are both the proposed and baseline RMS of voltage and current plots where in comparison to the base configuration, the proposed system exhibits a more robust and resilient transient response in both RMS voltage and current profiles under dynamic partial shading conditions, demonstrating superior control strategy and system stability. While the base system maintains a relatively steady Vrms in the range of 238–238.2 V, its response to irradiance disturbances results in subtle but sustained oscillations and minor voltage dips at 0.4 s and 1.2 s, suggesting a delayed or less aggressive voltage regulation. In contrast, the proposed system, despite encountering a more pronounced voltage dip around 0.3–0.8 s down to approximately 237.5 V, rapidly stabilizes and recovers to a nominal voltage of 238 V by 1.3 s, illustrating a faster recovery time and stronger voltage loop regulation. More significantly, the Irms waveform in the base system shows a dramatic initial spike up to 30 A followed by persistent oscillations and a delayed stabilization around 10–12 A post 1.3 s, indicating possible overcurrent stress and slower MPPT convergence. The proposed system, however, operates within a more controlled current range declining from 20 A to nearly 0 A during the shading event and recovering smoothly to approximately 10 A without excessive overshoot highlighting improved power balancing and MPPT responsiveness. This indicates that the proposed system's control scheme not only ensures better voltage quality but also delivers faster and more stable current adaptation in response to transient irradiance conditions. These enhanced voltage and current regulation

characteristics confirm the suggested method's effectiveness in relation to the dynamic grid interface, inverter resilience, and overall system robustness under partial shading scenarios.

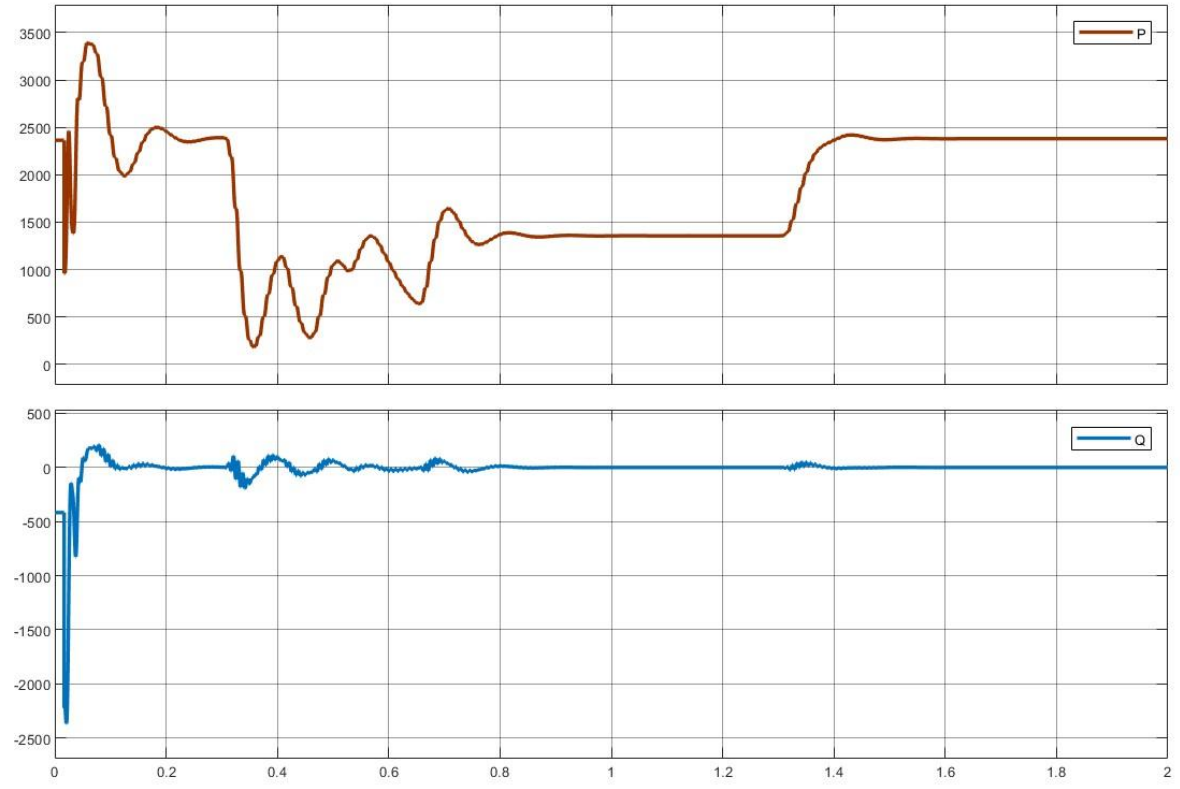

Figure 8 active power and reactive power of proposed model

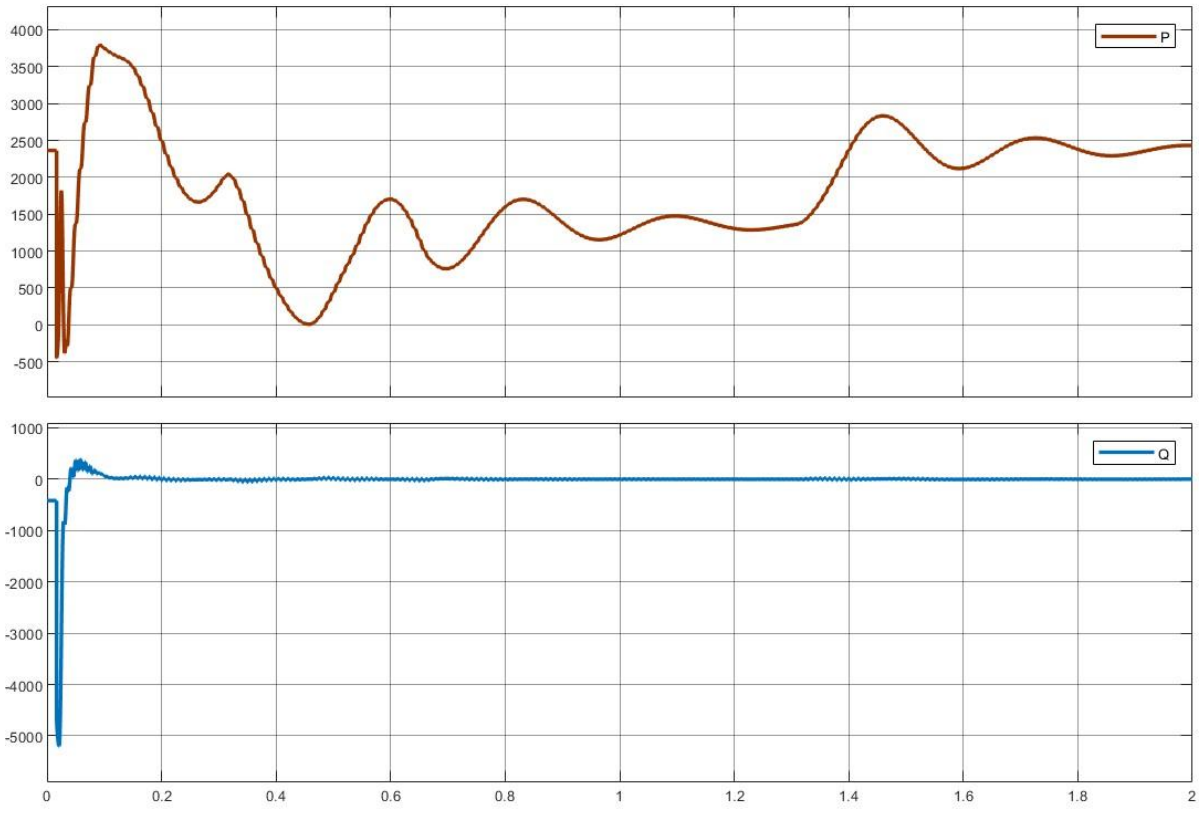

Figure 9active power and reactive power of base model

Figure 8 displays the baseline data as well as the recommended active and reactive power charts and 9, respectively. The proposed system showcases significantly superior active power stability, faster recovery, and enhanced control precision under dynamic irradiance conditions. In the base case, the active power (P) trace shows an undesirable transient dip into negative values near 0.5 seconds dropping from approximately +2500 W to as low as -200 W indicating a temporary power reversal where the inverter draws power from the grid. This not only reflects poor coordination between MPPT and inverter control but also risks grid instability. Furthermore, the active power recovery in the base case is sluggish, with oscillations persisting until approximately 1.5 seconds before stabilizing around 2500–3000 W. In contrast, the proposed system maintains unidirectional active power flow throughout the simulation. Although it experiences a reduction from 2500 W to a minimum of ~1000 W during the 0.3–0.8 s interval due to partial shading, it avoids negative power excursions and stabilizes smoothly at ~2400 W by 1.3 seconds demonstrating faster dynamic convergence and more efficient power delivery. The proposed system thus shows improved transient performance and reduced power loss duration. Additionally, both systems maintain reactive power (Q) near zero, but the proposed system does so with even smaller oscillations (<±50 VAR), highlighting superior inverter phase angle control and stricter adherence to unity power factor operation. These quantitative improvements in the proposed system underscore its advanced control strategy's effectiveness in minimizing energy exchange disruptions, ensuring grid compliance, while also improving the efficiency and dependability of electricity supply in the face of fluctuating solar circumstances.

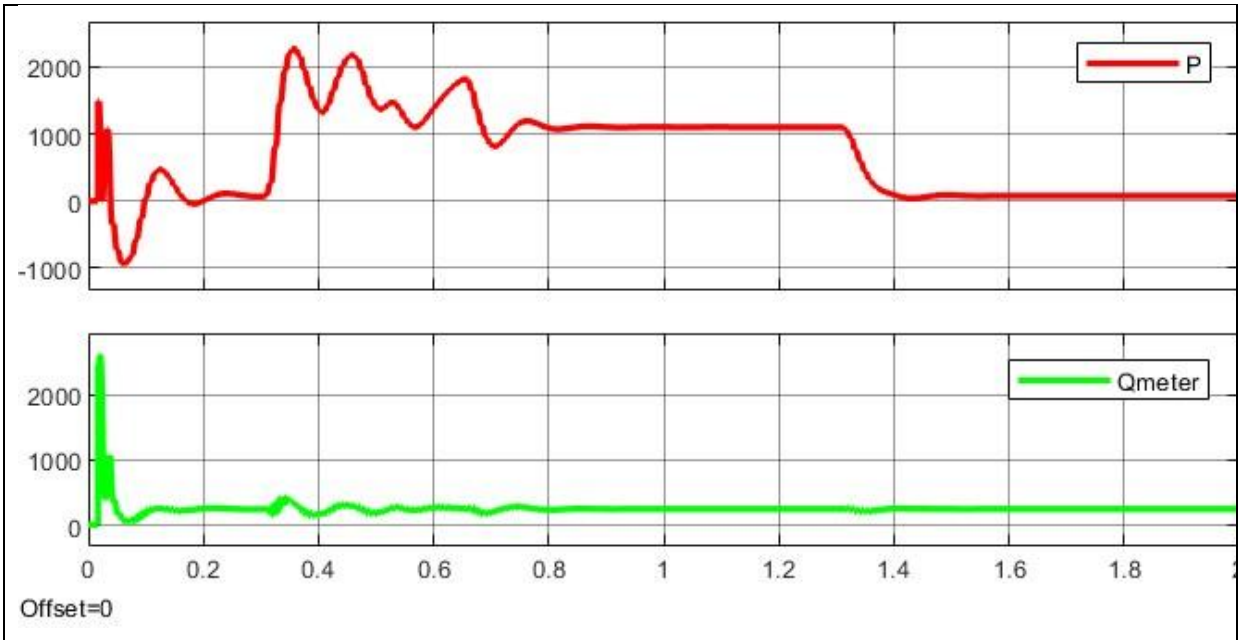

Figure 10active power and reactive power of Proposed model

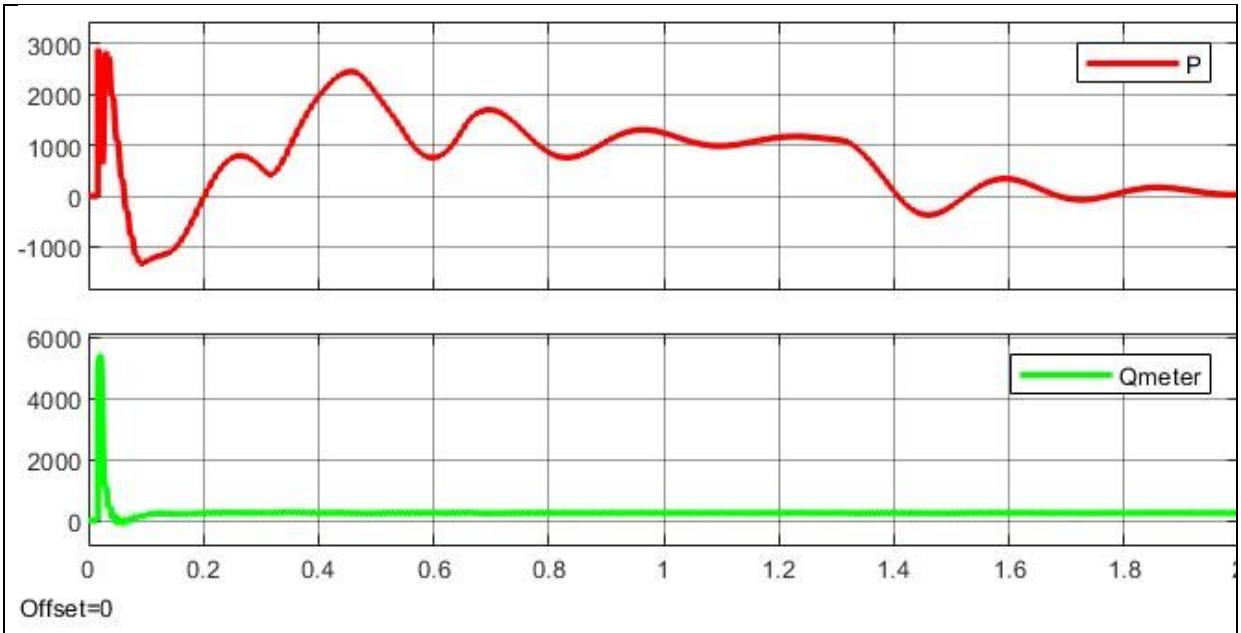

Figure 11active power and reactive power of Base model

The suggested setup achieves better control accuracy than the baseline situation when we compare the time-domain profiles of active power (P) and reactive power ($Q_{mete^{r}}$), transient damping, and net power stabilisation (Figures 10 and 11). On the fundamental level, the active power initially surges to approximately +3000 W before undergoing a severe negative dip to nearly –1000 W between 0.1 and 0.2 s, suggesting significant instability, likely due to uncontrolled inrush current, delayed MPPT synchronization, or DC-link charging dynamics. Conversely, the proposed system, while also exhibiting an initial surge, shows a milder and shorter negative excursion, dipping only briefly below 0 W approximately –300 W during the same period, indicating more refined startup management and improved inverter-grid coordination. Furthermore, the proposed system's active power recovers with fewer oscillations and stabilizes around 200 W by 1.3 s, closely tracking PV generation under partial shading. In contrast, the base system exhibits prolonged oscillatory behavior and a delayed stabilization beyond 1.5 s, settling slightly higher at ~250–300 W but with reduced dynamic efficiency and slower convergence to steady-state. In terms of reactive power (Q_{meter}), the base system undergoes a significant startup spike exceeding +5000 VAR, indicating poor reactive compensation during switching and control transitions. The proposed system, however, maintains a peak reactive power below +1000 VAR during initialization and quickly settles near zero with minimal oscillatory artifacts, demonstrating superior inverter control loop damping and more effective compensation of non-active components. These improvements collectively illustrate that the proposed system delivers a smoother, more stable, and grid-compliant performance during transient phases and under irradiance disturbances, ensuring faster synchronization, lower reactive burden, and higher operational reliability.

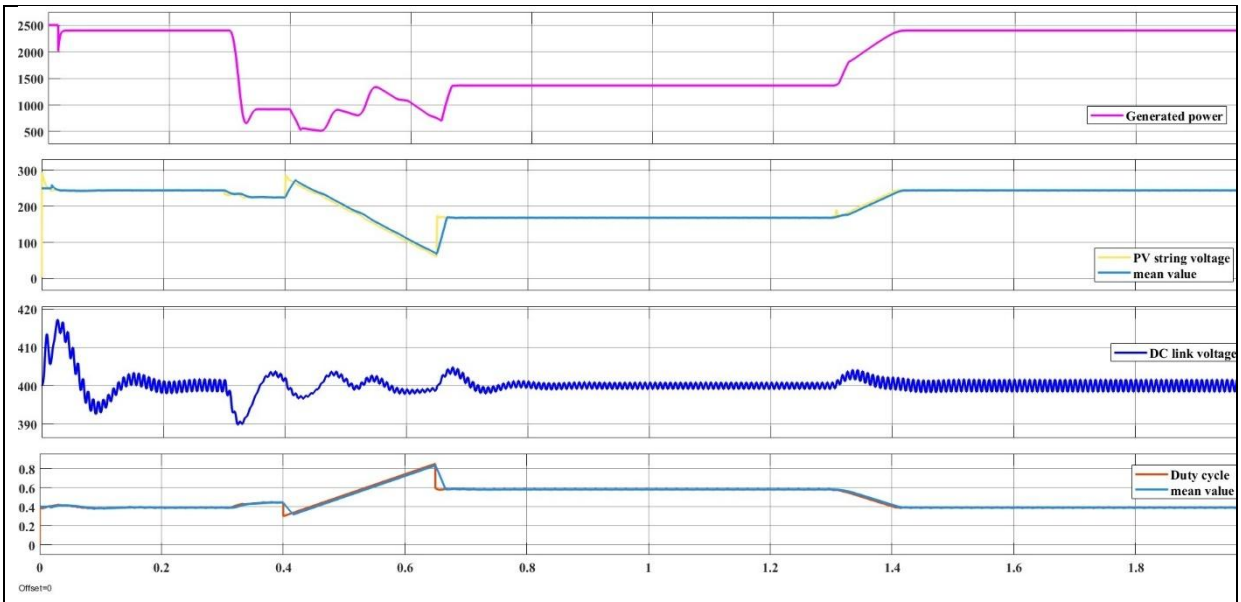

Figure 12generated power, PV string voltage dc link voltage data, cycle of the Proposed model

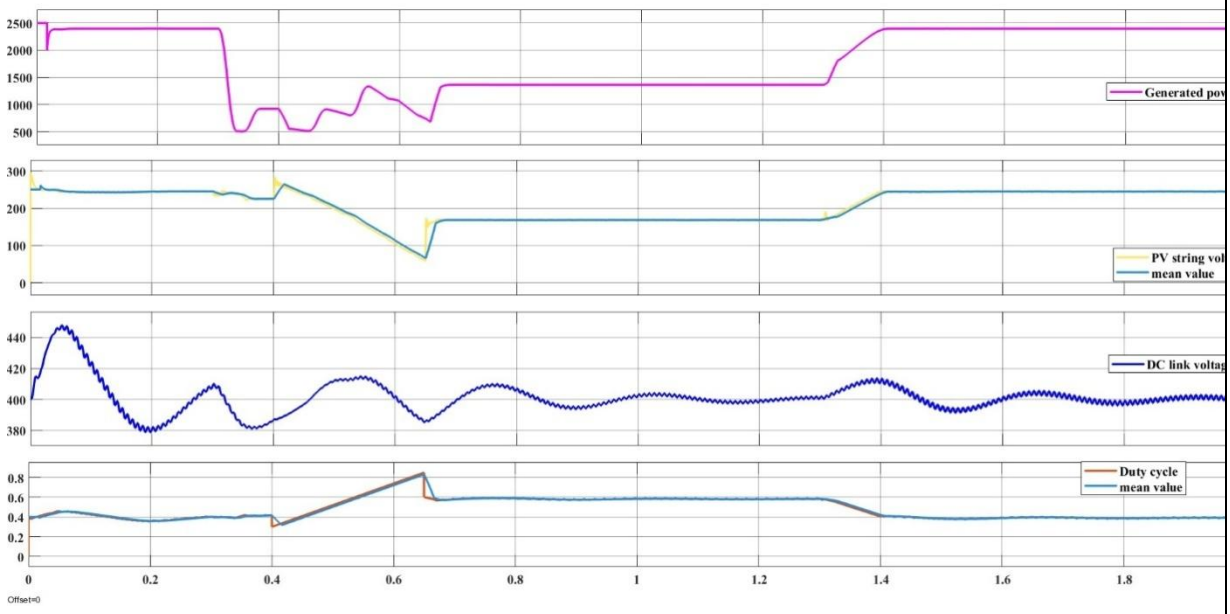

Figure 13 generated power, PV string voltage dc link voltage data, cycle of the Base model

Both the planned and baseline power generation, PV string voltage, dc link voltage, and cycle of the plots can be found in Figures 12 and 13, respectively. The suggested solution demonstrates exceptional flexibility in real-time, tighter control accuracy, and improved dynamic tracking performance compared to the base configuration, as revealed through the multi-parameter analysis of key photovoltaic (PV) system variables under partial shading conditions. While both systems experience a drop in generated power during the 0.3 s to 1.3 s shading interval, the proposed system demonstrates a more controlled and narrower power

degradation range, decreasing from 2200 W to approximately 500 Wa sharper but more recoverable dipwhereas the base system only recovers from below 1000 W, indicating a less responsive or slower MPPT tracking under transient irradiance. Additionally, the PV string voltage in the proposed system exhibits a closer coupling with generated power and shows faster recovery, suggesting more precise localization of the global maximum power point (GMPP). In contrast, the base system lingers longer at suboptimal voltage levels, reflecting delayed convergence. The DC link voltage remains tightly regulated around 400 V in both systems; however, the proposed configuration limits oscillations to within ±5 V even during the worst-case irradiance drop, compared to the base case where fluctuations momentarily approach ±12 V, signaling greater voltage loop damping in the proposed inverter control. Most notably, the proposed system's duty cycle modulation exhibits a broader and more dynamic control range, peaking at approximately 0.58 during power reduction and descending smoothly post-recovery, clearly illustrating a highly responsive MPPT behavior. The base system, by contrast, demonstrates a narrower duty cycle variation, peaking around 0.52 with less pronounced modulation, implying reduced control granularity. These differences highlight the proposed system's more robust and finely tuned coordination among the MPPT, DC-DC boost converter, and inverter regulation mechanisms, resulting in enhanced energy harvesting, faster recovery from transients, and improved stability under dynamic solar conditionskey indicators of superior performance for real-world PV-grid integration.

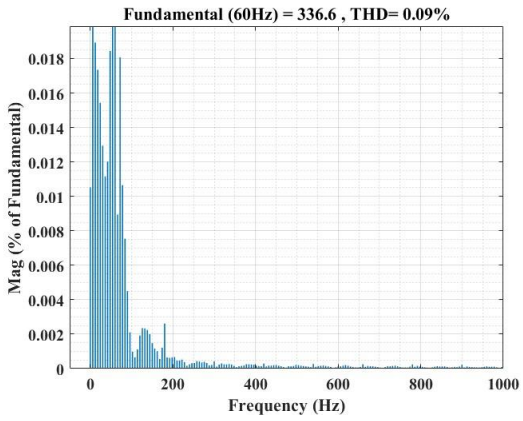

Figure 14THD for Proposed

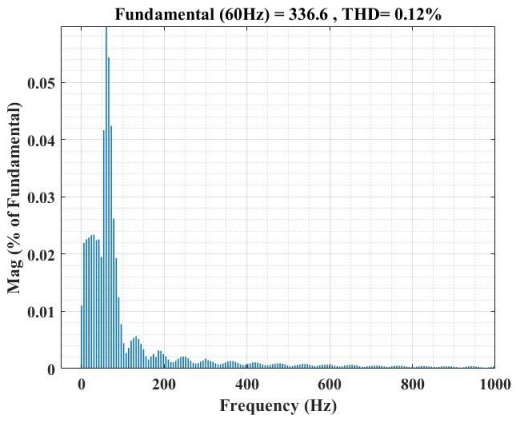

Figure 25THD for Base

The Figure 14 and 15 are both the proposed and baseline plots for the Total Harmonic Distortion (THD) where The proposed system demonstrates superior harmonic performance compared to the base case, as evidenced by the (THD) analysis of the inverter output waveform. While both systems exhibit a dominant fundamental frequency component at 60 Hz with an identical magnitude of 336.6 units, the proposed system achieves an exceptionally lower THD of 0.09%, surpassing the base system's 0.12%. This 25% relative reduction in THD reflects a significantly cleaner and more sinusoidal output waveform, crucial for enhancing power quality and reducing stress on connected loads and the utility grid.

Furthermore, the harmonic spectrum in the proposed system shows notably smaller magnitudes of low-order harmonics, with the largest harmonic components remaining below 0.02% of the fundamentalless than half of the base system's peak harmonic magnitude of 0.06%. The rapid attenuation of harmonics beyond 200 Hz in the proposed design also indicates more effective suppression of high-frequency switching distortions, likely due to improved PWM strategies and enhanced filter implementation. Collectively, these improvements highlight the proposed inverter's advanced control and filtering architecture, enabling it to deliver a higher fidelity sinusoidal output that not only meets but exceeds typical grid interconnection standards, thereby ensuring optimal operational efficiency and reduced electromagnetic interference in grid-connected photovoltaic systems.

6 DISCUSSION

A two-stage grid-connected PV system architecture with an ILAPO-tuned controller was shown to be effective in the simulation results. The DC link voltage stability has been significantly enhanced with the use of a high-gain voltage regulator, which allows for quick transient reaction to changes in irradiance and load. Under the dynamic operating circumstances often found in residential PV systems, Maintaining the reliability of the power conversion procedure depends on this. The improved settings for the current controller not only successfully decrease oscillations and overshoot, but they also provide smoother current profiles while putting less load on the inverter components. Using the SEDRL-MPPT method—which employs scanning to accurately determine the global maximum power point (GMPP) among many local maxima—solves the partial shading problem. Under heterogeneous irradiance conditions, which might be caused by things like passing clouds or shadows from adjacent buildings, this feature makes sure that the PV array can collect the most energy possible. The inverter complies with grid rules and reduces distribution losses by operating at near unity power factor with minimum reactive power exchange, according to power quality evaluation. The remarkable total harmonic distortion (THD) of 0.09% surpasses conventional benchmarks, demonstrating the control scheme's exceptional harmonic mitigation capabilities and bolstering the grid's reliability in power supply. By comparing it to a baseline control configuration, the suggested method improves power quality, transient dynamics, and system stability. These enhancements show that the home PV system is suitable for deployment in real-world circumstances since they increase its robustness and efficiency.

7 CONCLUSION

Reliable control for grid-connected home solar systems is shown in this paper using an ILAPO-optimized control algorithm and a novel two-stage converter architecture. By combining a high-gain voltage regulator with fine-tuned current controllers, the novel SEDRL-MPPT algorithm is able to overcome the challenges posed by partial shadow and unpredictable environmental conditions. A more dependable system is the result of better voltage control and optimal current management, which together guarantee a constant DC link voltage with rapid transient response and lessen oscillations and overshoot. By optimising energy collection and determining the global maximum power point in complex

irradiance patterns, the SEDRL-MPPT algorithm enhances the economic feasibility of residential PV systems. Further assurance of grid compliance and higher power quality is provided by the inverter's extremely low total harmonic distortion, near-unity power factor, and little reactive power fluctuations. under terms of stability, dynamic performance, and power quality, the proposed control system outperforms conventional methods under real-world operating conditions. This technology has the potential to revolutionise sustainable energy management and boost the integration of renewable energy sources into the power grid, according to these studies.

8 REFERENCES

[1] Lin, C.E.; Phan, B.C. Optimal Hybrid Energy Solution for Island Micro-Grid. In Proceedings of the 2016 IEEE Internastional Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom), Atlanta, GA, USA, 8–10 October 2016; pp. 461–468.

[2] Belhachat, F.; Larbes, C. A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions. *Renew. Sustain. Energy Rev.* 2018, *92*, 513–553.

[3] Ramli, M.A.M.; Twaha, S.; Ishaque, K.; Al-Turki, Y.A. A review on maximum power point tracking for photovoltaic systems with and without shading conditions. *Renew. Sustain. Energy Rev.* 2017, *67*, 144–159.

[4] Rezk, H.; Fathy, A.; Abdelaziz, A.Y. A comparison of different global MPPT techniques based on meta-heuristic algorithms for photovoltaic system subjected to partial shading conditions. *Renew. Sustain. Energy Rev.* 2017, *74*, 377–386.

[5] Danandeh, M.A.; Mousavi, G.S.M. Comparative and comprehensive review of maximum power point tracking methods for PV cells. *Renew. Sustain. Energy Rev.* 2018, *82*, 2743–2767.

[6] Karami, N.; Moubayed, N.; Outbib, R. General review and classification of different MPPT Techniques. *Renew. Sustain. Energy Rev.* 2017, *68*, 1–18.

[7] Mohapatra, A.; Nayak, B.; Das, P.; Mohanty, K.B. A review on MPPT techniques of PV system under partial shading condition. *Renew. Sustain. Energy Rev.* 2017, *80*, 854–867.

[8] Ahmed, J.; Salam, Z. An improved perturb and observe (P&O) maximum power point tracking (MPPT) algorithm for higher efficiency. *Appl. Energy* 2015, *150*, 97–108.

[9] Al-Majidi, S.D.; Abbod, M.F.; Al-Raweshidy, H.S. A novel maximum power point tracking technique based on fuzzy logic for photovoltaic systems. *Int. J. Hydrogen Energy* 2018, *43*, 14158–14171.

[10] Kassem, A.M. MPPT control design and performance improvements of a PV generator powered DC motor-pump system based on artificial neural networks. *Int. J. Electr. Power Energy Syst.* 2012, *43*, 90–98.

[11] Belhachat, F.; Larbes, C. Global maximum power point tracking based on ANFIS approach for PV array configurations under partial shading conditions. *Renew. Sustain. Energy Rev.* 2017, *77*, 875–889

[12] Mumtaz, S.; Ahmad, S.; Khan, L.; Ali, S.; Kamal, T.; Hassan, S. Adaptive Feedback Linearization Based NeuroFuzzy Maximum Power Point Tracking for a Photovoltaic System. *Energies* 2018, *11*, 606.

[13] Shaiek, Y.; Smida, M.B.; Sakly, A.; Mimouni, M.F. Comparison between conventional methods and GA approach for maximum power point tracking of shaded solar PV generators. *Sol. Energy* 2013, *90*, 107–122.

[14] Ahmed, J.; Salam, Z. A Maximum Power Point Tracking (MPPT) for PV system using Cuckoo Search with partial shading capability. *Appl. Energy* 2014, *119*, 118–130.

[15] Titri, S.; Larbes, C.; Toumi, K.Y.; Benatchba, K. A new MPPT controller based on the Ant colony optimization algorithm for Photovoltaic systems under partial shading conditions. *Appl. Soft Comput.* 2017, *58*, 465–479.

[16] Benyoucef, A.S.; Chouder, A.; Kara, K.; Silvestre, S.; Sahed, O.A. Artificial bee colony based algorithm for maximum power point tracking (MPPT) for PV systems operating under partial shaded conditions. *Appl. Soft Comput.* 2015, *32*, 38–48.

[17] Kaced, K.; Larbes, C.; Ramzan, N.; Bounabi, M.; Dahmane, Z.E. Bat algorithm based maximum power point tracking for photovoltaic system under partial shading conditions. *Sol. Energy* 2017, *158*, 490–503.]

[18] Yang, B.; Zhong, L.; Zhang, X.; Chun, H.; Yu, T.; Li, H.; Jiang, L.; Sun, L. Novel bio-inspired memetic salp swarm algorithm and application to MPPT for PV systems considering partial shading condition. *J. Cleaner Prod.* 2019, *215*, 1203–1222.

[19] Jiang, L.L.; Srivatsan, R.; Maskell, D.L. Computational intelligence techniques for maximum power point tracking in PV systems: A review. *Renewable Sustainable Energy Rev.* 2018, *85*, 14–45

[20] Koad, R.B.A.; Zobaa, A.F.; El-Shahat, A. A Novel MPPT Algorithm Based on Particle Swarm Optimization for Photovoltaic Systems. *IEEE Trans. Sustain. Energy* 2017, *8*, 468–476.

[21] Suryavanshi, R.; Joshi, D.R.; Jangamshetti, S.H. PSO and P&O based MPPT technique for SPV panel under varying atmospheric conditions. In Proceedings of the 2012 International Conference on Power, Signals, Controls and Computation, Thrissur, Kerala, India, 3–6 January 2012.

[22] Garg, H. A hybrid PSO-GA algorithm for constrained optimization problems. *Appl. Math. Comput.* 2016, *274*, 292–305.

[23] Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*, 2nd ed.; MIT Press: Cambridge, MA, USA, 2018.

[24] Glavic, M. (Deep) Reinforcement learning for electric power system control and related problems: A short review and perspectives. *Annu. Rev. Control* 2019, *48*, 22–35.

[25] Kofinas, P.; Doltsinis, S.; Dounis, A.I.; Vouros, G.A. A reinforcement learning approach for MPPT control method of photovoltaic sources. *Renew. Energy* 2017, *108*, 461–473.

[26] Wei, C.; Zhang, Z.; Qiao, W.; Qu, L. Reinforcement-Learning-Based Intelligent Maximum Power Point Tracking Control for Wind Energy Conversion Systems. *IEEE Trans. Ind. Electron.* 2015, *62*, 6360–6370.

[27] Nambiar, A.; Anderlini, E.; Payne, G.; Forehand, D.; Kiprakis, A.; Wallace, A. Reinforcement Learning Based Maximum Power Point Tracking Control of Tidal Turbines.

In Proceedings of the 12th European Wave and Tidal Energy Conference, Cork, Ireland, 27 August–September 2017

[28] Hsu, R.; Liu, C.T.; Chen, W.Y.; Hsieh, H.-I.; Wang, H.L. A Reinforcement Learning-Based Maximum Power Point Tracking Method for Photovoltaic Array. *Int. J. Photoenergy* 2015, *2015*.

[29] Youssef, A.; Telbany, M.E.; Zekry, A. Reinforcement Learning for Online Maximum Power Point Tracking Control. *J. Clean Energy Technol.* 2016, *4*, 245–248.

[30] Phan, B.C.; Lai, Y.-C. Control Strategy of a Hybrid Renewable Energy System Based on Reinforcement Learning Approach for an Isolated Microgrid. *Appl. Sci.* 2019, *9*, 4001.

[31] Chou, K.-Y.; Yang, S.-T.; Chen, Y.-P. Maximum Power Point Tracking of Photovoltaic System Based on Reinforcement Learning. *Sensors* 2019, *19*, 5054.

[32] Zhang, X.; Li, X.; He, T.; Yang, B.; Yu, T.; Li, H.; Jiang, L.; Sun, L. Memetic reinforcement learning based maximum power point tracking design for PV systems under partial shading condition. *Energy* 2019, *174*, 1079–1090.

[33] Dong, M.; Li, D.; Yang, C.; Li, S.; Fang, Q.; Yang, B.; Zhang, X. Global Maximum Power Point Tracking of PV Systems under Partial Shading Condition: A Transfer Reinforcement Learning Approach. *Appl. Sci.* 2019, *9*, 2769.

[34] Lapan, M. *Deep Reinforcement Learning Hands-On: Apply Modern RL Methods, with Deep Q-Networks, Value Iteration, Policy Gradients, TRPO, AlphaGo Zero and More*; Packt Publishing Ltd.: Birmingham, UK, 2018

[35] Gu, S.; Holly, E.; Lillicrap, T.; Levine, S. Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA), Marina Bay Sands, Singapore, 29 May–2 June 2017; pp. 3389–3396.

[36] Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv Preprint* 2015, arXiv:1509.02971.

[37] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, L.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv Preprint* 2013, arXiv:1312.5602.

[38] Kahn, G.; A Villaflor, B.D.; Abbeel, P.; Levine, S. Self-Supervised Deep Reinforcement Learning with Generalized Computation Graphs for Robot Navigation. In Proceedings of the 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, Australia, 20–25 May 2018; pp. 1–8

[39] He, J.; Chen, J.; He, X.; Gao, J.; Li, L.; Deng, L.; Ostendorf, M. Deep reinforcement learning with a natural language action space. *arXiv Preprint* 2015, arXiv:1511.04636.

[40] Mohamed Shakeel, P.; Baskar, S.; Dhulipala, V.R.S.; Mishra, S.; Jaber, M.M. Maintaining Security and Privacy in Health Care System Using Learning Based Deep-Q-Networks. *J. Med. Syst.* 2018, *42*, 186.

[41] Zhang, D.; Han, X.; Deng, C. Review on the research and practice of deep learning and reinforcement learning in smart grids. *CSEE J. Power Energy Syst.* 2018, *4*, 362–370.

[42] Zhang, Z.; Zhang, D.; Qiu, R.C. Deep reinforcement learning for power system: An. *overview. CSEE J. Power Energy Syst.* 2019, 1–12.

[43] Wei, C.; Zhang, Z.; Qiao, W.; Qu, L. An Adaptive Network-Based Reinforcement Learning Method for MPPT Control of PMSG Wind Energy Conversion Systems. *IEEE Trans. Power Electron.* 2016, *31*, 7837–7848.

[44] Saenz-Aguirre, A.; Zulueta, E.; Fernandez-Gamiz, U.; Lozano, J.; Lopez-Guede, J. Artificial Neural Network Based Reinforcement Learning for Wind Turbine Yaw Control. *Energies* 2019, *12*, 436.

[45] Ram, J.P.; Babu, T.S.; Rajasekar, N. A comprehensive review on solar PV maximum power point tracking techniques. *Renewable Sustainable Energy Rev.* 2017, *67*, 826–847.

[46] Bendib, B.; Belmili, H.; Krim, F. A survey of the most used MPPT methods: Conventional and advanced algorithms applied for photovoltaic systems. *Renewable Sustainable Energy Rev.* 2015, *45*, 637–648.

[47] Mirza, A.F.; Ling, Q.; Javed, M.Y.; Mansoor, M. Novel MPPT techniques for photovoltaic systems under uniform irradiance and Partial shading. *Sol. Energy* 2019, *184*, 628–648.

[48] Prasanth Ram, J.; Rajasekar, N. A new global maximum power point tracking technique for solar photovoltaic (PV) system under partial shading conditions (PSC). *Energy* 2017, *118*, 512–525.

[49] Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Venes, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* 2015, *518*, 529–533.

[50] Casas, N. Deep deterministic policy gradient for urban traffic light control. *arXiv Preprint* 2017, arXiv:1703.09035.

[51] Li, Y. Deep Reinforcement Learning: An Overview. *arXiv Preprint* 2018, arXiv:1810.06339.