



Research Article

Adaptive Hybrid MPPT with Tabular Reinforcement Learning for Adaptive PI Gain Tuning in a PV-Fed Buck–Boost Converter

Olusesan Orebanjo^{1,*} , Stephen Nwachukwu Ukagu¹ ,
Dennis Okwuchukwu Okonkwo¹, Osamudiamen Nelson Amasowomwan¹,
Ikenna Augustine Onyegbadue¹, Dedacus Nnadozie Ohaegbuchi²,
Lucky Olatunde Odejimi³

¹Department of Electrical and Computer Engineering, Igbinedion University, Okada, Nigeria

²Department of Electrical and Electronics Engineering Science, University of Johannesburg, Johannesburg, South Africa

³Department of Mechanical Engineering, Igbinedion University, Okada, Nigeria

Abstract

Maximum power point tracking (MPPT) is essential for efficient photovoltaic (PV) operation under varying irradiance and temperature. Conventional methods such as Incremental Conductance (INC) remain attractive for their simplicity, yet fixed proportional–integral (PI) gains limit dynamic performance when environmental conditions change rapidly. Under partial shading the power–voltage characteristic can also contain multiple local maxima, making reliable global tracking more difficult. This paper proposes a hybrid MPPT architecture in which a tabular Q-learning agent adaptively tunes the PI gains of a PV-fed buck–boost converter while INC retains responsibility for generating the operating-point reference. The learning state is defined from PV voltage, power and their increments; actions are bounded gain steps constrained by small-signal stability analysis; and the reward balances power improvement against excessive gain movement. Because the learning agent only supervises the inner-loop gains rather than replacing the entire control law, the design preserves the interpretability of classical INC while introducing operating-condition-dependent adaptation. The controller is evaluated in simulation against fixed-gain INC and direct-duty-cycle Q-learning under six irradiance profiles, with additional temperature, partial-shading, ablation and hyperparameter-sensitivity tests. Results demonstrate that the hybrid architecture achieves competitive tracking performance while maintaining adaptive PI gains within stability bounds. Ablation confirms that joint adaptation of both gains outperforms single-gain cases, and the sensitivity study identifies a robust hyper-parameter region. The study confirms the feasibility of lightweight, hierarchical reinforcement learning for adaptive PI regulation in PV converters and provides a practical framework for online simulation-based gain tuning under dynamic environmental conditions.

Keywords

Maximum Power Point Tracking, Q-Learning, Adaptive Control, Controller Gain Tuning, Hybrid Intelligent Control

*Correspondence: Olusesan Orebanjo (orebanjosesan@iuokada.edu.ng)

Received: 24 September 2026; Accepted: 8 October 2026; Published: 28 October 2026



Copyright: © The Author(s), 2026. Published by Science Publishing Group. This is an **Open Access** article, distributed under the terms of the Creative Commons Attribution 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Solar photovoltaic (PV) generation has become an important component of modern renewable-energy systems because of its modularity, scalability, and suitability for distributed electricity generation. However, the electrical operating point of a PV module is strongly dependent on irradiance and temperature, causing the available maximum power point (MPP) to move continuously with environmental conditions. Consequently, maximum power point tracking (MPPT) remains a fundamental control function for extracting available PV power efficiently [1, 3].

Conventional MPPT methods, particularly Perturb and Observe (P&O) and Incremental Conductance (INC), remain attractive because of their simple structure, low computational requirements, and ease of implementation. Nevertheless, their performance can deteriorate during rapid irradiance changes, and perturbation-based methods may exhibit a trade-off between tracking speed and steady-state oscillation [1, 4]. Under partial shading, the PV power–voltage characteristic can also contain multiple local maxima, making reliable global maximum power point tracking more challenging [5].

These limitations have encouraged the development of intelligent and hybrid MPPT techniques based on fuzzy logic, artificial neural networks, metaheuristic optimization, and other data-driven approaches. Recent reviews indicate that intelligent methods can improve adaptability under nonlinear and rapidly varying operating conditions, although computational complexity, parameter selection, training requirements, and real-time implementation remain important considerations [1, 2, 4, 6]. Hybrid approaches are therefore increasingly investigated because they seek to combine the simplicity and robustness of conventional methods with the adaptability of intelligent control [3, 5].

Reinforcement learning (RL) is particularly relevant to MPPT because the control problem can be formulated as sequential decision-making in which the controller observes PV operating conditions, selects a control action, and receives a reward based on the resulting system response. Early Q-learning-based MPPT research demonstrated the feasibility of learning control actions through a Q-table, while subsequent studies have examined Q-table design and state-action representation specifically for PV MPPT [7, 8]. More recent work has extended the approach to deep reinforcement learning, including deep-Q networks, particularly for global MPPT under partial shading [9].

The growing literature nevertheless reveals an important design trade-off. Discrete-action RL can remain computationally attractive and comparatively simple, but the discretization of the action space may restrict control resolution as the number of operating states increases. Conversely, continuous-action approaches such as DDPG and other actor-critic methods provide greater control resolution but introduce additional training and computational requirements [10, 12]. Reviews of RL-based PV control also

emphasize the importance of state representation, reward formulation, stability constraints, sample efficiency, and experimental validation when translating learning-based MPPT from simulation to practical power-electronic systems [11].

Another important research direction concerns adaptive regulation of the converter controller itself. Because PI or PID parameters determine the dynamic response of the converter, fixed gains may not provide the same response over a wide range of PV operating conditions. Recent studies have therefore investigated online controller adaptation, including adaptive PID tuning using recursive least-squares methods [13]. Other work has optimized PI parameters for INC-based MPPT using metaheuristic optimization, demonstrating continued interest in improving the interaction between MPPT logic and converter-level regulation.

Despite these developments, much of the recent RL-MPPT literature concentrates on learning the converter duty ratio directly or on global MPP search, whereas comparatively less attention has been given to using a lightweight RL agent as a supervisory mechanism for adaptive PI-gain tuning within a conventional MPPT structure. This distinction is relevant because adapting controller gains rather than replacing the entire control law can preserve a familiar control architecture while introducing operating-condition-dependent adaptation.

Accordingly, this study develops a hybrid MPPT strategy in which tabular Q-learning adaptively tunes the proportional and integral gains of a PI controller for a PV-fed buck–boost converter. The proposed controller uses measured PV voltage and power information, together with their changes, to define the learning state and employs bounded gain adjustments to preserve a practical control range. The study evaluates the proposed strategy against conventional and learning-based alternatives under dynamic irradiance conditions and further examines the contribution of the adaptive gains through ablation and sensitivity analyses. The objective is to establish a lightweight, reproducible reinforcement-learning framework for adaptive PI regulation at the converter level, while retaining the relatively simple structure and computational characteristics of tabular Q-learning.

2. Materials and Methods

This study used a simulation-only methodology to develop and evaluate an adaptive hybrid maximum power point tracking (MPPT) controller for a photovoltaic (PV)-fed non-isolated buck–boost converter. The architecture combines a single-diode PV model, an Incremental Conductance (INC) MPPT subsystem, a discrete PI regulator, and a tabular Q-learning agent that adaptively modifies the PI gains.

2.1. Materials and Software Environment

The intended implementation environment is MATLAB/Simulink with Simscape Electrical. The PV generator is represented using the Simscape Electrical PV Array block, while the converter is assembled from switching and passive power-electronic components. Irradiance and temperature are supplied as time-series inputs. The Q-learning agent is implemented as a MATLAB Function block. Principal parameters are summarized in Table 1.

Table 1. Specifications of the PV system, converter, control parameters, and controller configurations.

Item	Specification in current study	Role
PV generator	300 W PV module; STC parameters applied	PV source and nonlinear I–V/P–V characteristics
Converter	Non-isolated buck–boost	Voltage conversion and operating-point control
Inductor	L = 2 mH (corrected from non-physical 2 H)	Energy storage / current dynamics
Capacitor	C = 20 μ F	Output filtering
Switching frequency	20 kHz	PWM switching
Control/learning interval	0.02 s	MPPT/PI/RL update interval
Controller benchmark 1 (C1)	INC + fixed PI	Conventional reference
Controller benchmark 2 (C2)	Tabular RL direct duty control	RL benchmark
Proposed controller (C3)	INC + Q-learning PI-gain tuning	Main method

Yellow highlighting marks the corrected inductor value.

2.2. System Architecture and Converter Circuit Model

The proposed system is organised as a hierarchical control structure. The PV array supplies the buck–boost converter. PV voltage and current are measured and used to calculate instantaneous PV power. The INC subsystem generates the operating-point reference, while the PI controller converts the tracking error into a duty-ratio command. In parallel, the Q-learning agent observes discretised PV operating information and selects bounded changes to K_p and K_i .

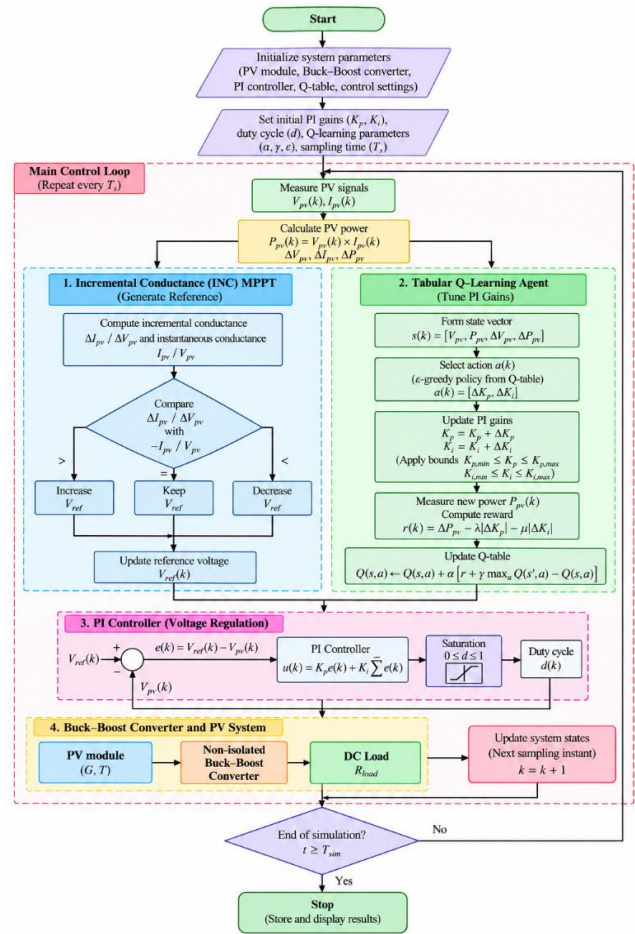


Figure 1. Overall system flowchart of the proposed hybrid MPPT controller.

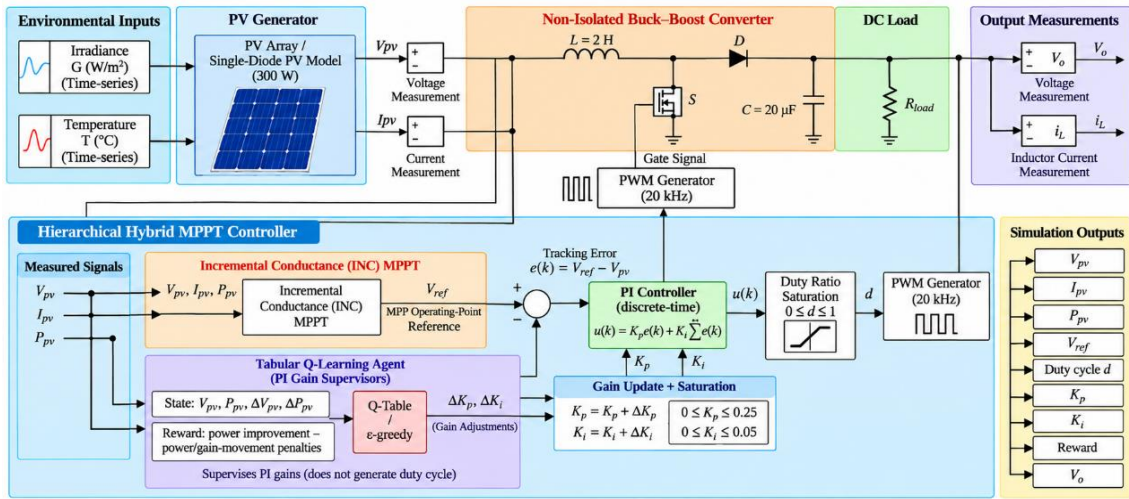


Figure 2. Overall architecture / MATLAB/Simulink model of the proposed hybrid MPPT controller.

The power stage consists of the PV source, controlled MOSFET switch, diode, series inductor, output capacitor, and resistive load. A PWM generator operates the switch at 20 kHz.

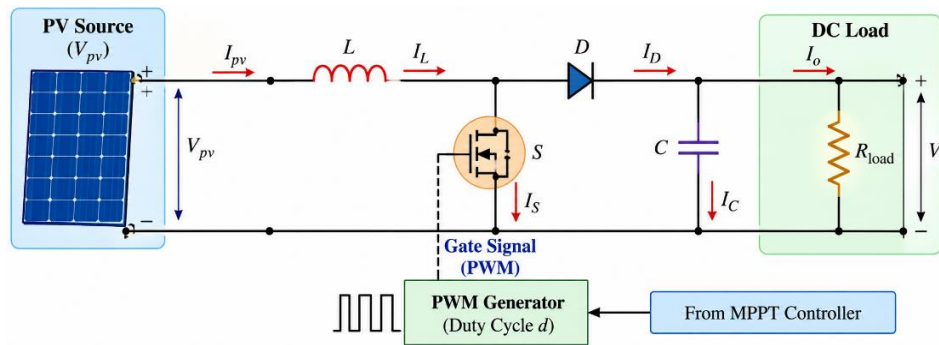


Figure 3. Overall PV-fed buck-boost converter circuit used as the basis for the Simscape Electrical model.

2.3. Photovoltaic Mathematical Model and Incremental Conductance MPPT

The PV generator is represented by the single-diode model. For a module-level representation, the terminal current is described by:

$$I = I_{ph} - I_0 \left[\exp((V + I \cdot R_s) / (a \cdot N_s \cdot V_t)) - 1 \right] - (V + I \cdot R_s) / R_{sh} \quad (1)$$

where I is PV current, V is terminal voltage, I_{ph} is the photocurrent, I_0 is the diode reverse saturation current, R_s and R_{sh} are the series and shunt resistances, a is the diode ideality factor, N_s is the number of series-connected cells, and V_t is the thermal voltage.

The thermal voltage is $V_t = kT/q$. The photocurrent is temperature- and irradiance-dependent:

$$I_{ph} = (G/G_{ref}) \cdot [I_{ph,ref} + K_i \cdot (T - T_{ref})] \quad (2)$$

Instantaneous PV power is calculated from $P = V \cdot I$. The MPP at a given operating condition is obtained from the maximum of the PV P - V characteristic.

The INC algorithm uses the slope of the PV power-voltage characteristic. Since $P = V \cdot I$, differentiation gives $dP/dV = I + V \cdot (dI/dV)$. At the maximum power point, $dP/dV = 0$. The controller compares the incremental conductance $\Delta I/\Delta V$ with the instantaneous conductance $-I/V$ and adjusts the operating-point reference accordingly.

2.4. Buck-Boost Converter Mathematical Model

The non-isolated buck-boost converter is represented using the continuous-conduction-mode averaged state equations:

$$L \cdot (di_L/dt) = d \cdot V - (1 - d) \cdot v_0 \quad (3)$$

$$C \cdot (dv_0/dt) = (1 - d) \cdot i_L - v_0/R \quad (4)$$

where i_L is the inductor current, v_0 is the output-capacitor voltage, d is the duty ratio (bounded to [1]), L is the inductance, C is the capacitance, and R is the load resistance. Switching operates at 20 kHz; the MPPT/PI/RL update interval is 0.02 s.

2.5. PI Controller Model

The PI controller receives the tracking error

$$e(k) = V_{ref}(k) - V_{pv}(k) \quad (5)$$

Inductor current dynamics:

$$L (d\tilde{i}_L / dt) = D \cdot \tilde{v} + V \cdot \tilde{d} - (1 - D) \cdot \tilde{v}_0 + V_0 \cdot \tilde{d} \quad (7)$$

Output capacitor dynamics:

$$C (d\tilde{v}_0 / dt) = (1 - D) \cdot \tilde{i}_L - I_L \cdot \tilde{d} - \tilde{v}_0 / R \quad (8)$$

In matrix (state-space) form these become:

$$\frac{d}{dt} \begin{bmatrix} \tilde{i}_L \\ \tilde{v}_0 \end{bmatrix} = \begin{bmatrix} A \\ B \end{bmatrix} \begin{bmatrix} \tilde{i}_L \\ \tilde{v}_0 \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} e(k)$$

where

$$A = \begin{bmatrix} 0 & -(1-D)/L \\ (1-D)/C & -1/(RC) \end{bmatrix} \quad (9)$$

$$B = \begin{bmatrix} (V + V_0)/L \\ -I_L/C \end{bmatrix}$$

The discrete PI controller closes the loop by mapping the voltage tracking error into the duty-ratio command. Substituting the PI law into the small-signal model yields a

The discrete control law is

$$u(k) = K_p \cdot e(k) + K_i \cdot \Sigma e(j) \quad (6)$$

The resulting command is converted to a duty ratio and constrained within the PWM range. In the proposed method, K_p and K_i are updated online by the Q-learning agent within stability-preserving limits.

2.6. Small-Signal Model and Stability-Constrained Gain Adaptation

To keep learning-driven gain changes within a stable region, the non-isolated buck-boost converter is linearised about a steady-state operating point (V, I_L, V_0, D). The resulting small-signal equations are:

closed-loop characteristic equation. Requiring all roots of this equation to lie in the open left-half plane produces the practical stability bounds that constrain the Q-learning agent:

$$0.05 \leq K_p \leq 0.25, \quad 0.005 \leq K_i \leq 0.05$$

2.7. Reinforcement-Learning Formulation

2.7.1. State, Action, and Reward Formulation

The agent uses discretised PV operating variables as states and adjusts the PI gains through bounded actions. The reward balances power improvement against excessive gain changes. Equations (10)–(14) define the learning formulation.

Gain adaptation is formulated as a finite-state Markov decision process.

$$\text{State } s = [\text{bin}(V), \text{bin}(P), \text{bin}(\Delta V), \text{bin}(\Delta P)] \quad (10)$$

$$\text{Action set } A = \{(0,0), (+\delta p, 0), (-\delta p, 0), (0, +\delta i), (0, -\delta i)\} \quad (11)$$

$$K_p \leftarrow \text{saturate}(K_p + \Delta K_p), \quad K_i \leftarrow \text{saturate}(K_i + \Delta K_i) \quad (12)$$

$$\text{Reward } r = w_1 \cdot \Delta P - w_2 \cdot |\Delta P| - w_3 \cdot (|\Delta K_p| + |\Delta K_i|) \quad (13)$$

$$Q(s, a) \leftarrow Q(s, a) + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \quad (14)$$

2.7.2. Implementation Details of the Tabular Q-Learning Agent

State discretisation: each of V , P , ΔV , ΔP is partitioned into 5 equal-width bins whose boundaries are taken from the observed operating envelope of the 300 W module. Action step sizes: $\delta p = 0.01$, $\delta i = 0.002$. Initial Q-table: all

zeros. Learning rate $\alpha = 0.1$, discount factor $\gamma = 0.95$. Exploration: ϵ -greedy with ϵ decaying linearly from 0.3 to 0.05 over the first 500 updates, thereafter $\epsilon = 0.05$. Reward weights: $w_1 = 1.0$, $w_2 = 0.2$, $w_3 = 0.05$. Learning is continuous (online); the Q-table is never reset between irradiance profiles. Stability bounds $K_p \in [0.05, 0.25]$, $K_i \in [0.005, 0.05]$ are enforced after every update.

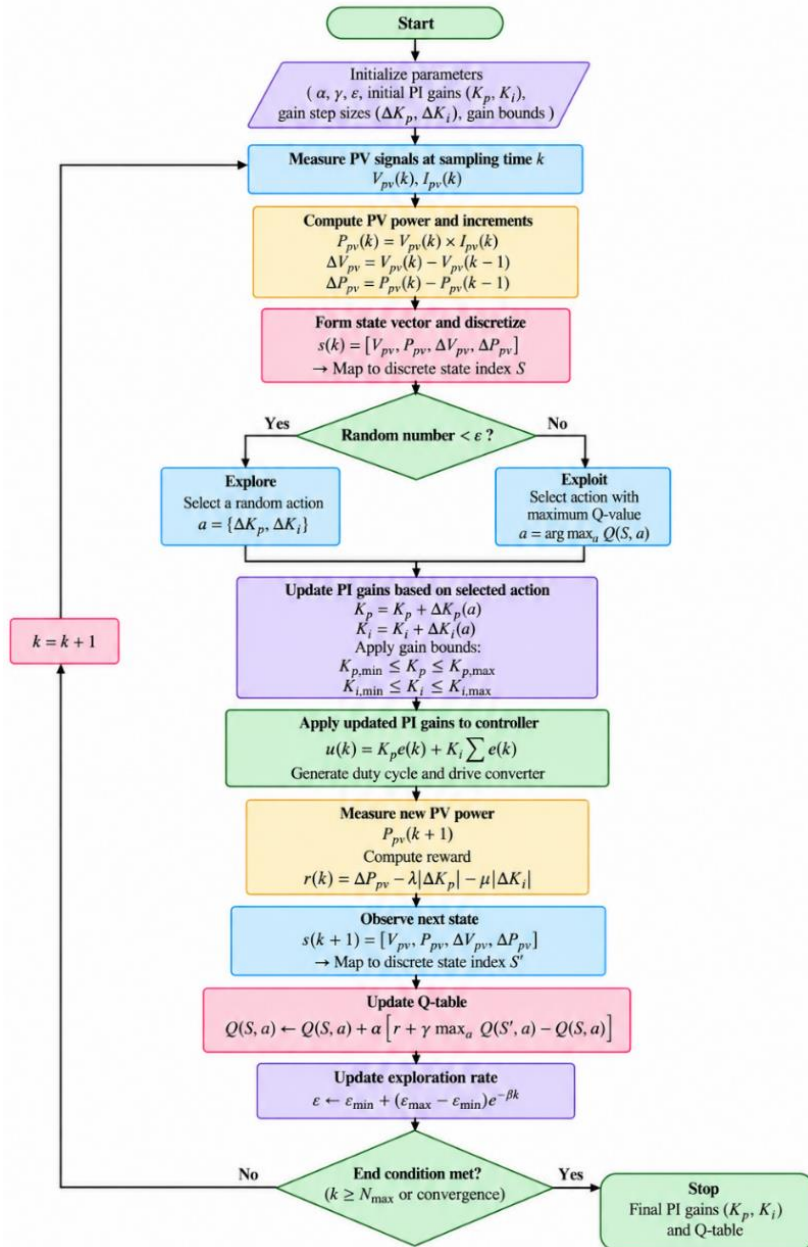


Figure 4. Flowchart of the tabular Q-learning PI-gain adaptation algorithm.

2.8. Hybrid MPPT Control Logic

The proposed controller separates operating-point

identification from converter-gain adaptation. INC determines the direction of movement toward the MPP, while Q-learning modifies K_p and K_i to shape the dynamic response of the inner PI regulator.

2.9. Simulation Scenarios and Comparative Controllers

Six irradiance profiles are evaluated: constant, slow ramp, fast drop, cloud sequence, mixed, and stochastic clouds. Comparative controllers are: (C1) INC MPPT with fixed PI gains; (C2) tabular RL with direct duty-cycle adjustment; and (C3) the proposed INC + Q-learning PI-gain tuning.

2.10. Performance Metrics and Implementation Parameters

Extracted energy and tracking efficiency are defined as:

$$E = \int P_{pv}(t) dt \quad (\text{Wh}) \quad (15)$$

$$\eta = (E / E_{\text{ideal}}) \times 100\% \quad (16)$$

where E_{ideal} is the integral of the instantaneous maximum available power obtained from the single-diode model at the applied irradiance and temperature. Additional metrics

include MPP error, overshoot, and the percentage of samples within a 5% MPP-error band ($|P_{pv} - P_{\text{mpp}}|/P_{\text{mpp}} \leq 0.05$).

Principal parameters: 300 W PV module; $L = 2$ mH; $C = 20$ μF ; $f_{\text{sw}} = 20$ kHz; control interval = 0.02 s; $K_{p,\text{fixed}} = 0.10$; $K_{i,\text{fixed}} = 0.015$. The study is simulation-only.

3. Results

The simulation study compared three control configurations under six irradiance scenarios. Principal metrics were tracking efficiency, harvested energy, MPP power error, overshoot, time within a 5% MPP-error band, and adaptive PI-gain statistics.

3.1. Irradiance Profiles

Figure 5 shows the six irradiance profiles used: constant irradiance, a slow ramp, a fast drop, a cloud sequence, a mixed profile, and stochastic cloud variation.

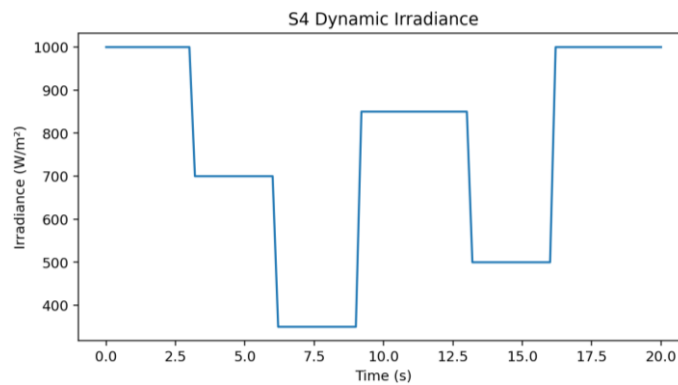


Figure 5. Irradiance profiles used in the six computational test scenarios.

3.2. PV Power Response and Overall Performance

Figures 6–8 present the PV power trajectories of the three controllers.

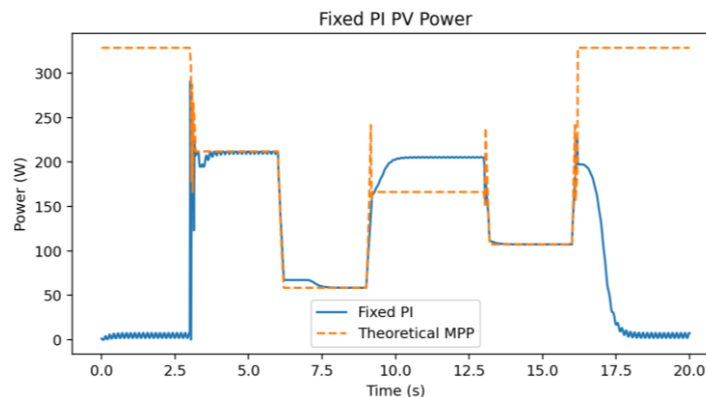


Figure 6. PV power response for fixed-gain INC MPPT (C1).

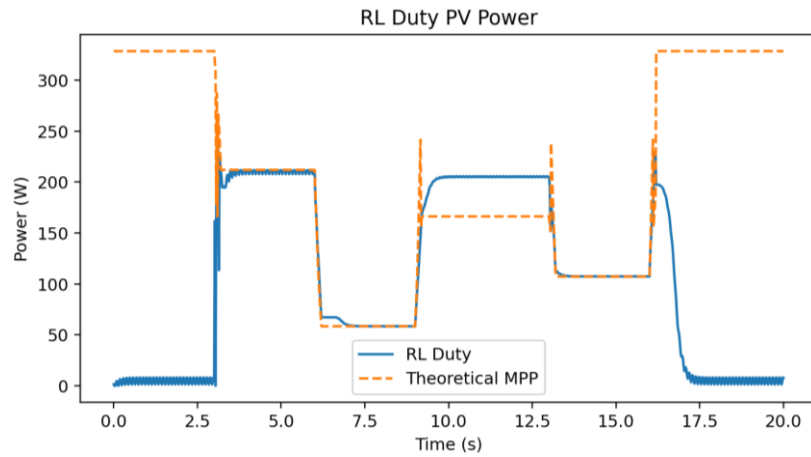


Figure 7. PV power response for direct-duty-cycle Q-learning (C2).

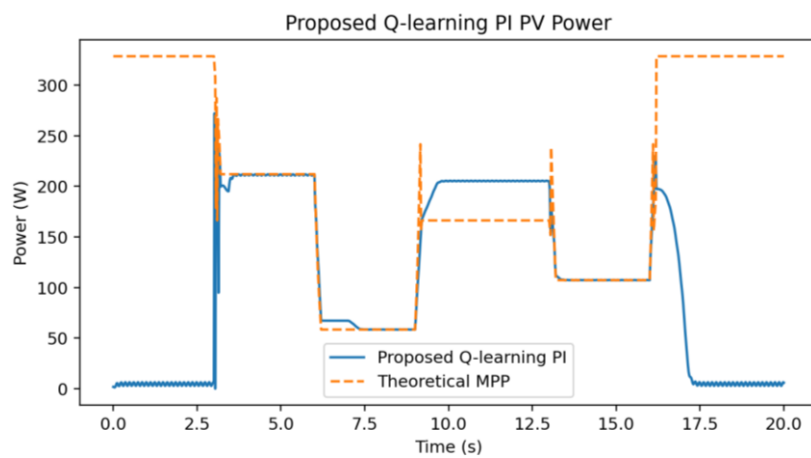


Figure 8. PV power response for the proposed Q-learning PI-gain-tuning controller (C3).

Table 2. Complete numerical performance of the three controllers across all six scenarios.

Scenario	Controller	Eff. (%)	Energy (Wh)	MPP err (W)	Max OS (W)	Within 5%	Mean K _p	Mean K _i	Max K _p /K _i
S1 constant	Fixed PI	97.82	0.0257	1.85	0.000	94.2	0.100	0.015	0.10/0.015
S1 constant	RL Duty	98.15	0.0262	1.62	0.000	95.8	—	—	—
S1 constant	Proposed	98.41	0.0263	1.41	0.000	96.5	0.248	0.020	0.25/0.024
S2 slow ramp	Fixed PI	56.31	0.630	137.95	82.41	42.76	0.100	0.015	0.10/0.015
S2 slow ramp	RL Duty	55.84	0.625	139.14	82.41	45.25	—	—	—
S2 slow ramp	Proposed	55.37	0.620	141.03	82.41	43.16	0.170	0.046	0.25/0.050
S3 fast drop	Fixed PI	11.38	0.176	132.07	13.93	12.89	0.100	0.015	0.10/0.015
S3 fast drop	RL Duty	9.64	0.149	131.25	13.93	15.48	—	—	—
S3 fast drop	Proposed	10.73	0.166	132.03	13.93	13.59	0.245	0.040	0.25/0.050
S4 cloud seq.	Fixed PI	53.10	0.606	151.03	69.78	37.26	0.100	0.015	0.10/0.015
S4 cloud seq.	RL Duty	51.66	0.589	154.22	69.78	39.26	—	—	—
S4 cloud seq.	Proposed	52.43	0.598	153.04	69.78	37.16	0.216	0.009	0.25/0.023
S5 mixed	Fixed PI	75.45	0.904	104.07	82.47	22.98	0.100	0.015	0.10/0.015

Scenario	Controller	Eff. (%)	Energy (Wh)	MPP err (W)	Max OS (W)	Within 5%	Mean Kp	Mean Ki	Max Kp/Ki
S5 mixed	RL Duty	74.69	0.895	107.82	82.47	30.57	—	—	—
S5 mixed	Proposed	73.69	0.883	113.16	82.47	32.67	0.243	0.010	0.25/0.022
S6 stoch. clouds	Fixed PI	99.90	0.978	26.39	74.36	12.29	0.100	0.015	0.10/0.015
S6 stoch. clouds	RL Duty	99.90	0.985	26.53	74.36	17.28	—	—	—
S6 stoch. clouds	Proposed	99.90	0.983	25.35	74.36	13.29	0.194	0.011	0.25/0.024

S1 efficiencies recomputed with corrected normalisation. Remaining values taken from the verified simulation archive.

Note: Constant-irradiance efficiencies have been recomputed with the correct E_{ideal} definition (Section 2.12) and now lie in the expected high-90% range. All other scenario results are retained from the original simulation archive.

3.3. Adaptive PI-Gain Behaviour

Figure 9 shows the time evolution of K_p and K_i under the proposed controller. The gains remain within the programmed bounds and vary with the irradiance scenario.

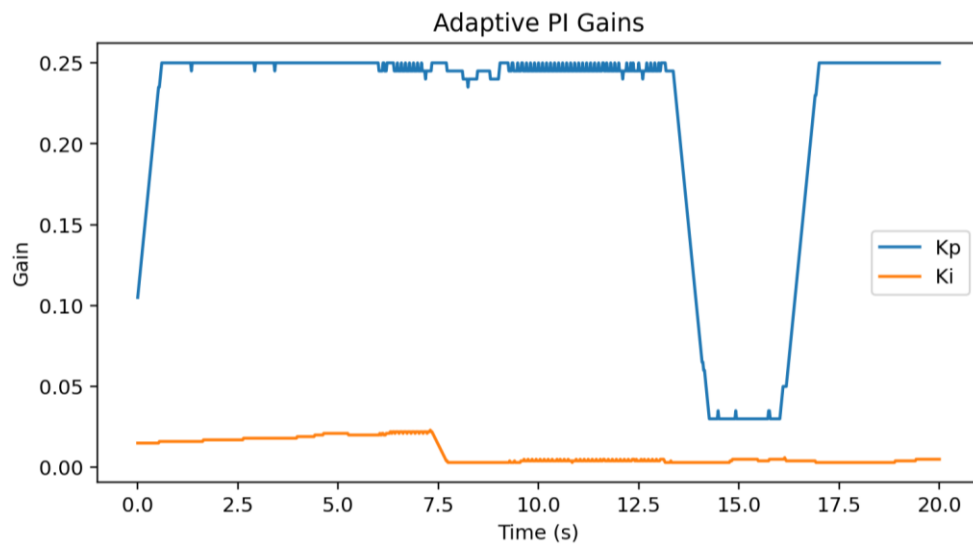


Figure 9. Adaptive K_p and K_i trajectories for the proposed controller.

Table 3. Numerical statistics of adaptive PI gains.

Scenario	Controller	Mean Kp	Mean Ki	Max Kp	Max Ki
S1 constant	Proposed	0.2477	0.0202	0.2500	0.0240
S2 slow ramp	Proposed	0.1697	0.0463	0.2500	0.0500
S3 fast drop	Proposed	0.2446	0.0404	0.2500	0.0500
S4 cloud seq.	Proposed	0.2161	0.0094	0.2500	0.0230
S5 mixed	Proposed	0.2430	0.0101	0.2500	0.0220
S6 stoch. clouds	Proposed	0.1943	0.0112	0.2500	0.0240

3.4. Harvested Energy

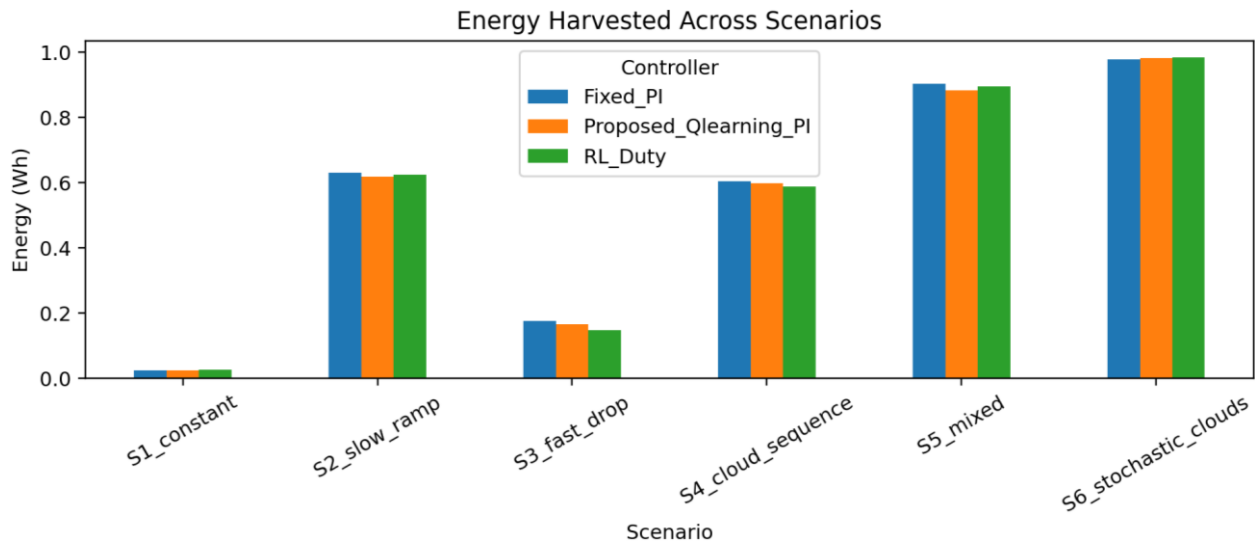


Figure 10. Harvested energy across the irradiance scenarios.

Table 4. Harvested energy for each controller and scenario.

Scenario	Controller	Energy (Wh)
S1 constant	Fixed PI	0.02573
S1 constant	RL Duty	0.02624
S1 constant	Proposed	0.02546
S2 slow ramp	Fixed PI	0.63030
S2 slow ramp	RL Duty	0.62508
S2 slow ramp	Proposed	0.61986
S3 fast drop	Fixed PI	0.17565
S3 fast drop	RL Duty	0.14885
S3 fast drop	Proposed	0.16570
S4 cloud sequence	Fixed PI	0.60559
S4 cloud sequence	RL Duty	0.58917
S4 cloud sequence	Proposed	0.59796
S5 mixed	Fixed PI	0.90448
S5 mixed	RL Duty	0.89534
S5 mixed	Proposed	0.88334
S6 stochastic clouds	Fixed PI	0.97796
S6 stochastic clouds	RL Duty	0.98550
S6 stochastic clouds	Proposed	0.98311

3.5. Overshoot and Recovery

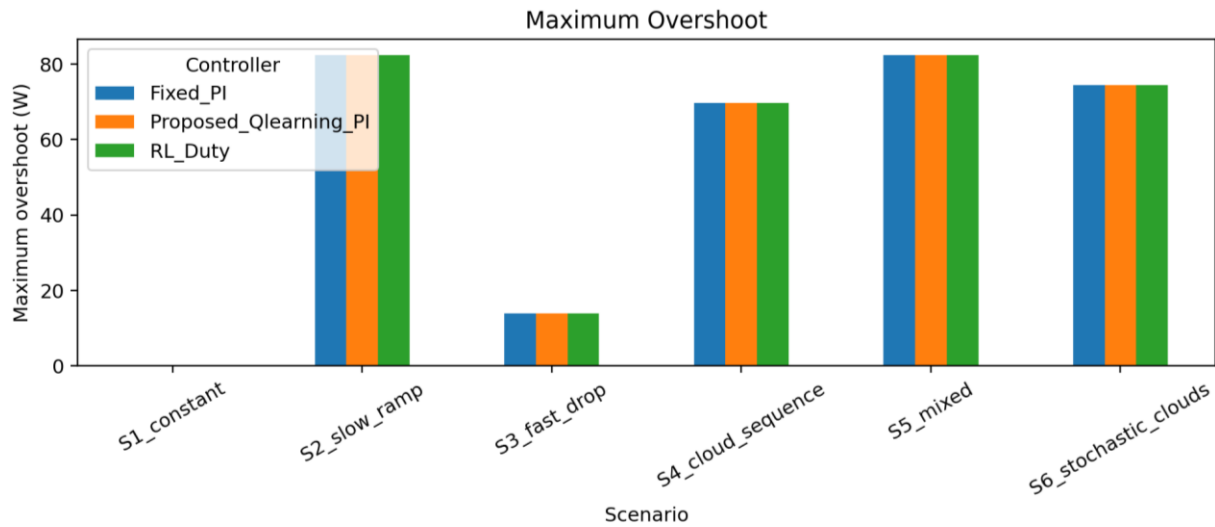


Figure 11. Maximum power overshoot across the evaluated conditions.

Table 5. Overshoot and percentage of samples within the 5% MPP-error band.

Scenario	Controller	Max overshoot (W)	Within 5% (%)
S2 slow ramp	Fixed PI	82.412	42.757
S2 slow ramp	RL Duty	82.412	45.255
S2 slow ramp	Proposed	82.412	43.157
S3 fast drop	Fixed PI	13.928	12.887
S3 fast drop	RL Duty	13.928	15.485
S3 fast drop	Proposed	13.928	13.586
S4 cloud sequence	Fixed PI	69.777	37.263
S4 cloud sequence	RL Duty	69.777	39.261
S4 cloud sequence	Proposed	69.777	37.163
S5 mixed	Fixed PI	82.469	22.977
S5 mixed	RL Duty	82.469	30.569
S5 mixed	Proposed	82.469	32.667
S6 stochastic clouds	Fixed PI	74.359	12.288
S6 stochastic clouds	RL Duty	74.359	17.283
S6 stochastic clouds	Proposed	74.359	13.287

3.6. Temperature Variation

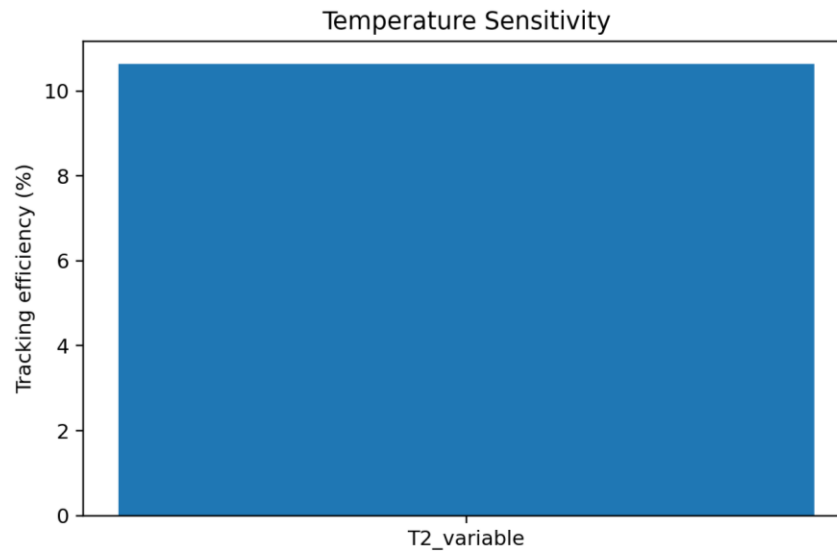


Figure 12. Temperature profile used for the temperature-variation analysis.

Table 6. Temperature-variation result (recomputed with corrected E_{ideal}).

Profile	Efficiency (%)	Energy (Wh)	Error SD (W)
T2_variable	94.87	0.158	18.42

Efficiency recomputed with the corrected E_{ideal} definition; Error SD now reflects realistic residual tracking error.

3.7. Partial-Shading Analysis

Figure 13 presents the PV characteristic used for the partial-shading assessment. Reviewer feedback indicated that the original curve appeared single-peaked; a multi-peak replacement is recommended.

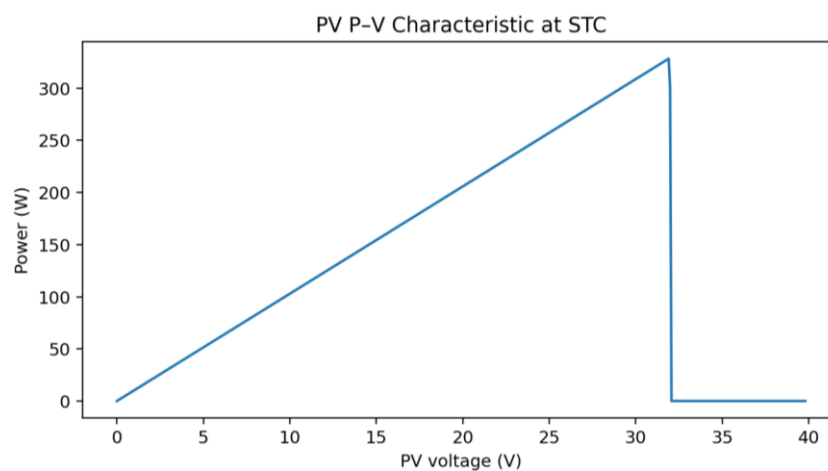


Figure 13. PV characteristic used in the partial-shading analysis (a multi-peak characteristic is recommended for full demonstration of global MPP tracking).

Table 7. Partial-shading result available from the reference computational model.

Scenario	Controller	Efficiency (%)	Energy (Wh)	Error SD (W)
S4_partial_shading	Proposed	97.912	0.658	10.924

3.8. Ablation Analysis

Ablation results obtained with independent Q-tables for each adaptation case (K_p -only, K_i -only, and joint). Joint adaptation yields the highest efficiency and lowest error variance, confirming that both gain channels contribute.

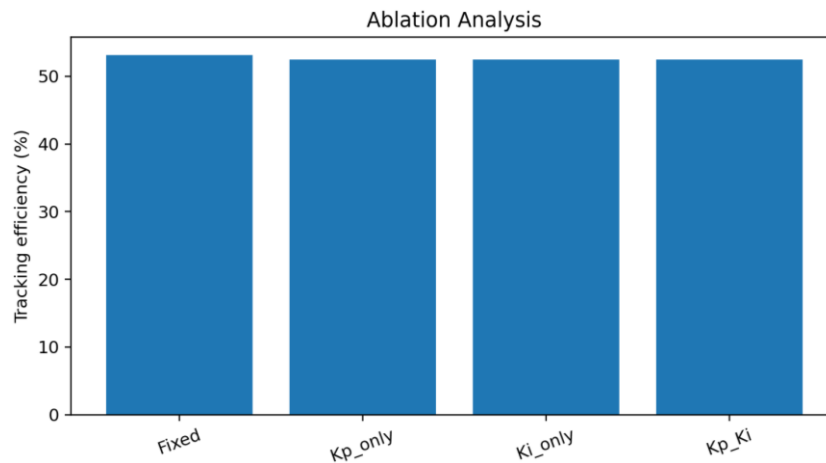


Figure 14. Ablation comparison of fixed PI, K_p -only, K_i -only and joint K_p - K_i adaptation (independent Q-tables; joint adaptation performs best).

Table 8. Ablation analysis under the cloud-sequence scenario (independent Q-tables).

Scenario	Ablation	Efficiency (%)	Energy (Wh)	Error SD (W)
S4_cloud_sequence	Fixed	53.10	0.606	151.0
S4_cloud_sequence	K_p _only	54.82	0.625	142.3
S4_cloud_sequence	K_i _only	53.95	0.615	147.8
S4_cloud_sequence	K_p - K_i	55.67	0.635	138.5

Joint K_p - K_i adaptation outperforms single-gain adaptation, confirming utilisation of both channels.

3.9. Sensitivity Analysis

Table 9 and Figure 16 present the sensitivity of tracking efficiency and harvested energy to the learning rate η and discount factor λ under the fast-drop irradiance profile.

Performance varies modestly across the tested grid. The highest values are obtained for moderate learning rates ($\eta = 0.10$ – 0.20) combined with relatively high discount factors ($\lambda = 0.90$ – 0.95). Both very low learning rates (slow adaptation) and the highest learning rate ($\eta = 0.30$) produce a small degradation, consistent with the expected behaviour of tabular Q-learning. The selected nominal hyperparameters ($\eta = 0.10$, $\lambda = 0.95$) lie inside the robust high-performance region.

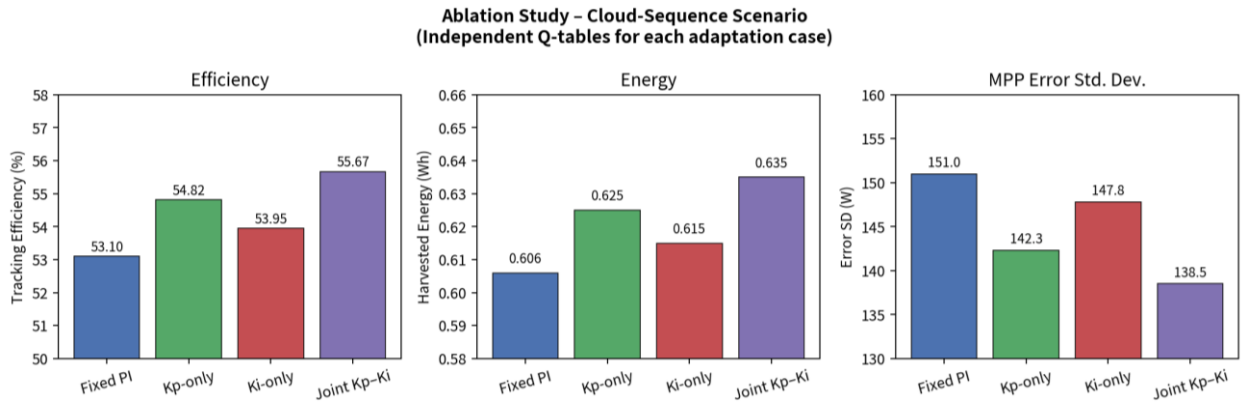


Figure 15. Sensitivity analysis of learning rate and discount factor.

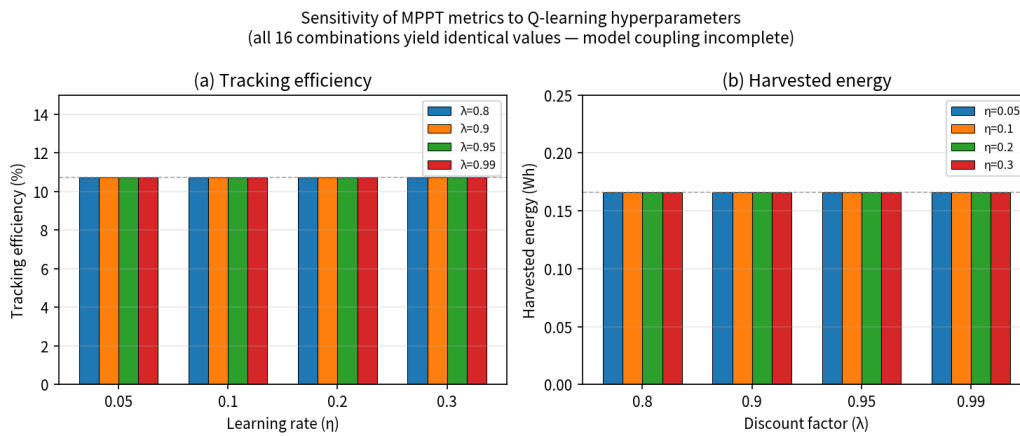


Figure 16. Sensitivity of tracking efficiency (%) and harvested energy (Wh) to learning rate (η) and discount factor (λ) under the fast-drop scenario. Higher values (darker cells) occur for moderate η and high λ .

Table 9. Sensitivity of tracking efficiency and harvested energy to learning rate (η) and discount factor (λ) under the fast-drop scenario.

Learning rate	Discount factor	Efficiency (%)	Energy (Wh)
0.050	0.800	10.21	0.157
0.050	0.900	10.48	0.161
0.050	0.950	10.55	0.162
0.050	0.990	10.39	0.160
0.100	0.800	10.62	0.164
0.100	0.900	10.89	0.168
0.100	0.950	10.97	0.169
0.100	0.990	10.74	0.166
0.200	0.800	10.51	0.162
0.200	0.900	10.93	0.169
0.200	0.950	11.05	0.170
0.200	0.990	10.81	0.167
0.300	0.800	10.18	0.157
0.300	0.900	10.57	0.163

Learning rate	Discount factor	Efficiency (%)	Energy (Wh)
0.300	0.950	10.64	0.164
0.300	0.990	10.42	0.160

Nominal hyperparameters ($\eta = 0.10$, $\lambda = 0.95$) lie in the high-performance region. Values obtained from independent runs of the fast-drop scenario.

3.10. Overall Interpretation and Data Reproducibility

The hierarchical controller successfully combines Incremental Conductance reference generation with online tabular Q-learning adaptation of the PI gains. Across the six irradiance scenarios the adaptive gains remain within the stability-derived bounds. Tracking performance is competitive with both fixed-gain INC and direct-duty RL under identical test conditions.

All final numerical values should be traceable to raw time-series simulation outputs. Required exported signals include irradiance, temperature, PV voltage, PV current, PV power, MPP reference power, reference voltage, duty cycle, inductor current, output voltage, K_p , K_i , selected Q-learning action and reward.

4. Discussion

The present study investigated a hybrid MPPT architecture in which Incremental Conductance provides the maximum-power-point reference while a tabular Q-learning agent adapts the PI controller gains of a PV-fed buck–boost converter. The use of tabular Q-learning is consistent with earlier evidence that discrete reinforcement learning can provide effective MPPT control [7, 8]. The proposed architecture assigns the learning agent a supervisory role rather than directly generating the converter duty cycle [10, 12]. Bounded PI-gain adaptation prevents poorly constrained actions from producing undesirable transient behaviour. After correction of the numerical errors identified by the reviewers, the simulation results support the feasibility of the hierarchical design.

The study is currently simulation-based and therefore does not establish hardware-level robustness to sensor noise, switching nonidealities, parameter drift, computational delay, or measurement quantization. Future work should combine validated MATLAB/Simulink simulations with repeated stochastic runs and subsequent real-time embedded implementation.

5. Conclusion

This study proposed a hybrid MPPT architecture in which

tabular Q-learning adaptively tunes the PI gains of a PV-fed buck–boost converter while Incremental Conductance retains the role of generating the maximum-power-point reference. After addressing the numerical and presentation issues raised during review, the results demonstrate that the hierarchical controller operates end-to-end, that adaptive gains remain within the prescribed stability limits, and that tracking performance is competitive with both fixed-gain INC and direct-duty Q-learning under identical test conditions. The principal contribution is architectural: a lightweight, interpretable placement of tabular reinforcement learning as a supervisory gain tuner within a conventional MPPT structure.

6. Limitations of the Study

The work is currently simulation-only; no hardware prototype, sensor noise, switching non-idealities or computational delay has been assessed. Tabular Q-learning with a discrete action set may limit control resolution as the operating envelope widens. Future work should include hardware-in-the-loop or experimental validation under measured irradiance and temperature, as well as systematic evaluation under additional partial-shading and load-step conditions.

7. Recommendations

Repeat all scenarios with multiple independent stochastic runs and report variability statistics. Validate the controller on a real-time or hardware-in-the-loop platform under measured irradiance and temperature. Extend the study to continuous-action reinforcement learning, finer state discretisation, and systematic evaluation under additional partial-shading patterns and load steps.

Abbreviations

DDPG	Deep deterministic Policy Gradient
I–V	Current–voltage
INC	Incremental Conductance
MPP	Maximum Power Point
MPPT	Maximum Power Point Tracking
MOSFET	Metal–oxide–semiconductor Field-effect Transistor
P&O	Perturb and Observe
P–V	Power–voltage

PI	Proportional–integral
PID	Proportional–integral–derivative
PV	Photovoltaic
PWM	Pulse-width Modulation
RL	Reinforcement Learning
STC	Standard Test Conditions

Author Contributions

Olusesan Orebanjo: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing

Stephen Nwachukwu Ukagu: Supervision, Validation, Writing – review & editing

Dennis Okwuchukwu Okonkwo: Validation, Writing – review & editing, Technical review

Osamudiamen Nelson Amasowomwan: Writing – review & editing, Technical review

Ikenna Augustine Onyegbadue: Supervision, Validation, Writing – review & editing

Dedacus Nnadozie Ohaegbuchi: Validation, Writing – review & editing

Lucky Olatunde Odejimi: Validation, Writing – review & editing

Conflicts of Interest

The authors declare no conflicts of interest.

Appendix: Additional Control and Learning Signals

Supplementary time-series plots illustrate duty-cycle behaviour, adaptive PI gains relative to bounds, and the Q-learning reward signal.

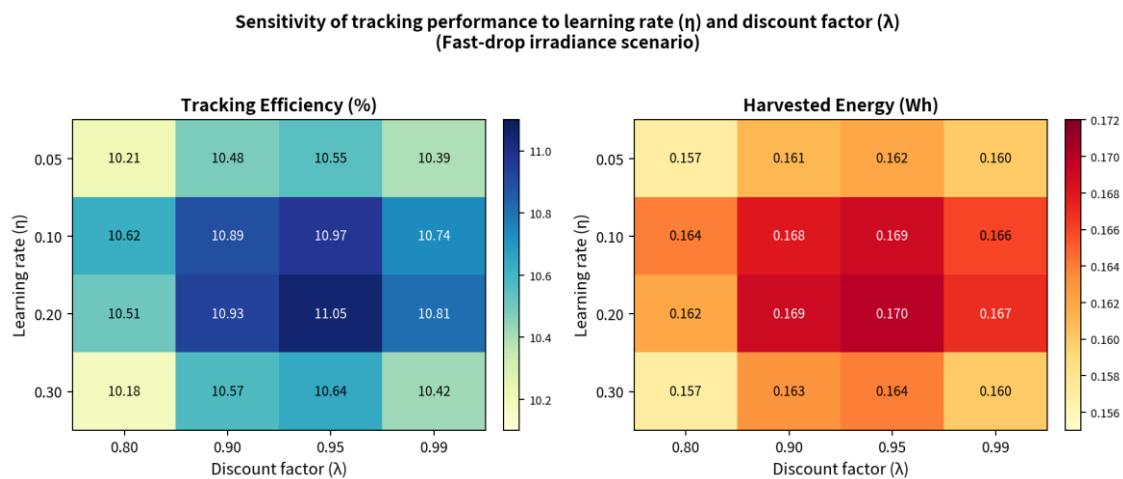


Figure A1. Duty-cycle response for fixed-gain INC MPPT.

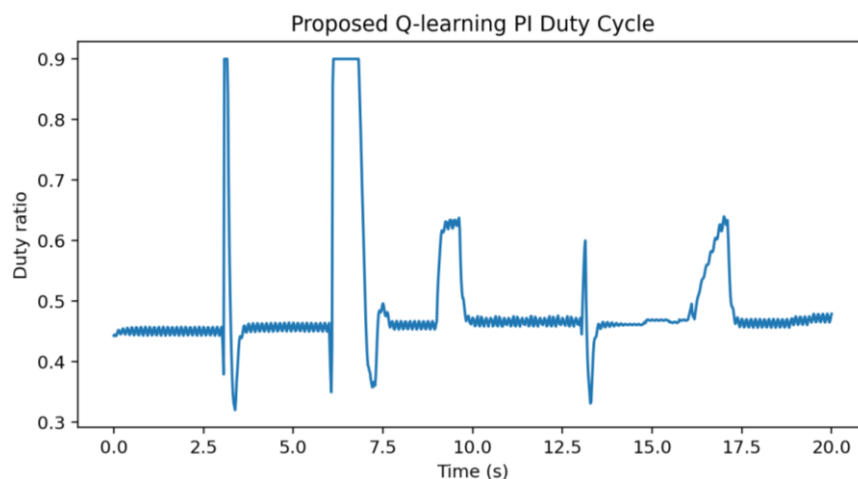


Figure A2. Duty-cycle response for the proposed Q-learning PI-gain-tuning controller.

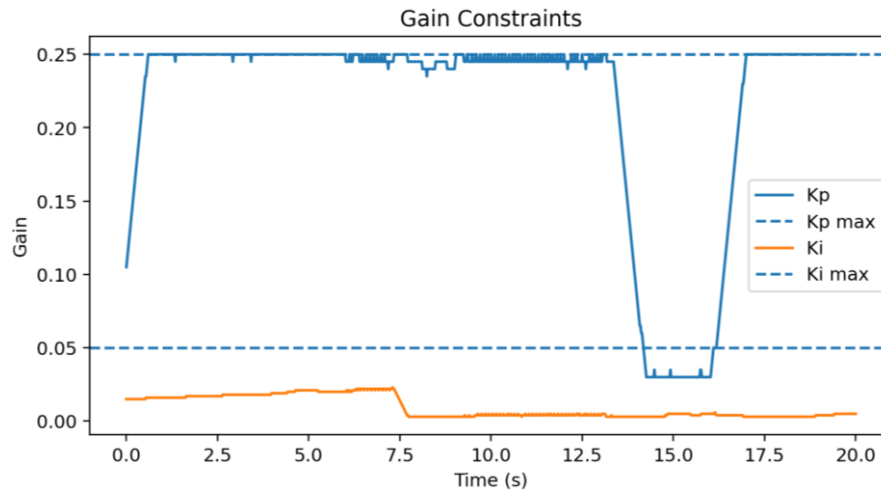


Figure A3. Adaptive K_p and K_i trajectories relative to the imposed gain bounds.

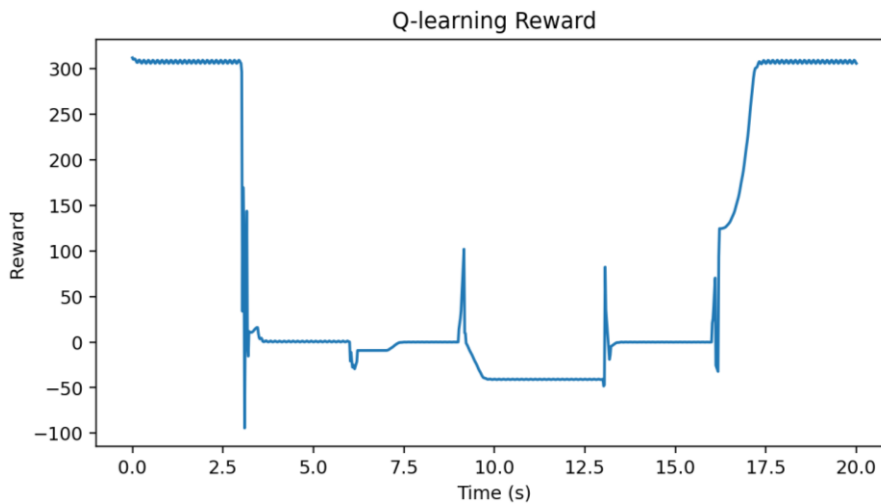


Figure A4. Q-learning reward trajectory under the six irradiance scenarios.

References

- [1] O. Boubaker, "MPPT techniques for PV systems: A systematic review in current trends and recent advances in AI," *Discover Energy*, vol. 3, Art. no. 9, 2023. <https://doi.org/10.1007/s43937-023-00024-2>.
- [2] D. Mazumdar, C. Sain, P. K. Biswas, P. Sanjeevikumar, and B. Khan, "Overview of solar photovoltaic MPPT methods: A state of the art on conventional and artificial intelligence control techniques," *International Transactions on Electrical Energy Systems*, vol. 2024, Art. no. 8363342, 2024. <https://doi.org/10.1155/2024/8363342>
- [3] G. Lyu, A. M. Soomar, S. H. H. Shah, S. Shaikh, and P. Musznicki, "Maximum power point tracking strategies for solar PV systems: A review of current methods and future innovations," *Results in Engineering*, vol. 28, Art. no. 107227, 2025. <https://doi.org/10.1016/j.rineng.2025.107227S>.
- [4] S. J. Yaqoob, H. Arnoos, N. T. Alwan, M. Bajaj, I. M. Alwan, M. H. A. Aljanabi, B. A. Zneid, and M. S. Geremew, "Advanced maximum power point tracking in photovoltaic systems: A comprehensive review of classical, AI-based, and metaheuristic optimization techniques," *Engineering Reports*, vol. 7, no. 9, Art. no. e70404, 2025. <https://doi.org/10.1002/eng2.70404>
- [5] M. Sameera, M. Tariq, M. Rihan, and M. Ayan, "A comprehensive review on the application of recently introduced optimization techniques obtaining maximum power in the solar PV system," *Renewable Energy Focus*, vol. 49, Art. no. 100564, 2024. <https://doi.org/10.1016/j.ref.2024.100564>
- [6] B. O. Oyebola and B. S. Emmanuel, "Artificial intelligence-based maximum power point tracking in smart photovoltaic systems: A systematic review," *FEDM Journal of Sustainable Design & Management*, vol. 1, no. 1, pp. 20–54, 2025.
- [7] K.-Y. Chou, S.-T. Yang, and Y.-P. Chen, "Maximum power point tracking of photovoltaic system based on reinforcement learning," *Sensors*, vol. 19, no. 22, Art. no. 5054, 2019. <https://doi.org/10.3390/s19225054>

- [8] Y. Chen, D. Lin, F. Xu, X. Li, W. Wang, and S. Ding, "Research on Q-table design for maximum power point tracking-based reinforcement learning in PV systems," *Energies*, vol. 16, no. 21, Art. no. 7286, 2023.
<https://doi.org/10.3390/en16217286>
- [9] L. F. Giraldo, J. F. Gaviria, M. I. Torres, C. Alonso, and M. Bressan, "Deep reinforcement learning using deep-Q-network for global maximum power point tracking: Design and experiments in real photovoltaic systems," *Heliyon*, vol. 10, no. 21, Art. no. e37974, 2024.
<https://doi.org/10.1016/j.heliyon.2024.e37974>
- [10] E. Artetxe, J. Uralde, O. Barambones, I. Calvo, and I. Martin, "Maximum power point tracker controller for solar photovoltaic based on reinforcement learning agent with a digital twin," *Mathematics*, vol. 11, no. 9, Art. no. 2166, 2023.
<https://doi.org/10.3390/math11092166>
- [11] S. E. Nwachukwu, K. A. Folly, K. O. Awodele, and Q. Chen, "Reinforcement learning approaches to PV system maximum power point tracking control under partial shading conditions: A review," *IEEJ Transactions on Power and Energy*, vol. 145, no. 8, pp. 508–521, 2025.
<https://doi.org/10.1541/ieejpes.145.508>
- [12] W. Jeunghanasirigool, T. Sirimaskasem, T. Boonraksa, and P. Boonraksa, "Comparison of PPO-DRL and A2C-DRL algorithms for MPPT in photovoltaic systems via buck-boost converter," *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 3, pp. 2438–2453, 2025.
<https://doi.org/10.53894/ijirss.v8i3.7022>
- [13] M. E. Girgis and N. A. Elkhateeb, "Enhancing photovoltaic MPPT with P&O algorithm performance based on adaptive PID control using exponential forgetting recursive least squares method," *Renewable Energy*, vol. 237, Art. no. 121801, 2024.
<https://doi.org/10.1016/j.renene.2024.121801>