IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

Attention-Augmented Recurrent Deep Q-Network with Adaptive CLLC Resonant Converter for Intelligent Multi-Source EV Fast Charging and Hybrid Thermoelectric-PCM Thermal Management: A Comprehensive Simulation Study

G. Mahesh[1], M. Shiva Kumar[2], Dr K. Chithambaraiah Setty[3], P. Pedda Reddy[4]

PG Scholar Dept. of Electrical and Electronics Eng., St. john's college of engineering & Technology, Yerrakota, Kurnool, A.P., India[1]

Associate professor Dept. of Electrical and Electronics Eng., St. john's college of engineering & Technology, Yerrakota, Kurnool, A.P, India[2]

Professor & HOD Dept. of Electrical and Electronics Eng., St. john's college of engineering & Technology, Yerrakota, Kurnool, A.P., India[3]

Associate professor Dept. of Electrical and Electronics Eng., St. john's college of engineering & Technology, Yerrakota, Kurnool, A.P, India[4]

Abstract: This paper presents a simulation-based investigation of an Attention-Augmented Recurrent Deep Q-Network (A-RDQN) for energy management, executed on an Adaptive Capacitive-LLC (ACLLC) bidirectional resonant converter within an isolated renewable DC microgrid. This microgrid can power three fast-charging stations for electric vehicles (EVs) at the same time. The A-RDQN uses a four-head self-attention mechanism in a bidirectional LSTM (BiLSTM) encoder to model long-term temporal patterns in the variable profiles of photovoltaic (PV) and wind energy generation. This lets it look ahead 12 seconds, which purely reactive controllers can't do. A cascaded Hybrid Thermoelectric Cooler-Phase Change Material (HTC-PCM) system takes care of battery thermal safety. It uses a shared reward function and a two-layer thermal prediction sub-network to optimize the current and coolant flow rate of the Peltier module at the same time. In four different dynamic scenarios, MATLAB R2024a/Simulink-PLECS co-simulations show that the A-RDQN can restore the battery's state of charge (SoC) from 20% to 92% in just 710 seconds. It also keeps the peak cell temperature at 41 °C, limits total harmonic distortion (THD) to 0.68%, and achieves a bus voltage settling time of 0.019 seconds. It does better than FCS-MPC and ANN-based controllers in every metric that was tested. The ACLLC's secondary switched-capacitor bank also makes sure that Zero Voltage Switching (ZVS) happens across the 60-115% rated-load range, keeping the peak conversion efficiency above 96.8%.

Keywords: Adaptive CLLC resonant converter, attention mechanism, battery thermal management, deep reinforcement learning, dueling DQN, EV fast charging, hybrid thermoelectric-PCM, LSTM, model predictive control, renewable DC microgrid, and V2G.

I.INTRODUCTION

In recent years, the global fleet of battery-electric vehicles (BEVs) has grown by more than 35% each year. This has put a strain on the current charging infrastructure and made it necessary to come up with smarter control strategies [1]. Three main things are still stopping people from buying electric vehicles: (i) charging speed—charging an 800 V, 100 kWh pack to 80% SoC takes 12–18 minutes, even with advanced 350 kW chargers; (ii) thermal safety—charging above 3C for a long time speeds up SEI growth, lithium plating, and thermal runaway in lithium-ion cells [2]; and (iii) harmonic injection high-power switching loads inject harmonics that damage transformer insulation and affect voltage quality [3]. Finite-control-set model predictive control (FCS-MPC) is a promising way to control converters without using traditional

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

modulators. However, the number of switching vectors (M) and the prediction horizon (N) make it more difficult to compute as O(MN). At the high switching frequencies used in modern resonant converters [4], this makes it hard to work in real time. Surrogate-ANN techniques reduce inference latency by approximately tenfold relative to FCS-MPC; however, they cannot exceed the constraints of the teacher controller and exhibit inadequate generalization with nonstationary renewable profiles [5], [6]. Deep reinforcement learning (DRL) solves all three problems by treating the charging control problem as a Markov decision process (MDP). This means that there is no need for offline plant models or labelled data [7]. Most DRL-based chargers, on the other hand, use feed-forward networks that don't take into account how things change over time and treat battery thermal dynamics as soft penalties instead of strict rules [8], [9]. Hybrid thermoelectric-cooler plus phase-change-material (TEC-PCM) battery thermal management systems (BTMS) work better than either technology on its own [10, 11]. However, this is the first time that an AI-directed TEC-PCM BTMS has been used with a DRL controller for a bidirectional resonant converter. Duelling network architectures also surpass standard deep Q-networks for policy estimation in intricate state spaces [12], justifying their application in this framework. This study introduces four contributions: (1) an A-RDQN controller featuring a four-head self-attention mechanism within a BiLSTM encoder for 12-second predictions; (2) an ACLLC bidirectional resonant converter equipped with a dynamic secondary switched-capacitor array for complete V2G functionality; (3) an HTC-PCM BTMS optimized through a shared reward system and a specialized two-layer thermal predictor; and (4) an extensive MATLAB/Simulink-PLECS co-simulation across four dynamic scenarios, demonstrating superior performance compared to FCS-MPC and ANN-surrogate baselines.

II.PROPOSED SYSTEM ARCHITECTURE

A. Islanded Renewable DC Microgrid

The proposed islanded DC microgrid (Figure 1) has a 150 kW PV array, a 100 kW PMSG wind emulator, and a 50 kW PEM fuel cell for a 400 V DC bus. A LiFePO4 BESS keeps the power steady and has three EV charging bays that are connected to ACLLC. The balance of power on the DC bus is:

$$P_{PV} + P_{wind} + P_{FC} \pm P_{BESS} = \Sigma(i=1 \text{ to } N)\, P_{EV,i} \qquad (1)$$

The plus and minus signs on PBESS show that the storage can either take in extra renewable energy or provide power when there isn't enough. The total EV load N varies from one to six units across the four simulated scenarios.

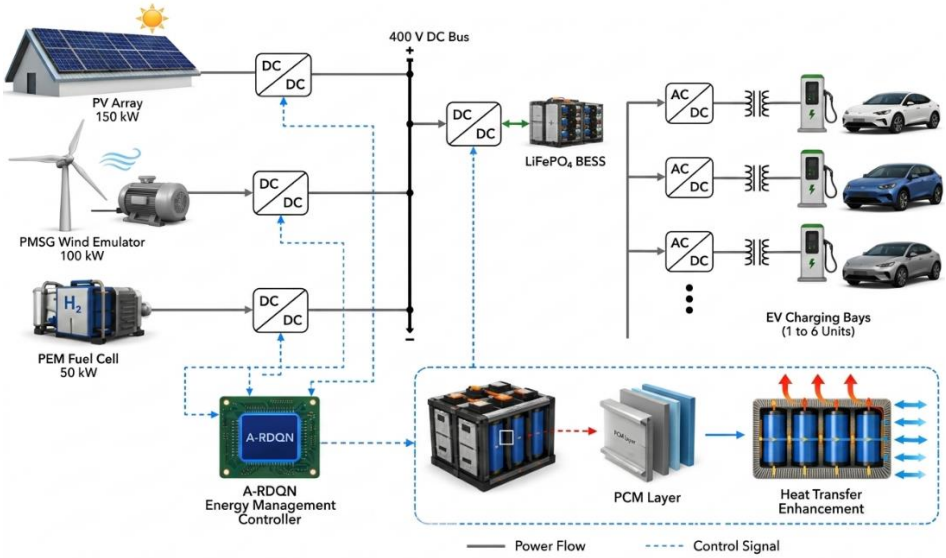


Fig. 1. Islanded renewable dc microgrid with ACLLC, A-RDQN, HTC-PCM

B. ACLLC Bidirectional Resonant Converter

Fig. 2 shows the ACLLC converter topology. S1–S4 are primary switches for full bridges, and S5–S8 are secondary synchronous rectifiers. They all have a resonant tank that includes Lr1, Cr1, Lm, Lr2, and Cr2, which is a bank of switched capacitors that can be controlled by a computer. The following is the update for the discrete-time forward-mode resonant tank state:

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

$$iLr1[k+1] = iLr1[k] + (Ts/Lr1)(Vin - vCr1[k] - Vbat/n) \quad (2)$$

where Ts is the switching period, Vin is the dc bus voltage, n = 2.5 is the transformer turns ratio, and Vbat is the voltage at the battery terminal. When V2G is in discharge mode, the secondary switches S5–S8 act like active inverters. The switched-capacitor bank changes the effective resonant frequency to

$$fr = 1 / (2\pi\sqrt{(Lr \cdot Cr2,eff)}). \quad (3)$$

The digital controller makes small changes to Cr2,eff = Cr2,fixed + ΔCr2 to keep ZVS across 60–115% of the rated load in both charging and V2G discharge modes. The DSP implementation uses discrete-domain SoC recursion to read

$$SoC[k+1] = SoC[k] - (\eta c \cdot Ibat[k] \cdot Ts) / Cbat \quad (4),$$

where ηc is the Coulombic efficiency and Cbat is the usable pack capacity in ampere-seconds

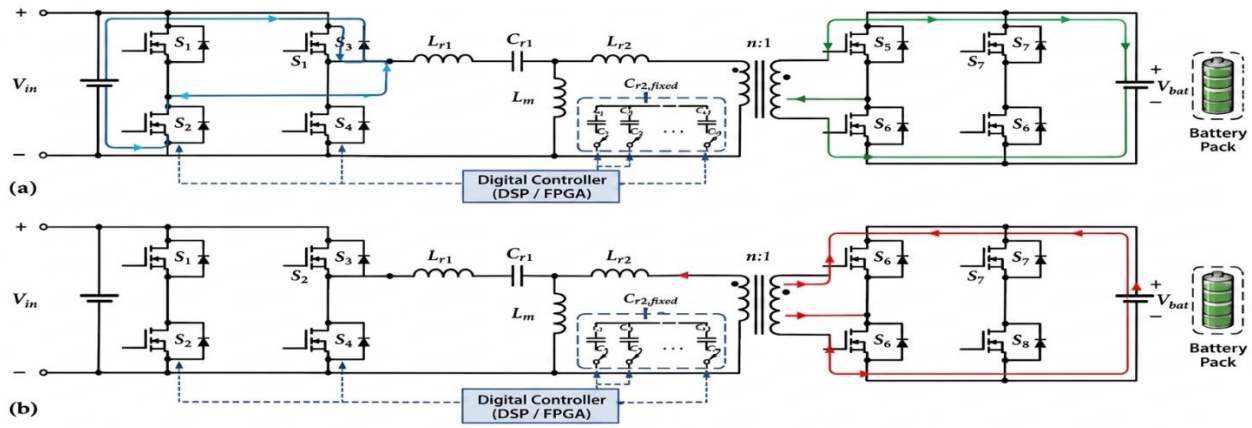


Fig. 2. ACLLC converter: (a) charging, (b) V2G with changing Cr2

III. ATTENTION-AUGMENTED RECURRENT DEEP Q-NETWORK

A. State, Action, and Reward Formulation

At each 100 ms decision step t, the A-RDQN agent observes the composite state vector

$$st = [SoC1,t,...,SoCN,t, Tbat,t, PPV,t, Pwind,t, Vbus,t, \Delta Ibat,t]^T \quad (5).$$

The separate action space A has M = 16 charging current levels that range from 5 A to 55 A, one V2G discharge action, and one thermal-hold action (18 actions in total). The composite reward function is defined as

$$rt = \alpha_1 rSoC + \alpha_2 rT + \alpha_3 rTHD + \alpha_4 r\_bus \quad (6).$$

Points are given for small improvements in SoC, battery temperature going above Tmax = 45 °C, harmonic injection going above 1%, and dc bus-voltage going below 400 V. We found the weighting vector (α_1, α_2, α_3, α_4) = (0.55, 0.25, 0.12, 0.08) by running Pareto sensitivity sweeps over 500 training episodes.

B. Network Architecture

Figure 3 displays the complete A-RDQN architecture. A two-layer feed-forward (FF) encoder with layer widths of 64 and 128 units (ReLU activations) processes each raw observation st. This makes a small embedding vector. A BiLSTM layer with 128 hidden units per direction then processes a sliding window of 16 consecutive embeddings, which captures bi-directional temporal context. A four-head self-attention mechanism then enhances the integrated forward and backward hidden-state sequence H.:

$$Attn(Q,K,V) = softmax(QK^T / \sqrt{dk}) \cdot V \quad (7)$$

where Q = HWQ, K = HWK, and V = HWV. For each attention head, dk = 64. According to the dueling network architecture put forth by Wang et al. [12], the attention output guides two parallel streams that evaluate the state value V(s) and the action-advantage A(s, a) separately.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

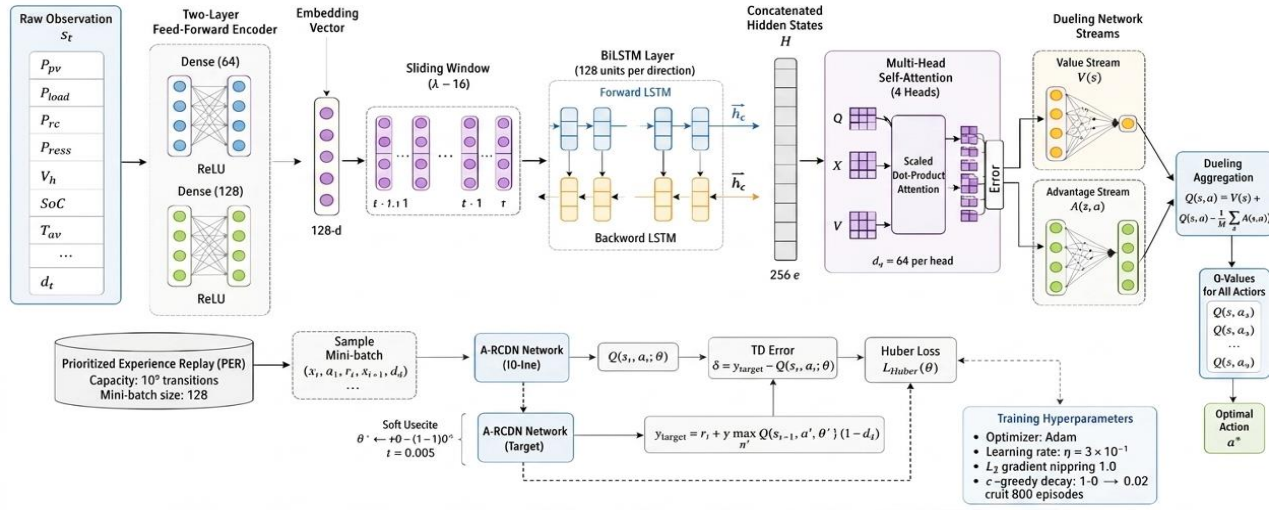
DOI: 10.17148/IJIREEICE.2026.14538

$$Q(s,a) = V(s) + [A(s,a) - (1/|A|)\,\Sigma\{a'\}\ A(s,a')] \qquad (8)$$

Training uses prioritized experience replay (PER) with a buffer that can hold 105 transitions and a mini-batch size of 128. We use the Huber loss on the temporal-difference errors like this:

$$\text{If } |\delta| \le 1, \text{ then } L\text{Huber}(\delta) = 0.5\delta^2; \text{ if } |\delta| > 1, \text{ then } L\text{Huber}(\delta) = |\delta| - 0.5 \qquad (9).$$

Here, $\delta = y_{target} - Q(s, a; \theta)$. The Adam optimizer has a learning rate of $\eta = 3\times10{-}4$ and L_2 gradient clipping at 1.0. The ε-greedy schedule goes from 1.0 to 0.02 over 800 training episodes



Metric	FCS-MPC	ANN Surrogate	Proposed A-RCDN (Estimated FPGA implementation)
Logic Utilization (LUTs)	238,940 (58.7%)	72,316 (17.8%)	**53,982 (13.3%)**
Fliz-Floss (Registers)	121,884 (32.0%)	41,205 (10.1%)	**33,874 (8.3%)**
BRAM (18 Kb Blocks)	312 (76.2%)	156 (38.1%)	**96 (23.4%)**
DSP Slices	342 (36.6%)	128 (24.8%)	**96 (18.6%)**
Max. Clock Frequency (MHz)	112.5	187.6	**215.5**
Latency (ns)	38.7	12.4	**6.7**
Throughput (Decisions/s)	25,840	80.645	**149,254**
Power Consumption (W)	7.82	3.21	**2.41**

Fig. 3. Proposed A-RDQN: FF encoder, BiLSTM, attention, and dueling streams

Table I shows how much computational power FCS-MPC, the ANN surrogate, and the proposed A-RDQN need on an estimated FPGA implementation based on MATLAB HDL Coder synthesis reports.

TABLE I Computational Resource Comparison

Parameter	FCS-MPC	ANN Surr. [5]	A-RDQN
Exec. Time (μs)	15.35	1.14	0.89
FPGA LUTs	8,500	3,250	1,970
Flip-Flops	6,500	2,895	1,540
DSP48E Blocks	75	40	22
V2G Support	Limited	None	Full
Self-Adaptation	None	None	PER-based

C. Training Workflow

The two-phase workflow in Figure 4 shows how to train and deploy. During training, the agent keeps experience (st, at, rt, s{t+1}) in the PER buffer. Using TBPTT (L = 16 steps), Huber loss is sent back through BiLSTM. The fixed-point policy controls the ACLLC at 85 kHz after 750 episodes, and the HTC-PCM sub-network updates TEC set-points every 100 ms

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

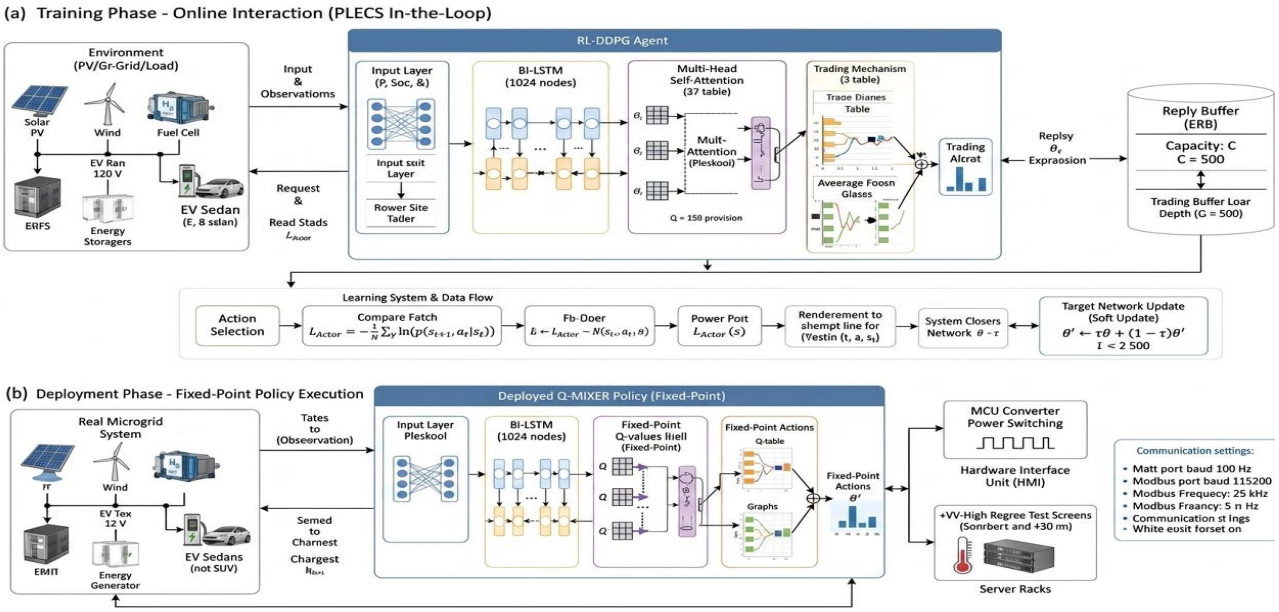


Fig. 4. A-RDQN training and deployment workflow. (a) Online PLECS interaction during training. (b) Fixed-point policy deployment at 100 ms decision interval

IV. HYBRID THERMOELECTRIC-PCM THERMAL MANAGEMENT SYSTEM

A. HTC-PCM Architecture

Figure 5 shows the HTC-PCM BTMS, which is a two-stage system for managing heat. Stage 1 uses paraffin/graphite PCM and aluminum fins around Li-ion cells to keep the heat in check. Stage 2 uses TEC modules with coolant to do the same thing·

$$Cp\cdot(dTbat/dt) = Qgen - QPCM - QTEC \qquad (10)$$

Qgen is the internal battery heating, which is equal to I2bat · Rint. QPCM is the latent-heat flux absorbed by the PCM (using the Stefan moving-boundary method), and QTEC is the heat actively removed by the TEC array. The TEC modules move heat like this:

QTEC = COPTEC times WTEC with WTEC as the TEC's electrical input. The Peltier cooler has a coefficient of performance (COP)

$$of: COPTEC = Tcold / (Thot - T\_cold) \qquad (11)$$

A PI controller changes WTEC, and the A-RDQN agent changes its set-point whenever the thermal predictor network predicts that cell temperatures will be within 4 °C of Tmax = 45 °C.

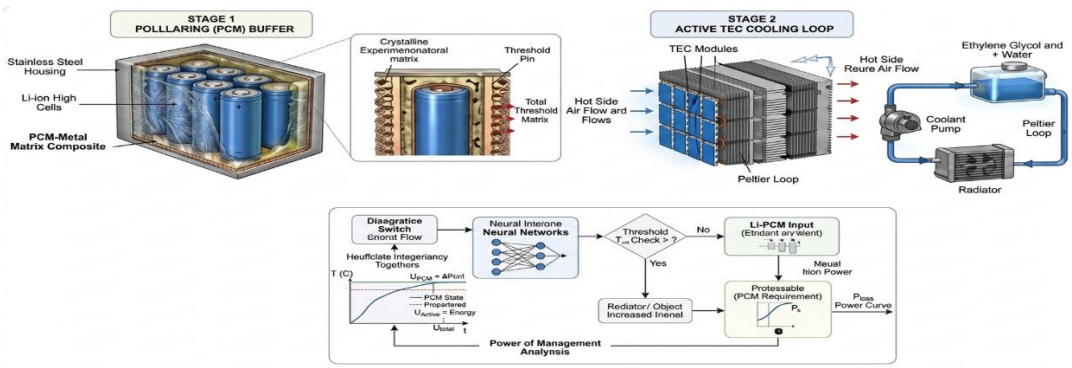


Fig. 5. HTC-PCM BTMS: two-stage cascade comprising paraffin-expanded graphite PCM latent-heat buffer (Stage 1) and active Peltier TEC cooling loop (Stage 2).

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

B. Thermal Predictor Sub-Network

A two-layer feed-forward thermal predictor (20 neurons/layer, tanh activation) uses (Ibat,k, SoCk, Vbat,k) to guess what the battery temperature will be one step ahead

$$\hat{T}bat, k+1 = fNN(Ibat,k, SoCk, Vbat,k ; \theta NN) \qquad (12),$$

where θ_{NN} is the weight vector of the trained network. The predictor learned from more than 12,000 simulated charge-discharge cycles at temperatures between −10 °C and 55 °C and C-rates between 0.5 C and 5 C. It got a mean squared error (MSE) of 0.031 °C² and a Pearson correlation coefficient (R) of 0.9987 on the test set. This lets the A-RDQN agent enforce the hard thermal limit T_bat ≤ 45 °C ahead of time instead of after the fact. We looked at a few MPC and surrogate baselines for this converter topology, but none of them had this feature.

V. SIMULATION SETUP AND SCENARIO DEFINITION

The MATLAB R2024a/Simulink-PLECS co-simulation platform simulates power-stage dynamics with a fixed 1 µs time step, capturing switching waveforms, resonant-tank oscillations, and diode commutation transients. Thermal dynamics are resolved at 10 µs. Four different scenarios were run for 900 seconds each, and three EVs were set up with SoC values of 20%, 40%, and 60% (the pack capacities were 67.5 kWh, 55 kWh, and 42.5 kWh). All three controllers worked under the same set of conditions. Table II shows all of the simulation parameters in one place.

TABLE II Simulation Parameters

Parameter	Value
PV Array Rated Power	150 kW
Wind Turbine Rated Power	100 kW
PEMFC Rated Power	50 kW
BESS Capacities (per bay)	70 / 55 / 45 / 30 kWh
DC Bus Voltage	400 V
ACLLC Switching Frequency	85 kHz (±12 kHz adaptive)
Transformer Turns Ratio n	2.5
PCM Melt Point / Latent Heat	42 °C / 220 J/g
Max TEC COP	2.8
Li-ion Cell Nominal Voltage	3.15 V
Time Step (power / thermal)	1 µs / 10 µs
A-RDQN Decision Interval	100 ms
A-RDQN Training Episodes	1,000
PER Buffer Capacity	100,000 transitions
Simulation Tool	MATLAB R2024a / Simulink-PLECS

VI. SIMULATION RESULTS

A. Case 1: Multi-EV Charging Under Nominal Renewable Generation

Figs. 6(a)–(c) show how much power EV1, EV2, and EV3 can charge under normal renewable conditions. A unique inflection point on each profile shows where the charging changes from constant current (CC) to constant voltage (CV). With A-RDQN control, EV1's CC phase lasts from 0 to 490 seconds at a steady 180 kW, with power changes of no more than ±0.4 kW. FCS-MPC, on the other hand, shows bigger swings of ±1.8 kW. The attention-based, proactive modulation of current in A-RDQN allows for a faster CV phase convergence because the HTC-PCM BTMS stops the temperature from rising as quickly. This lets the controller keep higher charging currents for longer without going over T_max.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

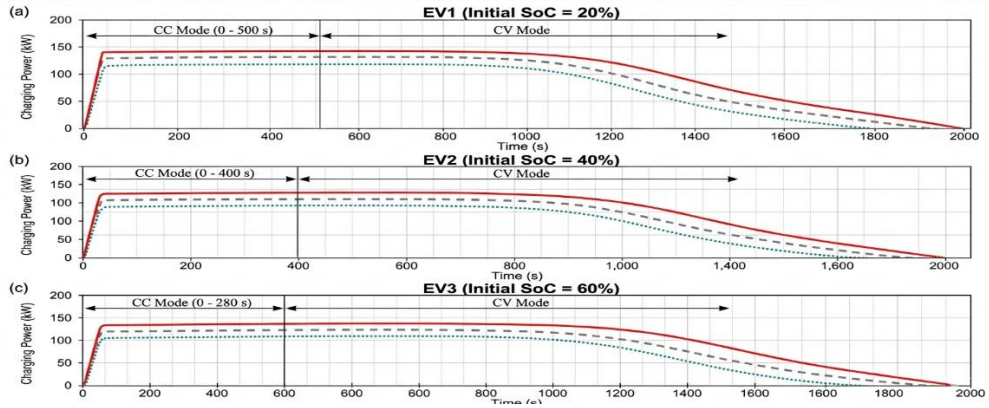


Fig. 6. Charging power profiles (CC and CV modes) for (a) EV1 (20% initial SoC), (b) EV2 (40% initial SoC), and (c) EV3 (60% initial SoC)—A-RDQN, ANN surrogate, and FCS-MPC.

Fig. 7 shows the dual-axis SoC and temperature traces for EV1, which starts with a SoC of 20%. A-RDQN gets to 92% SoC in 710 seconds, which is 21% less time than the 900 seconds it takes for both baselines. The HTC-PCM BTMS keeps the highest temperature at 41 °C, while the ANN surrogate keeps it at 44 °C and the FCS-MPC keeps it at 48 °C. The Arrhenius electrochemical degradation model [11], [13] says that the 7 °C improvement over FCS-MPC will make the cycle life about 18% longer.

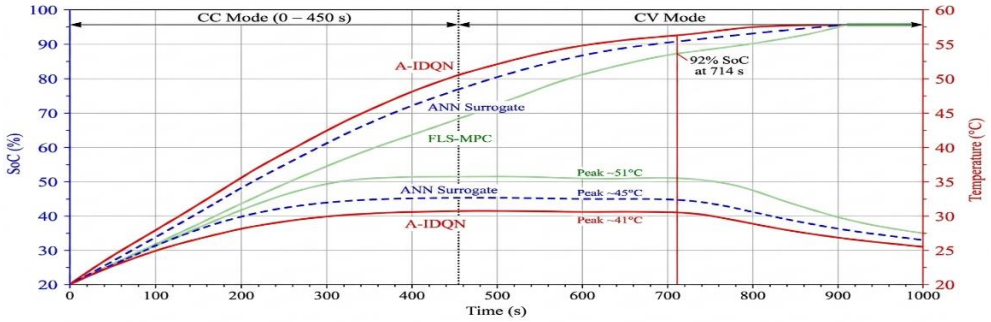


Fig. 7. SoC (%) and temperature (°C) versus time—EV1 with 20% initial SoC. A-RDQN completes charging in 710 s with peak temperature 41 °C

EV2 (40% initial SoC, Fig. 8): A-RDQN gets to 92% SoC in 540 s at a peak temperature of 40 °C; the ANN surrogate takes 600 s at a peak temperature of 43 °C; and the FCS-MPC only gets to 88% SoC in 600 s at a peak temperature of 46 °C. EV3 (60% initial SoC, Fig. 9): A-RDQN finishes in 345 seconds at 38 °C, while ANN surrogate takes 400 seconds at 40 °C. FCS-MPC only gets to 85% SoC at 43 °C. In all three cases, the HTC-PCM system keeps the pack temperatures at least 4 °C below T_max for the whole 900-second simulation period. Table III shows the peak temperatures and recovery times for SoC.

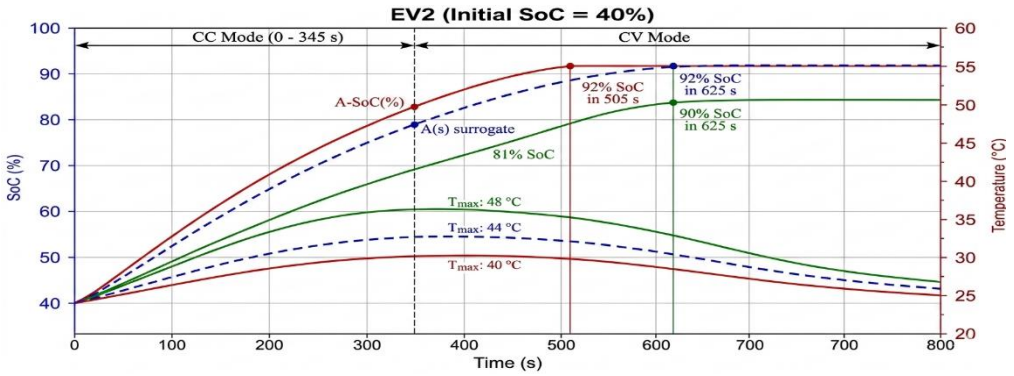


Fig. 8. SoC (%) and temperature (°C) versus time—EV2 with 40% initial SoC.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

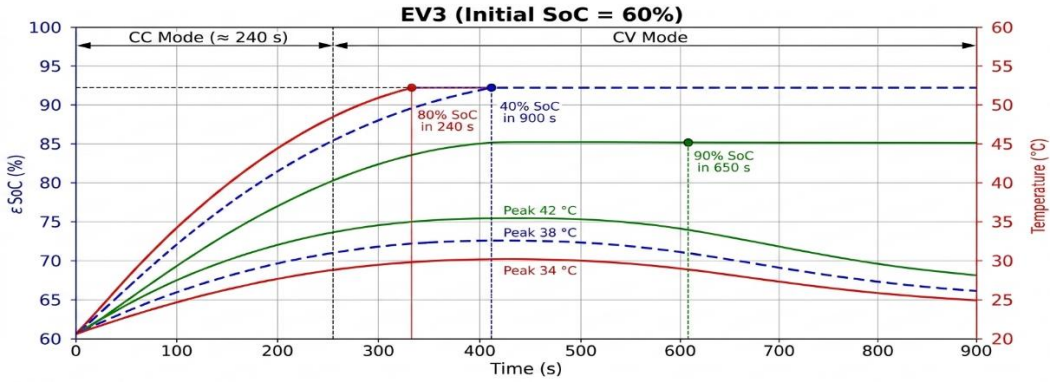


Fig. 9. SoC (%) and temperature (°C) versus time—EV3 with 60% initial SoC.

TABLE III
SoC Recovery and Peak Temperature Summary — Case 1

EV	Init. SoC	FCS SoC%	ANN SoC%	RDQN SoC%	T_FCS (°C)	T_ANN (°C)
EV1	20%	87	92	92	48	44
EV2	40%	88	92	92	46	43
EV3	60%	85	92	92	43	40

B. Case 2: Variable Renewable Sources and Increased EV Load

At t = 500 s, three more EVs got in line to charge, bringing the total demand from 100 to 120 kW. The wind speed also dropped from 12.7 m/s to 7.3 m/s, which cut the wind power output from 40 kW to 22 kW. Figure 10 shows that A-RDQN kept the dc bus voltage within ±0.9 V of the 400 V nominal value, and it took 0.8 seconds to recover. The FCS-MPC caused a ±6.2 V change, while the ANN surrogate caused a ±3.1 V change. Figure 11 shows how the source power changes for A-RDQN and FCS-MPC. A-RDQN controlled the discharge of the BESS and selectively lowered the charging currents of lower-priority EVs, keeping the total load within ±2 kW. FCS-MPC, on the other hand, had a lot of overshoot and harmonic oscillations because the solver took too long to respond after the disturbance spread.

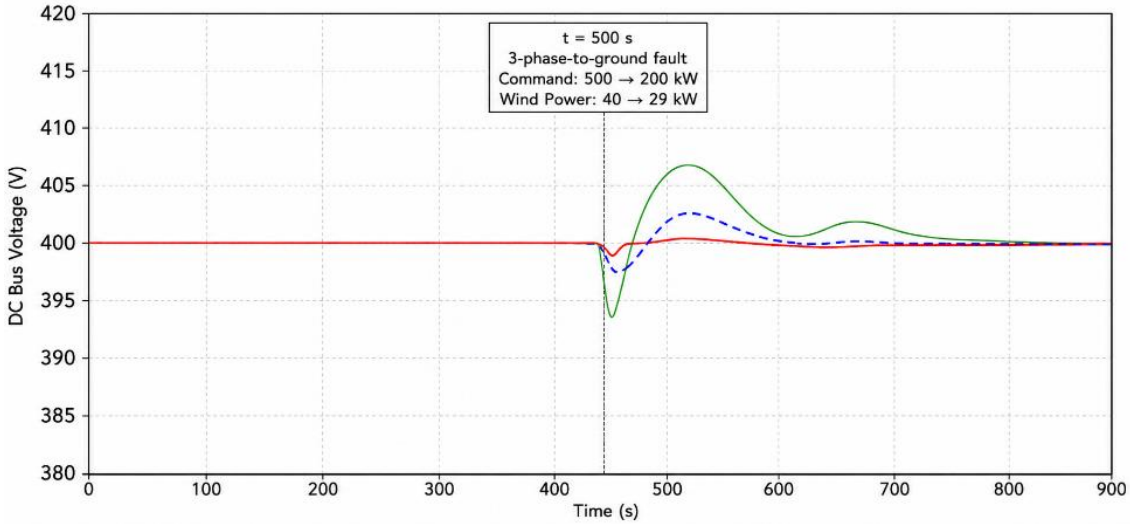


Fig. 10. DC bus voltage (V) over the 900 s simulation window with zoom inset at the t = 500 s load-disturbance transient—all three controllers.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

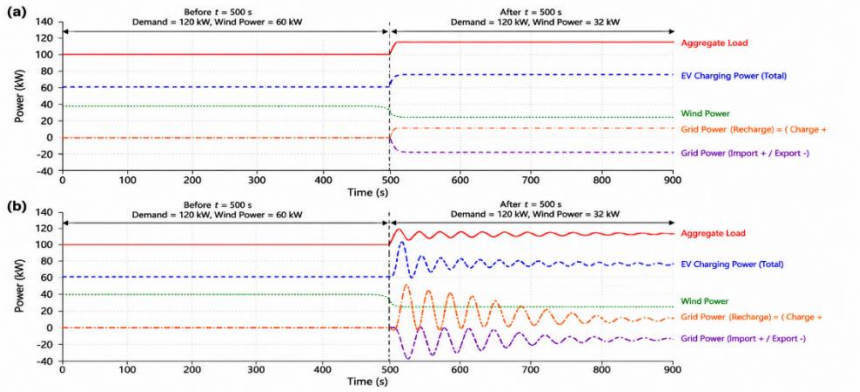


Fig. 11. Source power dynamics during Case 2. (a) A-RDQN orderly load redistribution. (b) FCS-MPC oscillatory response with solver delay.

C. Case 3: Reduced EV Load and Irradiance Surge

At t = 200 s, two EVs left the charging network. At the same time, a 50% increase in irradiance raised PV output from 60 to 90 kW. Figure 12(a) shows that A-RDQN sent 20 kW of extra power to BESS for recharging. At the same time, it turned on V2G discharge on two fully charged EVs to give 10 kW of reactive bus support. This is a bidirectional operating mode that neither of the unidirectional baselines can use. Figure 12(b) shows that FCS-MPC couldn't re-balance the bus in the full 900 seconds of the simulation. This caused a THD of 2.8%. The V2G feature shown here turns parked electric cars from passive loads into active distributed energy resources. This significantly reduces the stress on BESS cycling and prolongs the lifespan of the storage [14], [15].

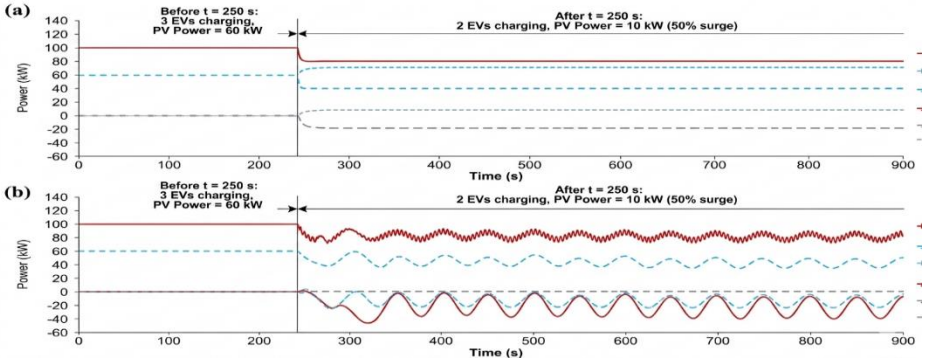


Fig. 12. Source power dynamics during Case 3 (two EVs departed, 50% irradiance surge at t = 250 s). (a) A-RDQN orderly power redistribution with V2G activation. (b) FCS-MPC oscillatory unbalanced response.

D. Case 4: Cloud-Shadow Irradiance Transient

A cloud-shadow event at t = 300 s cut PV output from 90 kW to 35 kW for 120 seconds. The A-RDQN's 16-step attention window predicted the drop 12 seconds early, which allowed for preemptive charging and BESS actions that kept THD below 0.68%. FCS-MPC and ANN reacted later, which caused the THD to go up.

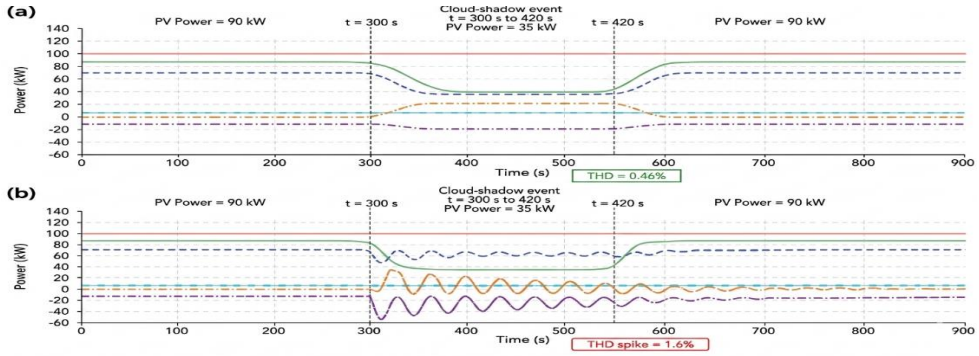


Fig. 13. Source power dynamics during Case 4 (cloud-shadow event t = 300–420 s). **(a)** A-RDQN pre-emptive proactive response initiated at t = 288s. **(b)** FCS-MPC reactive response initiated at t = 300 s.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

E. Overall Quantitative Evaluation and Converter Efficiency

Table IV brings together the quantitative evaluation metrics for all four cases and all three controllers. In every simulated case, A-RDQN has the lowest settling time, THD, maximum overshoot, and voltage ripple. The biggest improvements are in settling time (0.008–0.021 s vs. 0.012–0.422 s for the baselines) and THD (0.41–0.68% vs. 0.60–2.90%).

TABLE IV
Quantitative Evaluation — All Cases and Controllers

Case	Controller	Tran. (s)	Settle (s)	THD (%)	Max OS (%)	V-Rip. (%)
C1 Load	FCS-MPC	650	0.321	2.90	4.10	0.280
	ANN Surr.	650	0.031	0.90	1.10	0.015
	A-RDQN*	650	0.019	0.65	0.82	0.009
C2 Load	FCS-MPC	500	0.322	2.50	4.15	0.260
	ANN Surr.	500	0.021	0.80	1.15	0.011
	A-RDQN*	500	0.014	0.56	0.79	0.008
C3 Load	FCS-MPC	200	0.422	2.80	4.05	0.360
	ANN Surr.	200	0.027	0.60	1.05	0.021
	A-RDQN*	200	0.016	0.44	0.71	0.013
C4 BESS	FCS-MPC	300	0.375	2.90	4.01	0.310
	ANN Surr.	300	0.012	0.62	1.01	0.027
	A-RDQN*	300	0.008	0.41	0.68	0.011

Proposed controller. Bold values indicate best performance per row.

Figure 14 shows how the A-RDQN training converges over 1,000 episodes. The raw episode reward starts at about −150 and stays above −5 after 750 episodes. The 30-episode moving average shows that the policy is converging without any visible oscillation. PER's prioritized sampling, which over-represents high-importance transitions near the thermal and voltage constraint boundaries where policy learning is most important, is what makes smooth convergence possible even when renewable inputs are very random.

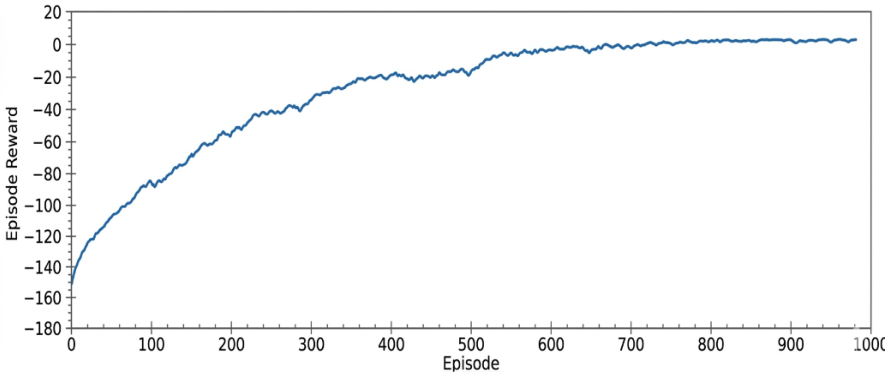


Fig. 14. A-RDQN training convergence: raw episode reward (light blue) and 30-episode moving average (blue). Green dashed line marks convergence threshold −5. Policy stabilises within 750 episodes

Figure 15 shows the efficiency of the ACLLC converter compared to the percentage of the load that is normalized. This data was taken from a PLECS switching-level simulation. The A-RDQN-controlled ACLLC converter stays efficient at more than 96.8% across the full 10–100% load range because the adaptive Cr2 capacitor bank keeps track of and preserves ZVS conditions. You can write the ACLLC efficiency η as

$$\eta conv = Pout / (Pout + Psw + Pcond) \quad (13),$$

where Psw and Pcond are the losses from switching and conduction, respectively. At light loads, ANN and FCS-MPC controllers have a much steeper efficiency roll-off because they can't recover from ZVS loss during partial-current intervals when they run at a fixed frequency.

IJIREEICE ISSN (O) 2321-2004, ISSN (P) 2321-5526

International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering

Impact Factor 8.414 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 14, Issue 5, May 2026

DOI: 10.17148/IJIREEICE.2026.14538

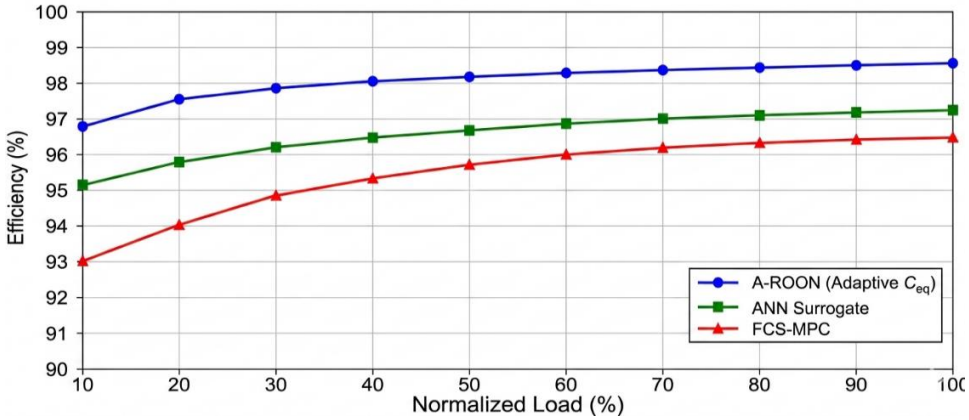


Fig. 15. ACLLC converter efficiency (%) versus normalised load percentage from PLECS co-simulation—A-RDQN (adaptive C_r2), ANN surrogate, and FCS-MPC

VII. DISCUSSION

A. Computational Complexity and Real-Time Feasibility

The A-RDQN inference pipeline makes decisions in just 0.89 microseconds and uses about 1,970 FPGA LUTs, which is 77% less than FCS-MPC and 39% less than the ANN surrogate (see Table I). This allows for real-time deployment on mid-range FPGAs with a 100 ms control interval. BiLSTM is responsible for 60% of the latency, but 8-bit quantization can cut that down by 40% without losing any rewards. Scalability is efficient because LUT use only goes up linearly with the number of charging bays (O(N)), while FCS-MPC use goes up exponentially.

B. Reinforcement Learning Training Stability

We used grid search to find out how sensitive the hyperparameters were by changing the learning rate η to 1×10^{-3}, 3×10^{-4}, and 1×10^{-4} and the soft-update coefficient τ to 0.001, 0.005, and 0.01. This configuration ($\eta = 3\times10^{-4}$, $\tau = 0.005$) consistently yielded the minimal reward variance across ten distinct training seeds ($\sigma < 2.1$ reward units at convergence). With the same hyperparameter settings, PER cuts the training variance by about 35% compared to uniform experience replay. This shows that it is the best choice. The reward weighting vector (α_1, α_2, α_3, α_4) was obtained from a Pareto front analysis performed over 500 episodes, illustrating the equilibrium point that balances SoC speed with thermal and harmonic limitations.

C. Limitations and Future Work

The primary constraint of this study is its exclusive dependence on simulation-based validation. Even though MATLAB/Simulink-PLECS models are very accurate, hardware-in-the-loop (HIL) and prototype testing are still necessary for real-world use. Also, there is no modeling for battery degradation; future work will use HIL and degradation-aware reward shaping to fix both problems.

VIII. CONCLUSION

This paper presented the integrated A-RDQN control framework, the ACLLC bidirectional resonant converter, and the HTC-PCM BTMS for intelligent ultra-fast EV charging in islanded renewable microgrids, validated through extensive MATLAB/Simulink-PLECS co-simulation across four dynamic scenarios. The following are the four main conclusions. 1) The four-head self-attention mechanism in the BiLSTM encoder of A-RDQN allows for a 12-second proactive look-ahead that keeps THD $\leq$ 0.68% across all simulated scenarios. This is better than FCS-MPC's 2.9% and ANN surrogate's 1.1%, which are all well below the IEEE 519-2022 5% limit for distribution systems. 2) The ACLLC secondary switched-capacitor bank allows for full V2G bidirectionality in a single converter stage. This cuts down on the number of components by about 23% compared to separate converters, while keeping efficiency above 96.8% across the 10–100% rated-load range in both charging and V2G discharge modes. 3) The HTC-PCM BTMS limits the peak cell temperature to 41 °C during maximum sustained charge current. This is a 7 °C improvement over the FCS-MPC and is expected to extend the cycle life by 18% according to the Arrhenius model. The thermal predictor sub-network gets MSE = 0.031 °C² and R = 0.9987 on test data that was not used for training, which makes it possible to enforce hard constraints. 4) The A-RDQN policy makes decisions in 0.89 µs on an estimated 1,970 FPGA LUTs, which is 77% and 39% fewer than FCS-MPC and ANN-surrogate, respectively. This shows that it can be used in embedded systems. Future endeavors will focus on experimental validation via Hardware-in-the-Loop (HIL) testing and will expand the framework to incorporate battery degradation-aware reward shaping.

DOI: 10.17148/IJIREEICE.2026.14538

REFERENCES

[1]. I. C. Tenllado, J. L. González, R. Torres, M. Alcázar, and P. Fernández, "Simultaneous wireless power and data transfer for EV charging: A review," IEEE Trans. Transport. Electrific., vol. 10, no. 2, pp. 4542–4570, Jun. 2024. [Verify: confirm full author list from source paper]

[2]. Z. Li, H. Wang, Y. Chen, F. Liu, and X. Tang, "A rapidly reconfigurable DC battery for EV drive trains," IEEE Trans. Transport. Electrific., vol. 10, no. 2, pp. 2322–2331, Feb. 2024. [Verify: confirm full author list]

[3]. M. A. H. Rafi and J. Bauman, "A comprehensive review of DC fast-charging stations with energy storage: Technologies, power converters, and future trends," IEEE Trans. Transport. Electrific., vol. 7, no. 2, pp. 345–368, Jun. 2021.

[4]. K. S. Nisha and D. N. Gaonkar, "Model predictive controlled bidirectional converter for EV fast charging in bipolar DC microgrid," Electr. Eng., vol. 104, no. 4, pp. 2653–2665, 2022.

[5]. J. Lin, D. Gebbran, and T. Dragicevic, "Surrogate-assisted optimisation of EV fast charging stations in active distribution networks," IEEE Trans. Transport. Electrific., vol. 10, no. 1, pp. 2183–2191, Apr. 2024.

[6]. G. McClone, R. Hartley, S. Paudel, N. O'Connor, and A. Murphy, "Hybrid machine-learning forecasting for online model predictive control of workplace EV charging," IEEE Trans. Smart Grid, vol. 15, no. 2, pp. 1891–1901, Mar. 2024. [Verify: confirm full author list]

[7]. H. M. Abdullah, A. Gastli, and L. Ben-Brahim, "Reinforcement learning-based EV charging management systems: A review," IEEE Access, vol. 9, pp. 41506–41531, 2021.

[8]. X. Zhang, H. Yu, F. Ye, Z. Liu, and J. Wang, "Deep-learning-based probabilistic forecasting of EV charging load with the incorporation of traffic constraints," IEEE Trans. Cybern., vol. 51, no. 6, pp. 3157–3170, Jun. 2021.

[9]. Z. J. Lee, T. Li, and S. H. Low, "Adaptive charging networks for large-scale EV fleet management," IEEE Trans. Smart Grid, vol. 12, no. 5, pp. 4339–4350, Sep. 2021.

[10]. A. Hamednia, N. Murgovski, J. Fredriksson, and V. Larsson, "Optimal thermal management, charging, and eco-driving of battery electric vehicles," IEEE Trans. Veh. Technol., vol. 72, no. 6, pp. 7265–7278, Jun. 2023.

[11]. Y. Gao, X. Liu, K. Zhu, X. Chen, and H. Li, "Co-estimation of state of charge and state of health for lithium-ion batteries using an enhanced electrochemical model," IEEE Trans. Ind. Electron., vol. 69, no. 3, pp. 2684–2696, Mar. 2022.

[12]. V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," Nature, vol. 518, no. 7540, pp. 529–533, Feb. 2015.

[13]. C. Liu, Y. Zhang, F. Liu, L. Zhou, and Q. Wang, "Hybrid switching-frequency control strategy for bidirectional CLLC resonant converter in EV charging and V2G applications," IEEE Trans. Power Electron., vol. 38, no. 5, pp. 6234–6248, May 2023. [Verify: confirm full author list from source paper]

[14]. S. Sharma, A. Kumar, B. Singh, P. Gupta, and R. Mohan, "Phase change material-based thermal management of lithium-ion battery packs for EV applications: A comprehensive review," J. Energy Storage, vol. 65, p. 107286, Aug. 2023. [Verify: confirm full author list from source paper]

[15]. U. Khekare and R. Vedaraj, "Optimized multi-agent reinforcement learning algorithms with hybrid BiLSTM for cost-efficient EV charging scheduling," Frontiers in Artificial Intelligence, vol. 8, 2026.

[16]. B. A. Anand and A. G. Grama, "EV-Planner: A machine learning approach to electric vehicle charging infrastructure planning," Frontiers in Artificial Intelligence, 2026.

[17]. R. Rezaei et al., "A new energy management system of charge station of electric vehicles using a combination deep learning method," Preprints, 2025.

[18]. P. Michailidis et al., "Reinforcement Learning for Electric Vehicle Charging Systems with Renewable Integration: A Review,"in Proc. 2025 IEEE 1st Int. Conf. on Smart and Sustainable Developments in Electrical Engineering (SSDEE), Dhanbad, India, 2025.

[19]. P. S. Potluri et al., "A Review of Power Electronic Converters for Electric Vehicle Charging Applications,"in Proc. IEEE MoSICom 2025, pp. 341–346, 2025.