Adaptive Control of Wireless Power Transfer for Electric Vehicle Charging under Variable Coupling Conditions Using Proximal Policy Optimization

DuVan Nguyen

Faculty of Mechanical Engineering, Thu Duc College of Technology, Ho Chi Minh City, Vietnam

DOI: https://doi.org/10.68050/JAMS.2026.448

Received: 10 Sept 2026; Accepted: 11 Sept 2026; Published: 14 September 2026

ABSTRACT

Wireless power transfer (WPT) offers a convenient and flexible solution for electric vehicle (EV) charging, but its performance is highly sensitive to variations in coil separation and magnetic coupling. This paper proposes a Proximal Policy Optimization (PPO) adaptive control strategy for resonant WPT EV charging under variable coupling conditions. A series-series compensated WPT system is modeled, and a PPO controller is developed to regulate the charging current and operating frequency according to the instantaneous coupling coefficient, transfer efficiency, battery voltage, and charging current. The proposed controller is trained and evaluated in MATLAB/Simulink under coil separation distances ranging from 5 to 20 cm. The PPO agent begins to converge at approximately 1200 training episodes and reaches a stable learning regime after around 1600 episodes. Compared with conventional fixed-frequency control, the proposed strategy improves the WPT transfer efficiency from 89.3% to 91.0% at a 5 cm coil separation and from 68.2% to 73.9% at 20 cm. The improvement becomes more pronounced as the coupling weakens. The dynamic-response evaluation gives a settling time of approximately 0.48 s, with a maximum deviation of 5.2% and a steady-state error below 2.8%. These results indicate that PPO-based adaptive control can improve both transfer efficiency and dynamic adaptability of wireless EV charging under variable coupling conditions. Overall, the results demonstrate that PPO-based adaptive control can effectively improve efficiency and dynamic regulation of wireless EV charging systems, particularly when variations in coil separation lead to unfavorable coupling conditions. The proposed approach provides a promising foundation for intelligent WPT controllers capable of maintaining reliable charging performance under practical operating variations.

Keywords: Wireless power transfer, Electric vehicle charging, Proximal Policy Optimization, Reinforcement learning, Adaptive control, Coupling coefficient, Power-transfer efficiency.

INTRODUCTION

The rapid adoption of electric vehicles (EVs) is increasing the demand for charging technologies that are efficient, reliable, and convenient for users. Conventional conductive charging systems require physical electrical connections between the vehicle and the charging infrastructure, which can limit user convenience and introduce concerns associated with connector wear, environmental exposure, and automated charging. Wireless power transfer (WPT) offers an alternative approach in which electrical energy is delivered through magnetic coupling between transmitting and receiving coils. This characteristic makes WPT particularly attractive for automated EV charging and future intelligent transportation infrastructure (S. A. Sagar et al. 2023), (Xue et al. 2025).

Despite its advantages, the practical performance of WPT-based EV charging is strongly influenced by the electromagnetic coupling between the two coils. In particular, changes in the separation distance or misalignment between the transmitting and receiving coils modify the mutual inductance and coupling coefficient. As a consequence, the transferred power and overall system efficiency can deteriorate as the vehicle position deviates from the nominal charging condition (Z. Zhao et al. 2023), (S. Zhao et al. 2024). Such variations are difficult to avoid in practical applications because the vehicle may not be parked at exactly the same position during every charging event.

Several approaches have therefore been investigated to improve the robustness and efficiency of WPT systems. Existing studies have considered coil optimization, resonant compensation, system modeling, parameter identification, and advanced control strategies (S. A. Benfadhel et al. 2025), (Chen et al. 2023), (Feng et al. 2024). These approaches have contributed significantly to improving WPT performance. Nevertheless, many conventional controllers rely on predetermined operating parameters or offline tuning. Their effectiveness can consequently deteriorate when the coupling condition changes significantly during operation.

The development of intelligent control techniques has created new opportunities for addressing this limitation. Reinforcement learning (RL) is particularly attractive because a controller can learn an operating policy through interactions with the system rather than relying entirely on an explicitly optimized control law. Recent studies have demonstrated the applicability of RL and deep reinforcement learning (DRL) to EV charging, energy management, and charging coordination problems (Xu et al. 2024), (Jamjuntr and Suanpang 2024), (Park and Moon 2022). Among the available policy-based algorithms, Proximal Policy Optimization (PPO) has attracted considerable attention because of its relatively stable policy updates and suitability for continuous-control problems.

Although adaptive WPT control and reinforcement learning-based EV charging have both received increasing attention, these research directions have largely developed along separate paths. Existing WPT studies mainly address parameter variation through analytical models, system identification, or model-based control, whereas RL-based EV studies predominantly focus on charging scheduling, energy management, and grid-level coordination (S. Zhang et al. 2023), (Z. Zhang, Liu, and Sun 2025). Consequently, the use of continuous-control RL to directly regulate the electromagnetic operating variables of a WPT system under changing coil coupling remains insufficiently explored. In particular, there is limited evidence on whether a learned policy can simultaneously regulate charging current and operating frequency while maintaining transfer efficiency as coil separation increases. This study addresses this gap by developing and evaluating a PPO-based controller directly at the WPT control layer.

The main contributions of this study are summarized as follows:

i) A series–series compensated WPT model is developed to represent the electromagnetic power-transfer process for wireless EV charging under variable coil coupling conditions.

ii) A PPO-based adaptive control strategy is developed to jointly regulate the charging current and operating frequency according to the instantaneous WPT and battery states.

iii) The proposed controller is evaluated under progressively weaker coupling conditions by varying the coil separation from 5 to 20 cm, allowing its ability to mitigate efficiency degradation to be systematically examined.

iv) The proposed strategy is benchmarked against conventional fixed-frequency control in terms of learning convergence, transfer efficiency, and transient response.

SYSTEM MODEL AND PPO-BASED ADAPTIVE CONTROL

The wireless charging system considered in this study is based on resonant Wireless Power Transfer (WPT), in which electrical energy is delivered to the EV battery through magnetic coupling rather than a direct conductive connection. As illustrated in Figure 1, the system consists of an AC power source, an AC/DC rectifier, a high-frequency inverter, a primary resonant compensation network, a transmitting coil, a receiving coil, a secondary compensation network, a receiving-side rectifier, and the EV battery. This configuration provides the energy conversion and magnetic coupling stages required to transfer power from the grid-side source to the vehicle without a physical electrical connector.

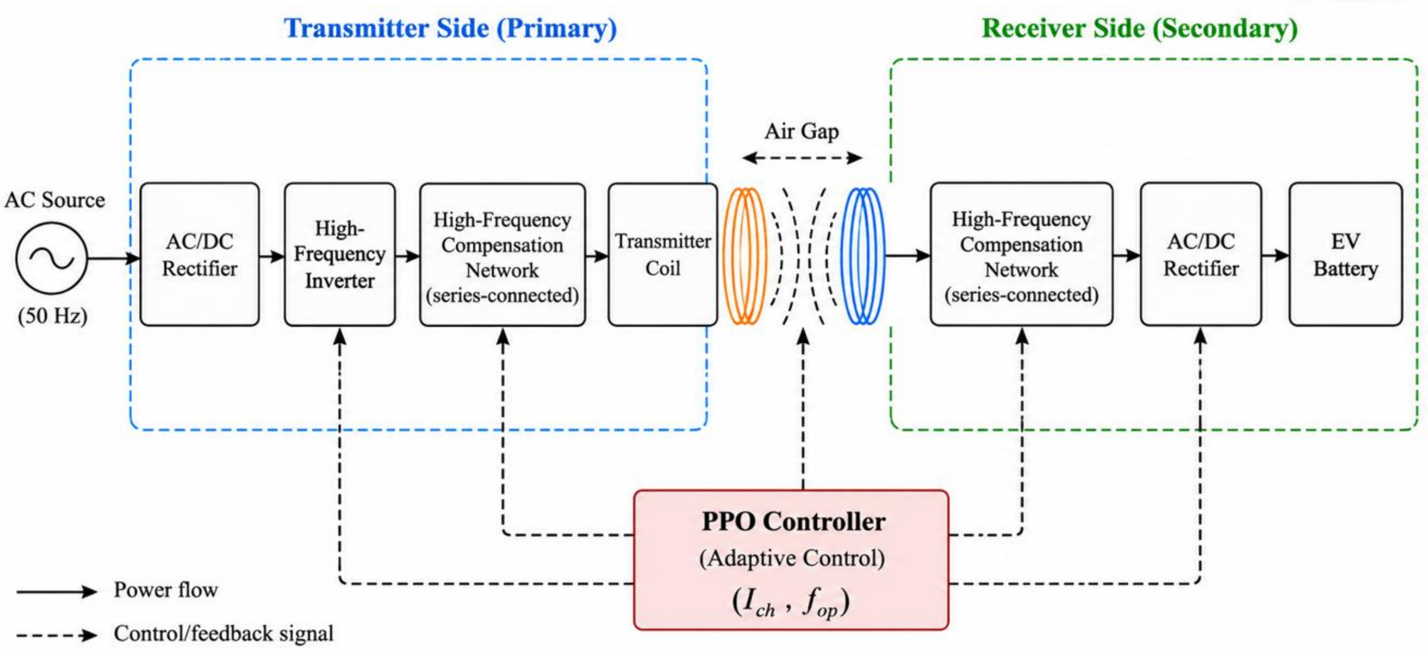

Figure 1. PPO Adaptive Control Architecture for Wireless EV Charging

A series-series (SS) compensation topology is adopted in the WPT system. This topology is selected because it provides a suitable resonant operating condition and helps reduce the influence of reactive power on the power-transfer process. It is also widely considered for resonant WPT applications because of its favorable power-transfer characteristics under varying operating conditions (S. A. Sagar et al. 2023), (Xue et al. 2025).

The resonant frequency of the compensation network is determined by the equivalent inductance and capacitance and is expressed as

$$f_r = \frac{1}{2\pi\sqrt{L_{eq}C_{eq}}} \tag{1}$$

where f_r denotes the resonant frequency (Hz), L_{eq} is the equivalent inductance (H), and C_{eq} represents the equivalent capacitance (F).

The resonant condition is important because the operating point of the WPT system directly affects the ability of the coupled coils to transfer energy efficiently. However, resonance alone does not guarantee high efficiency when the relative position of the two coils changes. The electromagnetic coupling between the coils must also be considered.

The coupling strength between the transmitting and receiving coils is characterized by the coupling coefficient:

$$k = \frac{M}{\sqrt{L_pL_s}} \tag{2}$$

where k is the coupling coefficient, M is the mutual inductance between the transmitting and receiving coils (H), and L_p and L_s are the primary and secondary coil inductances (H), respectively.

For the series-series compensated WPT system, the primary and secondary circuits are magnetically coupled through the mutual inductance M. The electrical behavior of the coupled system depends on the coil self-inductances, winding resistances, compensation capacitors, operating frequency, mutual inductance, and equivalent battery-side load. Therefore, variations in coil separation are represented in the model through the corresponding change in M, and consequently in the coupling coefficient k. The resulting changes in primary and secondary currents determine the transferred power and WPT efficiency.

The coupling coefficient provides an important physical link between coil positioning and WPT performance. When the separation distance between the two coils increases, or when their relative positions become misaligned, the mutual inductance generally decreases. Consequently, the effective coupling becomes weaker and the amount of power that can be transferred through the magnetic field is reduced. This behavior has been investigated in recent WPT studies under variations in mutual inductance and operating conditions (Z. Zhao et al. 2023), (S. Zhao et al. 2024), (Chen et al. 2023)

The overall power-transfer efficiency is defined as

$\eta = \frac{P_{out}}{P_{in}} . 100\%$	(3)

where η is the WPT transfer efficiency (%), P_{in} is the input power at the transmitting side (W), and P_{out} is the power delivered to the receiving side (W).

Equations (2) and (3) highlight the central control challenge considered in this study. A change in coil separation modifies the mutual inductance and coupling coefficient, which subsequently changes the power-transfer characteristics of the WPT system. Therefore, a controller designed for only one nominal coupling condition may not remain optimal when the vehicle position changes during charging. This issue is particularly relevant to practical EV wireless charging because exact alignment between the transmitting and receiving coils cannot always be guaranteed.

For this reason, the control problem is formulated as an adaptive decision-making problem rather than a fixed-parameter regulation problem. Instead of maintaining a predetermined operating frequency and charging current, the proposed controller continuously observes the WPT operating condition and modifies these control variables according to the instantaneous system state. The objective is consequently not only to obtain high efficiency at a nominal operating point, but also to preserve satisfactory transfer performance when the magnetic coupling changes.

To achieve this objective, a Proximal Policy Optimization (PPO)-based adaptive controller is introduced in the next subsection.

PPO Adaptive Control Strategy

The proposed control framework employs Proximal Policy Optimization (PPO) as the adaptive decision-making mechanism. The motivation for using PPO is related to the nature of the WPT control problem: both the electromagnetic operating condition and the charging process can vary continuously, while the appropriate control action depends on the instantaneous coupling and battery states. PPO provides an Actor–Critic reinforcement learning framework that can learn such a nonlinear control policy through repeated interaction with the WPT environment.

In contrast to a conventional controller with parameters determined in advance, the PPO agent does not rely on a single fixed operating point. Instead, it learns how the control actions should change in response to variations in the observed system state. This feature is particularly useful when the coil separation changes or when the electrical condition of the EV battery varies during charging. The use of PPO for continuous control is also consistent with recent developments in reinforcement learning for EV charging and energy management (Xue et al. 2025), (Z. Zhao et al. 2023).

At each control instant t, the state vector is defined from variables that describe both the wireless power-transfer condition and the battery operating condition:

$S_t = [k_t, \eta_t, V_{bat,t}, I_{bat,t}]$	(4)

where k_t is the coupling coefficient; η_t is the power-transfer efficiency (%); $V_{bat,t}$ is the battery voltage (V); and $I_{bat,t}$ is the battery current (A).

This state representation is important because it allows the agent to distinguish between different operating conditions that may require different control actions. In particular, the coupling coefficient provides information about the electromagnetic transfer condition, while the transfer efficiency reflects the immediate effectiveness of the energy-transfer process. Battery voltage and charging current, meanwhile, provide information about the electrical charging state. By combining these variables, the agent can make control decisions based on both WPT performance and charging requirements rather than optimizing only one quantity.

Based on the observed state, the PPO agent generates the control action

$A_t = [I_{ch,t}, f_{op,t}]$	(5)

The action vector is defined as A_t where $I_{ch,t}$ is the charging current and $f_{op,t}$ is the operating frequency of the resonant converter.

To ensure physically feasible control commands, the PPO actions are constrained within predefined operating limits. The charging current is bounded by the allowable battery-charging range, while the operating frequency is restricted to a feasible range around the nominal resonant frequency. These constraints prevent the learned policy from generating control commands beyond the operating capability of the WPT converter and battery.

The simultaneous control of these two variables is a key feature of the proposed strategy. A change in coil coupling modifies the electromagnetic operating condition, and therefore maintaining the same control settings may lead to a deterioration in transfer efficiency. By allowing the agent to adjust both the charging current and operating frequency, the controller has greater flexibility to respond to changes in the WPT environment.

The control objective is formulated through a reward function that seeks to improve transfer efficiency while maintaining a stable charging response:

$R_t = \omega_1 . \eta_t - \omega_2 . \lvert I_{ref} - I_{ch,t} \rvert - \omega_3 . \lvert f_{ref} - f_{op,t} \rvert$	(6)

where the reward at time *t* reflects the WPT efficiency together with deviations in the charging current and operating frequency. I_{ref} denotes the reference charging current, f_{ref} represents the nominal resonant frequency, and the weighting coefficients determine the relative importance of the individual reward components.

In the implemented PPO environment, the reward weights are fixed throughout the training and testing processes. The selected values are determined during preliminary simulations and are subsequently kept unchanged during the final evaluation to ensure a consistent comparison between the proposed and conventional controllers. The values of $\omega_1, \omega_2, \omega_3$, together with the reference charging current and nominal operating frequency, should be explicitly reported in the simulation-parameter table. The reward weights were selected as ω_1=0.80, ω_2=0.15, and ω_3=0.05, assigning the highest priority to WPT transfer efficiency while maintaining appropriate charging-current tracking and limiting excessive frequency variation. The selected weights were determined through preliminary simulation trials and were kept fixed throughout the final training and evaluation processes.

The formulation is consistent with the general use of performance-related rewards and control-error penalties in reinforcement-learning-based EV charging strategies reported in (S. Zhang et al. 2023), (Z. Zhang, Liu, and

Sun 2025). However, in the present work, the reward is applied directly to the WPT control problem, where the principal disturbance is associated with changes in the magnetic coupling condition.

PPO Policy Optimization

PPO employs an Actor–Critic architecture to learn the control policy. The Actor generates the control actions, while the Critic estimates the value of the current system state. The advantage function is used to determine whether the selected action produces a better outcome than the expected value associated with the current state:

$A_t = R_t + \gamma V(S_{t+1}) - V(S_t)$	(7)

where A_t denotes the advantage function, R_t is the instantaneous reward, $V(S_t)$ and $V(S_{t+1})$ represent the value estimates of the current and subsequent states, respectively, and γ is the discount factor.

A major difficulty in policy-gradient reinforcement learning is that excessively large policy updates can destabilize training. PPO addresses this issue through a clipped surrogate objective:

$L(\theta) = E[min(r_t(\theta)A_t, clip(r_t(\theta), 1-\varepsilon, 1+\varepsilon)A_t)]$	(8)

where the policy probability ratio is defined by

$r_t(\theta) = \dfrac{\pi\theta(a_t \mid s_t)}{\pi\theta_{old}(a_t \mid s_t)}$	(9)

where $L(\theta)$ is the PPO clipped objective function; $r_t(\theta)$ is the probability ratio between the new and old policies; $\pi\theta$ denotes the current policy; $\pi\theta_{old}$ denotes the previous policy; and ε is the clipping coefficient.

The clipping mechanism is particularly relevant to the present application because the WPT system operates under continuously changing conditions. If the learned policy were allowed to change excessively during a single update, the resulting control actions could vary significantly and lead to unstable training. By restricting the policy update to a controlled region, PPO provides a more conservative learning process and reduces the possibility of abrupt changes in the learned control behavior. This property has contributed to the application of PPO and related reinforcement-learning methods in energy-control problems (Xu et al. 2024), (Hao and Xu 2024).

Both the Actor and Critic are implemented as fully connected feed-forward neural networks. The same hidden-layer dimension is used for the two networks, while the Actor produces the continuous control actions and the Critic estimates the state value. The input and output variables are normalized before being processed by the networks, and the generated actions are subsequently mapped to their physical operating ranges.

PPO Training Procedure

The PPO training process consists of repeated interactions between the agent and the WPT simulation environment. At each control step, the agent receives the current system state, generates charging-current and operating-frequency commands, and applies these actions to the WPT model. The resulting state transition and reward are stored and subsequently used to estimate the advantage and update the Actor and Critic networks. The selected action is applied to the WPT environment, which subsequently returns a new state and a reward determined by the control objective. The collected experience is used to calculate the advantage and update both the Actor and Critic networks. The complete training process is formulated as an iterative interaction between the PPO agent and the WPT environment. At each time step, the agent receives the current system state and generates a control action. The WPT environment then executes this action and returns a new state together with the corresponding reward. The collected experience is subsequently used to update the Actor and Critic networks.

The training procedure can be summarized as follows:

Step 1: Initialize the Actor and Critic networks.

Step 2: Measure or obtain the current state of the WPT system.

Step 3: Use the Actor network to generate the charging-current and operating-frequency commands.

Step 4: Apply the control action to the WPT environment and obtain the subsequent state and reward.

Step 5: Calculate the advantage estimate and PPO objective.

Step 6: Update the Actor and Critic network parameters.

Step 7: Repeat the interaction and policy-update process until the predefined convergence criterion is satisfied or the maximum number of training episodes is reached.

Step 8: Evaluate the trained policy in terms of reward convergence, WPT efficiency, and dynamic response.

This iterative structure enables the controller to progressively learn the relationship between changes in magnetic coupling and the control actions required to preserve charging performance. The corresponding PPO learning flow is presented in Figure 2.

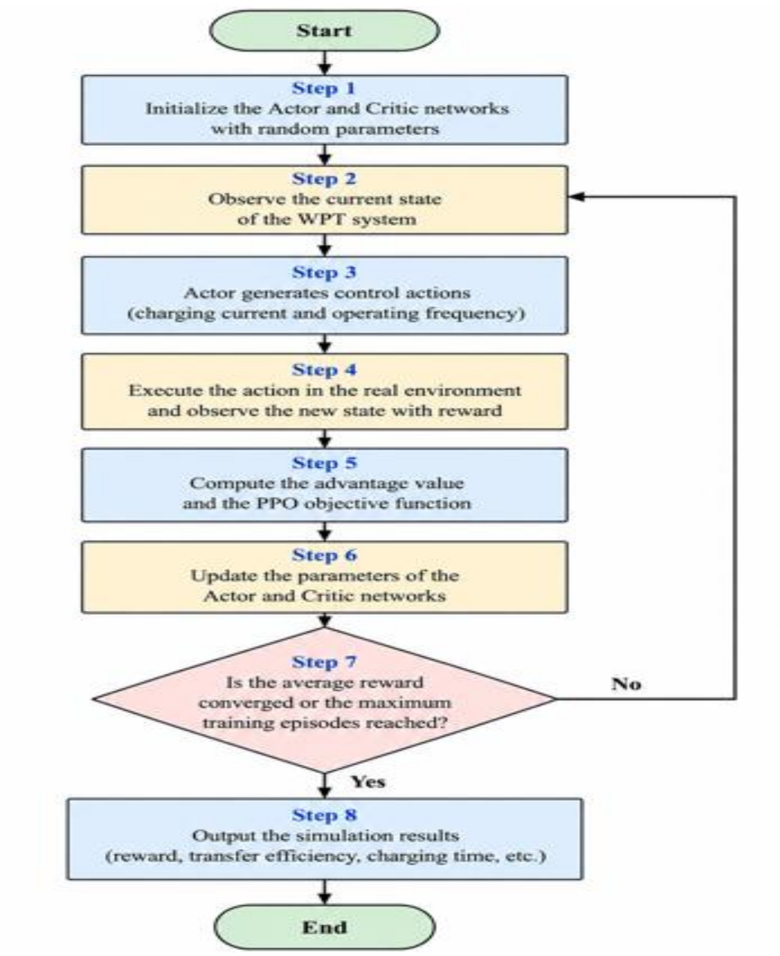

Figure 2. Flowchart of the PPO Algorithm

This closed-loop learning mechanism allows the controller to progressively learn how the operating variables should be coordinated under different coupling conditions. Consequently, the proposed approach focuses not merely on optimizing WPT efficiency at a single operating point, but on learning a control policy capable of maintaining efficient and stable wireless charging as the operating condition changes.

RESULTS AND DISCUSSION

Simulation Setup

To evaluate the effectiveness of the proposed PPO-based adaptive controller, the WPT charging system is modeled and simulated in MATLAB/Simulink R2024a. A series-series (SS) compensation topology is adopted with a rated charging power of 3.3 kW, representing a wireless charging configuration suitable for personal electric vehicles. To examine the adaptability of the controller under changing electromagnetic conditions, the separation distance between the transmitting and receiving coils is varied from 5 to 20 cm. This range is used to represent practical deviations in vehicle parking position and coil alignment, which can significantly affect the coupling condition and, consequently, the power-transfer performance.

The main parameters of the WPT system used in the simulations are summarized in Table 2.

Table 2. Simulation parameters of the WPT system

Parameter	Value
Battery capacity	50 Ah
Supply voltage	400 V
Rated charging power	3.3 kW
WPT operating frequency	85 kHz
Coil separation distance	5-20 cm
Coupling coefficient	0.15-0.45
Number of electric vehicles	50
PV source capacity	250 kW
Transmitting-coil inductance, Lp	200 µH
Receiving-coil inductance, Ls	200 µH

The reference charging current is set to I_{ref}=20 A, while the nominal operating frequency is f_{ref}=85 kHz. The PPO agent is constrained to generate physically feasible actions within $0 \leq I_{ch} \leq 200$ A and $80 \leq f_{op} \leq 100$ kHz. The battery initially operates at 50% SOC and is constrained within the range of 20–90% SOC. These limits are maintained consistently during both PPO training and performance evaluation.

The PPO controller is implemented using an Actor-Critic architecture composed of two feed-forward neural networks. The Actor network maps the observed system state to the control action, whereas the Critic network estimates the value of the current state and provides the evaluation signal required for policy improvement. This architecture is consistent with reinforcement-learning-based EV charging and dynamic wireless charging studies reported by (Jamjuntr and Suanpang 2024), (Hao and Xu 2024).

The PPO hyperparameters are selected through preliminary simulation trials with the objective of obtaining a reasonable compromise between learning speed and training stability. The adopted training configuration is given in Table 3.

Table 3. PPO training parameters

Parameter	Value
Number of training episodes	2000
Learning rate	1×10^{-4}
Discount factor (γ)	0.99
Clipping coefficient (ϵ)	0.2
Batch size	64
Entropy coefficient	0.01
Number of update epochs	10
Hidden-layer neurons	128

During the training and evaluation processes, the PPO agent continuously receives the current operating state of the WPT system and determines the charging current I_{ch} and operating frequency f_{op}. This closed-loop interaction enables the controller to modify its actions according to changes in the coupling condition rather than maintaining a predetermined operating point. The effectiveness of the learned policy is evaluated using four complementary criteria: the evolution of the average training reward, convergence behavior, WPT transfer efficiency, and the ability to maintain satisfactory efficiency over different coil separation distances.

To provide a meaningful benchmark, the proposed PPO controller is compared with a conventional control strategy based on a fixed operating frequency. Both approaches are evaluated under the same WPT configuration and coil-separation conditions. This comparison is particularly important because the principal objective of the proposed method is to improve adaptability to coupling variations, rather than merely to achieve a high efficiency at a single nominal operating point. The comparative evaluation therefore provides a direct indication of whether the PPO controller can effectively compensate for changes in the magnetic coupling and maintain more favorable power-transfer performance across a wider operating range.

Convergence Performance of the PPO Agent

The learning performance of the proposed PPO controller is first evaluated through the evolution of the average reward during the training process. The reward curve provides an important indication of whether the agent can progressively identify an effective control policy and whether the learned policy remains stable after training. Figure 3 presents the average reward obtained over 2000 training episodes.

As shown in Figure 3, the learning process can be divided into three representative stages. During the early stage, approximately the first 500 episodes, the average reward increases rapidly from an initial value of 12.4 to nearly 70. This pronounced improvement indicates that the PPO agent is progressively learning the relationship between the observed WPT operating conditions and the corresponding control actions. At this stage, the agent explores different combinations of charging current and operating frequency and gradually eliminates control actions that result in lower transfer efficiency or undesirable charging deviations.

A second learning stage can be observed between approximately 500 and 1200 episodes. The reward continues to increase, although at a considerably lower rate than during the initial stage. This behavior indicates that the agent has already acquired the basic characteristics of the WPT environment and is subsequently refining its control policy. The smaller improvement in reward during this stage is therefore not an indication of poor learning; rather, it reflects the transition from broad policy exploration toward more precise control-policy refinement.

After approximately 1200 episodes, the average reward remains within a relatively narrow range and approaches its maximum value in Table 4. The maximum observed reward is approximately 100.2, while the reward becomes consistently stable after around 1600 episodes. The remaining fluctuations are below approximately 2.5%, indicating that further training produces only limited changes in the learned policy.

Table 4. Convergence characteristics of the PPO agent

Performance indicator	Value
Number of training episodes	2000
Initial average reward	12.4
Maximum average reward	100.2
Approximate onset of convergence	1200 episodes
Stable convergence	1600 episodes
Post-convergence fluctuation	<2.5%

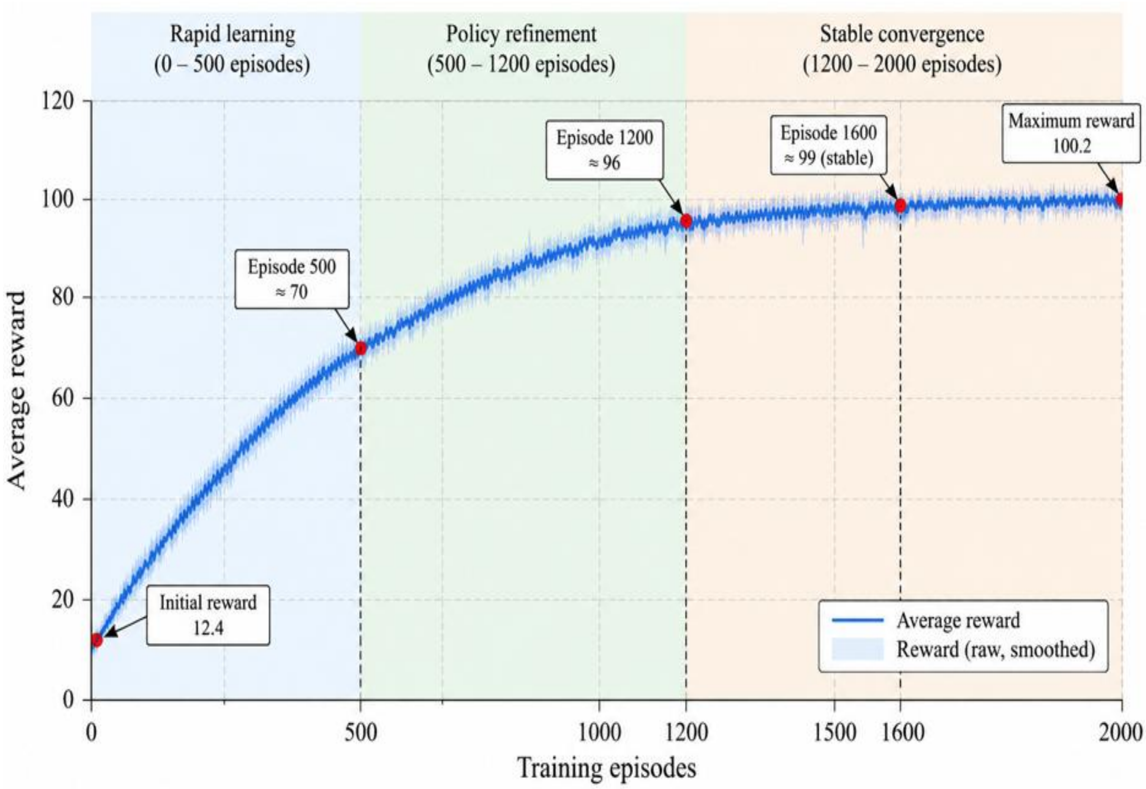

Figure 3. Convergence curve of the PPO agent

The convergence characteristics summarized in Table 4 demonstrate that the PPO agent is capable of learning a stable control policy within the prescribed training horizon. In particular, the substantial increase in reward during the first part of training confirms that the agent effectively extracts useful control information from the WPT system state. The subsequent reduction in reward variation indicates that the learned policy becomes increasingly consistent as training progresses.

From a control perspective, the convergence behavior is important because the proposed controller operates in a continuous state and action space. The agent must simultaneously determine suitable values of the charging current and operating frequency while responding to variations in the WPT coupling condition. The relatively smooth increase in reward and the limited fluctuations after convergence suggest that the PPO policy can adapt to this nonlinear control environment without exhibiting persistent instability.

The observed behavior can also be attributed to the clipped policy-update mechanism employed by PPO. By restricting the extent to which the updated policy can deviate from the previous policy, the clipping mechanism prevents excessively large policy changes during training. This provides a balance between exploration and policy refinement and helps avoid abrupt changes that could otherwise lead to unstable learning. The resulting convergence behavior is consistent with the role of PPO in continuous-control applications.

Importantly, the objective of the training process is not simply to maximize the numerical reward as quickly as possible. A useful policy must also produce physically meaningful control actions and remain effective when the WPT operating condition changes. Therefore, the convergence result in Figure 3 should be considered together with the subsequent efficiency and dynamic-response analyses. The latter tests determine whether the policy learned during training can translate its high reward into improved WPT performance under different coil separation distances.

WPT Transfer Efficiency under Different Coil Separation Distances

To evaluate the transient adaptability of the proposed controller, an abrupt increase in coil separation from 10 cm to 20 cm is introduced at $t=1\ s$ during the charging process. This disturbance represents a sudden deterioration in the magnetic coupling condition and is used to evaluate the ability of the PPO controller to restore the charging system to a stable operating condition. The proposed strategy is compared with conventional fixed-frequency control under identical system conditions, with the results summarized in Table 5.

Table 5. WPT transfer efficiency under different coil separation distances

Coil separation distance (cm)	Conventional control (%)	Proposed PPO (%)	Improvement (percentage points)
5	89.3	91	1.7
10	84.7	87.2	2.5
15	76.5	81.4	4.9
20	68.2	73.9	5.7

As shown in Table 5, the transfer efficiency decreases progressively as the coil separation distance increases for both control strategies. Under conventional fixed-frequency control, the efficiency decreases from 89.3% at 5 cm to 68.2% at 20 cm. The same general trend is observed with the proposed PPO controller, for which the efficiency decreases from 91.0% to 73.9% over the same distance range.

This reduction is consistent with the electromagnetic characteristics of the WPT system. According to (2), the coupling coefficient depends on the mutual inductance between the transmitting and receiving coils. Increasing the separation distance weakens the magnetic interaction and consequently reduces the mutual inductance. The resulting reduction in coupling adversely affects the power-transfer capability of the system, which is reflected directly in the efficiency defined in (3). Therefore, the downward trend in Figure 4 is physically expected rather than a limitation specific to either control strategy. More importantly, the proposed PPO controller consistently achieves higher efficiency than the conventional controller at all investigated distances. At 5 cm, PPO improves the efficiency from 89.3% to 91.0%, corresponding to an improvement of 1.7 percentage points. When the distance increases to 10 cm, the improvement increases to 2.5 percentage points, with PPO achieving 87.2% compared with 84.7% for the conventional strategy.

The advantage of adaptive control becomes more evident under weaker coupling conditions. At 15 cm, the proposed controller maintains an efficiency of 81.4%, whereas the conventional controller achieves only 76.5%, resulting in a difference of 4.9 percentage points. At the maximum investigated separation of 20 cm, PPO maintains an efficiency of 73.9%, compared with 68.2% for fixed-frequency control. The resulting improvement of 5.7 percentage points represents the largest performance difference observed in the tested operating range. These results provide an important indication of the behavior of the proposed controller. The benefit of PPO is relatively modest when the coils are closely aligned and strongly coupled, because the conventional controller can already operate under a favorable electromagnetic condition. As the separation distance increases, however, the operating condition becomes progressively more challenging, and the limitation of a fixed-frequency strategy becomes more pronounced. The PPO controller is able to compensate for part of this deterioration by modifying the charging current and operating frequency according to the observed system state.

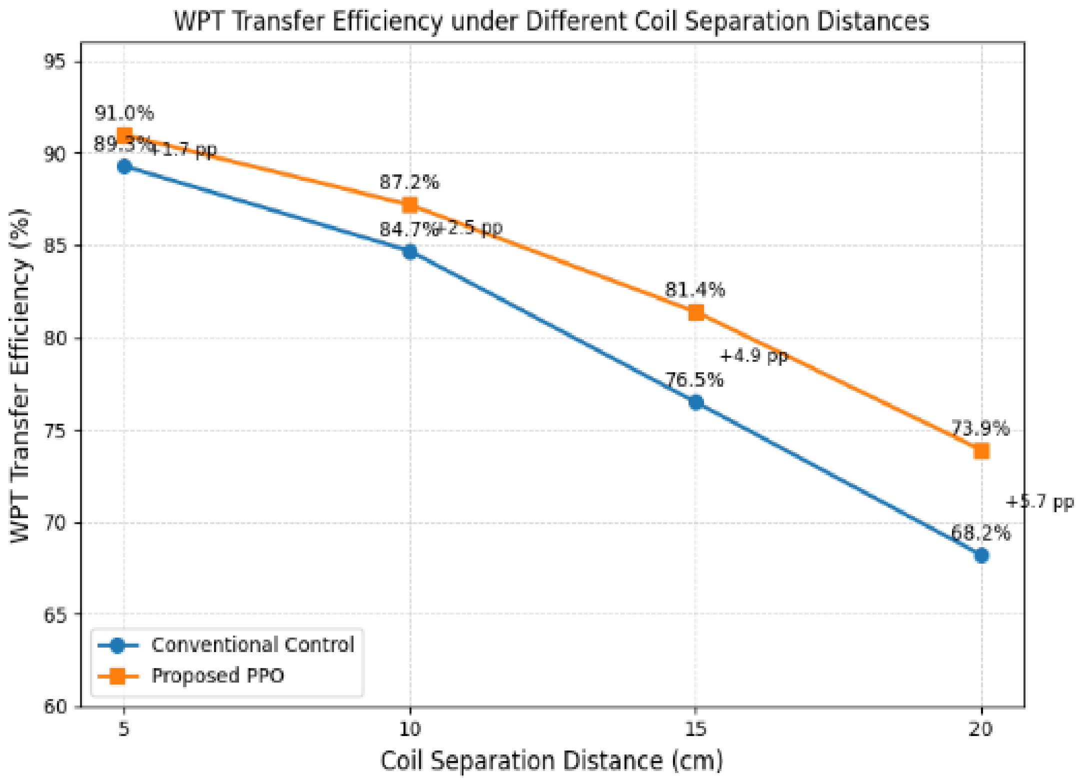

Figure 4. WPT transfer efficiency under different coil separation distances.

Figure 4 compares the transfer efficiency of the conventional and PPO-based controllers for coil separation distances of 5-20 cm. As expected, efficiency decreases with increasing separation because of the deterioration of magnetic coupling. However, the PPO controller consistently maintains higher efficiency than the conventional strategy, with the performance gap becoming more pronounced at larger distances. This improvement results from the adaptive nature of PPO. Unlike fixed-frequency control, the PPO controller uses the current WPT state to jointly adjust the charging current and operating frequency, allowing the operating point to respond to changes in coupling conditions.

Importantly, PPO does not eliminate the physical efficiency loss associated with weaker coupling; instead, it mitigates this degradation. At 20 cm, for example, PPO achieves 73.9%, compared with 68.2% for conventional control, corresponding to an improvement of 5.7 percentage points. The increasing advantage at larger separation distances demonstrates that adaptive control becomes more valuable under unfavorable coupling conditions. Overall, the results in Table 5 and Figure 4 confirm that the proposed PPO strategy provides a more robust efficiency response across the investigated 5-20 cm range, supporting its suitability for wireless EV charging under variable coil positioning.

Adaptive Charging Response

The dynamic behavior of the proposed PPO controller is further evaluated by introducing an abrupt change in the coil separation distance during the charging process. This test is designed to examine whether the learned control policy can respond effectively to a sudden change in the electromagnetic coupling condition and restore the charging system to a stable operating state. Unlike the steady-state efficiency assessment presented in Section convergence performance of the PPO agent, this experiment focuses on the transient behavior of the controller. In a practical wireless EV charging system, the relative position between the transmitting and receiving coils may change during operation because of parking-position errors, vehicle movement, or mechanical disturbances. Such changes can modify the coupling coefficient and cause a temporary deviation in the transferred power and charging current. Therefore, the ability to recover rapidly from these disturbances is an important characteristic of an adaptive WPT controller.

The settling time is defined as the time required for the charging current to enter and remain within a ±2% band around its reference value following the coil-separation disturbance. The maximum deviation is evaluated relative to the pre-disturbance steady-state charging current, while the steady-state error is calculated from the final charging-current value relative to I_{ref}. The main dynamic-response indicators obtained from the simulation are summarized in Table 6.

Table 6. Dynamic response characteristics of the proposed PPO controller

Performance indicator	Value
Settling time	0.48 s
Maximum deviation	5.20%
Steady-state error	<2.8%
Charging-current oscillation	<2.0%
Stability loss	No instability observed

As shown in Table 6, the proposed PPO controller exhibits a settling time of approximately 0.48 s following the change in the coil separation distance. The maximum transient deviation is limited to 5.2%, after which the charging current gradually returns toward its desired operating level. The steady-state error remains below 2.8%, indicating that the controller can recover the charging condition with relatively small residual deviation.

The charging-current oscillation is also maintained below approximately 2.0%. This result is important because rapid adaptation alone is not sufficient for practical WPT operation. A controller that responds rapidly but produces excessive oscillations could adversely affect the power electronic converters, resonant circuit, and battery charging process. The limited oscillation observed in this study indicates that the PPO policy provides a reasonable compromise between response speed and control smoothness.

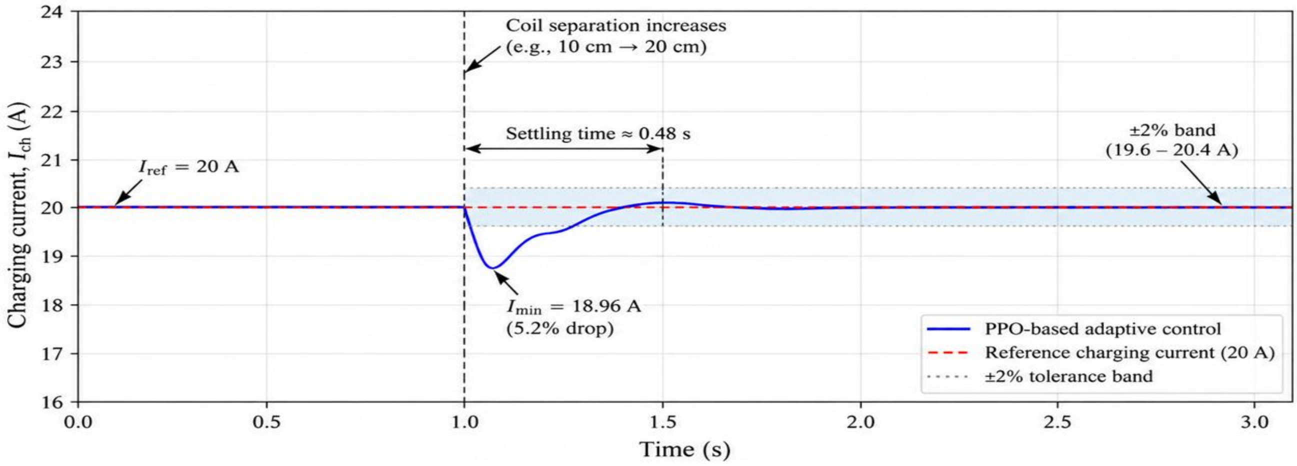

Figure 5. Adaptive charging-current response of the PPO controller.

Figure 5 shows that the charging current responds immediately after the change in coil separation and subsequently converges to a stable value. During this process, the PPO agent adapts both the charging current and operating frequency according to the updated system state. This state-dependent control enables the WPT system to compensate for part of the performance degradation caused by the changed coupling condition.The transient deviation is not completely eliminated because a sudden change in coil separation inherently alters the electromagnetic characteristics of the WPT system. Nevertheless, the relatively short settling time and limited oscillation demonstrate that PPO can rapidly mitigate the disturbance without compromising system stability. the results confirm that the proposed PPO controller provides effective dynamic adaptation under variable coupling conditions.

To further demonstrate the control actions generated by the PPO agent, Figure 6 presents the corresponding charging-current and operating-frequency responses following a sudden change in coil separation. The two control variables are jointly adjusted according to the updated WPT operating state. The charging current changes rapidly to maintain the desired charging condition, while the operating frequency is simultaneously modified to compensate for the variation in magnetic coupling. This coordinated adjustment confirms that the PPO agent does not rely on a fixed-frequency strategy but actively adapts both control variables to the prevailing WPT condition.

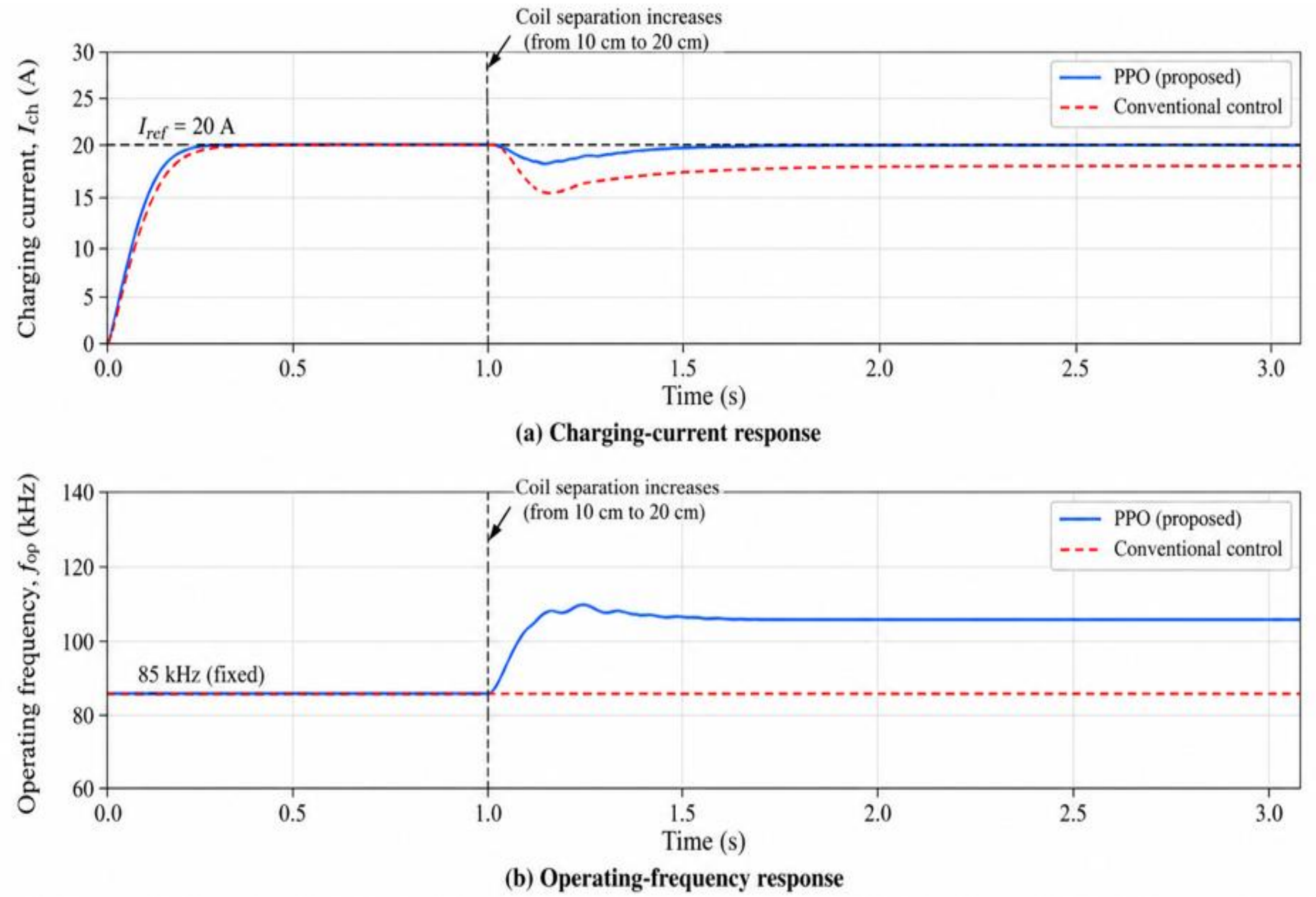

Figure 6. Adaptive charging current and operating-frequency responses under a sudden change in coil separation.

Comparison with Existing Studies

To further assess the effectiveness and research relevance of the proposed approach, the obtained results are compared with representative recent studies addressing wireless power transfer and reinforcement-learning-based EV charging. The comparison is summarized in Table 7. Rather than comparing numerical values that represent fundamentally different performance metrics, the comparison focuses on the main methodological approach and the principal outcome reported in each study.

Table 7. Comparison of the proposed approach with representative related studies

Study	Main approach	Key reported outcome
(S. A. Benfadhel et al. 2025)	Optimal WPT control	Maximum transfer efficiency of approximately 90.5%
(S. Zhao et al. 2024)	Model-based adaptive WPT control	Stable operation under variations in mutual inductance and load
(S. Zhang et al. 2023)	Safe reinforcement learning for EV charging	Reduction of operational constraint violations
(Z. Zhang, Liu, and Sun 2025)	Safe deep reinforcement learning for EV energy management	Approximately 92% of energy demand satisfied
This study	PPO adaptive WPT control	91.0% at 5 cm; 87.2%, 81.4%, and 73.9% at 10, 15, and 20 cm, respectively

The reported numerical values are not intended as a direct one-to-one benchmark because the compared studies employ different system configurations, operating conditions, and performance metrics. Table 7 compares the proposed PPO-based controller with representative studies addressing WPT control and intelligent EV charging. Existing WPT studies mainly rely on model-based optimization or adaptive control. Benfadhel et al., for example, reported a maximum transfer efficiency of approximately 90.5%, while Zhao et al. demonstrated stable WPT operation under variations in mutual inductance and load. These approaches provide effective control solutions but generally depend on explicit system models and predefined control relationships.

Recent studies have also introduced reinforcement learning into EV charging. S. Zhang et al. employed Safe RL to reduce operational constraint violations, whereas Z. Zhang et al. applied Safe DRL to EV energy management and reported approximately 92% of energy demand satisfaction. However, these studies primarily address charging management and energy scheduling rather than the electromagnetic control of the WPT process itself.

In contrast, the proposed method applies PPO directly to the physical WPT control layer. The controller uses the coupling coefficient, transfer efficiency, battery voltage, and charging current as the system state, while jointly regulating charging current and operating frequency. This enables the control policy to respond to changes in the electromagnetic coupling condition without relying exclusively on predefined control relationships.

The proposed controller achieves a transfer efficiency of 91.0% at a 5 cm coil separation, slightly exceeding the 90.5% reported by Benfadhel et al. More importantly, the performance remains competitive as the separation distance increases, with efficiencies of 87.2%, 81.4%, and 73.9% at 10, 15, and 20 cm, respectively. The results therefore demonstrate that the proposed approach is not limited to optimizing performance at a favorable operating point but can continuously adapt to progressively weaker coupling conditions.

DISCUSSION

The increasing performance gap at larger separation distances suggests that the learned policy does not merely optimize a nominal operating point. Instead, it modifies the control variables according to the changing electromagnetic state, allowing part of the coupling-induced performance degradation to be mitigated. The results demonstrate that the main benefit of PPO is its ability to mitigate efficiency degradation caused by changing coil separation. As the distance increases, the physical coupling between the transmitting and receiving coils becomes weaker, leading to an expected reduction in transfer efficiency. Nevertheless, PPO consistently maintains higher efficiency than the conventional controller across the investigated range.

The improvement becomes more pronounced under unfavorable coupling conditions. At 15 cm, PPO increases efficiency from 76.5% to 81.4%, corresponding to an improvement of 4.9 percentage points. At 20 cm, the efficiency increases from 68.2% to 73.9%, giving a 5.7-percentage-point improvement. This trend indicates that adaptive control provides greater value when the WPT system deviates from its nominal operating condition.

The dynamic response results further support the adaptability of the proposed controller. Following an abrupt change in coil separation, the PPO controller rapidly adjusts the charging current and operating frequency and reaches a new operating condition with a settling time of approximately 0.48 s. The transient response remains controlled, with a maximum deviation of 5.2% and steady-state error below 2.8%. No stability loss is observed during the tested operating conditions.

In addition, the generalization of the trained PPO policy to coupling conditions outside the training range has not been evaluated. Therefore, the reported performance should be interpreted within the operating range represented in the training and testing environment.

Overall, the comparison indicates that the contribution of the proposed method lies not simply in applying reinforcement learning to EV charging, but in extending learning-based control to the electromagnetic WPT layer. By jointly adapting the charging current and operating frequency according to the instantaneous system state, PPO provides a practical mechanism for maintaining charging performance under variable magnetic coupling. Nevertheless, the present evaluation is simulation-based and does not yet consider factors such as lateral coil misalignment, measurement noise, component tolerances, temperature variation, or battery-condition changes. These aspects should be investigated through experimental validation in future work.

CONCLUSION

This study proposed a PPO-based adaptive control strategy for wireless power transfer (WPT) in electric vehicle charging systems operating under variable coil coupling conditions. Unlike conventional control approaches that rely on predetermined operating parameters, the proposed method enables the charging system to adapt its operating frequency and charging current according to the instantaneous WPT condition. The simulation results demonstrate that the PPO controller can learn an effective control policy and maintain stable operation despite changes in coil separation. More importantly, the proposed strategy consistently improves power-transfer efficiency compared with conventional control, particularly when the magnetic coupling becomes weaker. This indicates that the controller is able to compensate, to a practical extent, for performance degradation caused by changes in the charging geometry rather than relying on a fixed operating point. The dynamic response results further show that the proposed controller can respond rapidly to changes in coupling while maintaining smooth charging behavior and avoiding significant oscillations or instability. The combination of adaptive control and reinforcement learning therefore provides a promising solution for WPT systems in which coil position and coupling conditions cannot be perfectly maintained. Overall, the main contribution of this work is to demonstrate the feasibility of using PPO as an intelligent control layer for adaptive wireless EV charging. The proposed strategy improves the ability of the WPT system to maintain efficient and stable operation under variable coupling conditions, providing a useful foundation for more flexible and robust wireless charging systems. Future work will focus on experimental validation and on extending the control framework to more realistic operating conditions, including lateral coil misalignment, measurement noise, parameter uncertainties, temperature variations, and different battery operating conditions. These developments will further assess the robustness and practical applicability of the proposed PPO-based WPT control strategy.

REFERENCES

1. Chen, F, H Hu, L Zhao, A Padilla, and J Hou. 2023. “A Linear Parameter-Varying Hammerstein Model for Dynamic Modeling of WPT Systems.” *IEEE Transactions on Power Electronics* 38(12): 16230–16244.
2. Feng, Y, F Chen, Y Sun, T Lin, F Meng, and H Hu. 2024. “Identification and Control of LCC-S WPT Systems Using a Linear Parameter Varying Model.” *IEEE Transactions on Power Electronics* 39(9): 11862–11873.

3. Hao, L, and Y Xu. 2024. "Reinforcement Learning of Threshold Policy for Dynamic Wireless Charging with Distributed Energy Resources." In *Proc. IEEE PES General Meeting (PESGM) Seattle, WA, USA, 2024*, Seattle, WA, USA, 1–5.
4. Jamjuntr, P, and P Suanpang. 2024. "Adaptive Multi-Agent Reinforcement Learning for Electric Vehicle Charging Networks." *World Electric Vehicle Journal* 15(10): 453–468.
5. Park, K, and I Moon. 2022. "Multi-Agent Deep Reinforcement Learning Approach for EV Charging Scheduling in a Smart Grid." *Applied Energy* 328: 120111.
6. S. A. Benfadhel, A, A J M Cardoso, and M Trabelsi. 2025. "Intelligent Optimization and Real-Time Control of Wireless Power Transfer for Electric Vehicles." *IEEE Access* 13: 45872–45886.
7. S. A. Sagar, A. Kashyap, M. A. Nasab, S. Padmanaban, M. Bertoluzzo, A. Kumar, and M. Al-Haddad. 2023. "A Comprehensive Review of the Recent Development of Wireless Power Transfer Technologies for Electric Vehicle Charging Systems." *IEEE Access* 11: 83703–83751.
8. Xu, X, Y Jia, Y Xu, Z Xu, and C S Lai. 2024. "Multi-Agent Reinforcement Learning-Based Data-Driven Energy Management for Electric Vehicle Charging Systems." *IEEE Transactions on Smart Grid* 15(2): 1324–1338.
9. Xue, Z, W Liu, C Liu, and K T Chau. 2025. "Critical Review of Wireless Charging Technologies for Electric Vehicles." *World Electric Vehicle Journal* 16(2): 65.
10. Zhang, S, R Jia, H Pan, and Y Cao. 2023. "A Safe Reinforcement Learning-Based Charging Strategy for Electric Vehicles in Residential Microgrid." *Applied Energy* 348: 121490.
11. Zhang, Z, W Liu, and H Sun. 2025. "Model-Free Safe Deep Reinforcement Learning for Grid-to-Vehicle Charging Management." *Engineering Applications of Artificial Intelligence* 145: 112529.
12. Zhao, S, F Chen, C Tang, P Deng, and C Duan. 2024. "Modeling and Control of WPT Systems in the Presence of Load and Mutual Inductance Variations." *IEEE Transactions on Power Electronics* 39(11): 15315–15328.
13. Zhao, Z, S Tang, C Chen, D Zhao, and P Deng. 2023. "Modeling and Control of the Wireless Power Transfer System Subject to Input Nonlinearity and Communication Delay." *IEEE Transactions on Power Electronics* 38(11): 14776–14787.