energies

MDPI

Article

Temporal-Feature-Enhanced Reinforcement Learning Control with Adaptive Speed-Loop Gain for PMSM Drives

Hongquan Zhang [1,†], Jingjun Cui [1], Zhihan Wu [2,3,*,†], Anqi Situ [2] and Jiaqiang Yang [2,*]

[1] COSMOPlat AIoT Technology, Qingdao 266100, China; zhanghongq@haier.com (H.Z.); cuijingjun@haier.com (J.C.)
[2] College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China; 22410189@zju.edu.cn
[3] Polytechnic Institute, Zhejiang University, Hangzhou 310015, China
* Correspondence: wuzhihan@zju.edu.cn (Z.W.); yangjiaq@zju.edu.cn (J.Y.)
† These authors contributed equally to this work.

Abstract

The high-performance control of permanent magnet synchronous motor (PMSM) drives is challenged by operating uncertainties and coupled electromechanical dynamics. This paper proposes a temporal-feature-enhanced reinforcement learning control approach for a PMSM drive system with adaptive regulation of the speed-loop gain. The learning agent is integrated into a dual-loop control structure to generate voltage control commands and update the speed-loop gain in real time. A multiobjective learning criterion is designed to balance tracking accuracy, torque smoothness, and gain boundedness. Temporal error features are further embedded in the state representation to capture short-term transient trends with limited additional complexity. Comparative simulations with conventional PI control and standard TD3 indicate that the proposed method improves transient response, reduces speed tracking error, and suppresses torque ripple. These results demonstrate the effectiveness of temporal-feature-enhanced reinforcement learning for coordinated control of PMSM drives.

Keywords: permanent magnet synchronous motor (PMSM); reinforcement learning (RL); twin delayed deep deterministic policy gradient (TD3); temporal feature enhancement; multiobjective optimization control

1. Introduction

Permanent magnet synchronous motor (PMSM) drives have been extensively utilized in industrial automation, electric transportation, and high-performance servo systems because of their high efficiency, compact structure, and fast dynamic response [1,2]. In practical applications, field-oriented control (FOC) with cascaded proportional–integral (PI) regulators remains a standard solution for PMSM drives, since it offers a clear control structure and efficient embedded implementation [3–5]. However, the efficacy of this scheme still depends heavily on the precise calibration of the controller and the accurate determination of motor model parameters. Under conditions involving parameter variations, load disturbances, and changes in operating conditions, the fixed-gain structure may no longer adequately maintain the desired current and speed responses.

To enhance dynamic performance and robustness, a range of advanced control strategies have been investigated for PMSM drives. Adaptive control has been employed to compensate for variations in operating conditions [3,4]. Sliding-mode control (SMC) has

Academic Editor: Krzysztof Górecki

Received: 27 June 2026
Revised: 27 July 2026
Accepted: 4 August 2026
Published: 5 August 2026

been examined for its resilience to disturbances and uncertain parameters [6,7]. Additionally, the model predictive control (MPC) method has been implemented in PMSM drives, with the objective of enhancing tracking performance [8]. Among these methods, MPC has attracted considerable attention due to its ability to address multivariable coupling and control constraints within a finite-horizon optimization framework. Nevertheless, its practical performance still relies on the accuracy of the prediction model and the online computational resources, which limits its flexibility in uncertain operating conditions [9,10]. These challenges have encouraged increasing interest in data-driven control, where machine learning (ML) techniques are utilized to extract informative features or control policies directly from system data [11,12].

Among learning-based methods, reinforcement learning (RL) provides a promising framework for motor-drive control because it can optimize control policies through interaction with the environment, rather than relying on labeled data or predesigned expert controllers. A range of studies have explored the application of RL to PMSM torque control [13], current regulation [14] and speed control [15], indicating its potential to reduce model dependence and alleviate the manual effort required for tuning. Some early applications of RL to motor control depend on value-based methods, such as deep Q-learning, which are well-suited to discrete switching decisions [16]. For control tasks with continuous action variables, actor–critic methods such as deep deterministic policy gradient (DDPG) have been employed to generate voltage or correction commands directly [17,18], and twin delayed deep deterministic policy gradient (TD3) further improves learning stability by mitigating value overestimation and smoothing policy updates [19,20]. Additionally, studies on feature embedding have demonstrated that transforming controller inputs into richer state representations enhances RL training and testing performance, while memory mechanisms such as long short-term memory (LSTM) have been introduced to capture temporal information in motor-control tasks [10,21]. These findings imply that enhancing the observation design effectively improves the state awareness and control capability of RL-based drive controllers.

Despite the encouraging progress of RL-based PMSM control, its integration with cascaded FOC structures still offers several opportunities for further refinement:

(i) Methods that focus on specific control objectives have verified the effectiveness of RL in PMSM control, including current regulation, torque control, and speed-loop compensation. Concurrently, learning-based optimization has also demonstrated promising performance in balancing multiple control objectives. For instance, a recent study on multiobjective MPC employed TD3 to adjust the weighting factors in a customized cost function involving torque regulation, flux tracking, and switching-sequence optimization [22]. Motivated by these studies, this paper further considers coordinating the dynamic coupling among current regulation, torque generation, and speed control in the cascaded FOC structure, since these processes are inherently interconnected in PMSM drives. The integration of these coupling relationships into the learning objective and controller coordination design may enable the control strategy to more effectively capture the coupled electromechanical behavior.

(ii) Furthermore, the observation design of RL agents plays an important role in determining the quality of the learned control policy. Some RL-based PMSM controllers are formulated primarily based on immediate observations, providing a compact and efficient state representation. Concurrently, some researchers have indicated that richer feature representations can improve RL performance, and memory cells can help capture the temporal characteristics of motor-drive dynamics [10,21–23]. For PMSM drives operating under reference changes and load disturbances, the recent evolution of key tracking errors contains valuable information regarding transient

trends. Therefore, incorporating the dynamic progression of key elements into the observation space becomes a reasonable way to enhance the agent's decision-making capability without significantly increasing input complexity.

To address these considerations, this paper proposes a memory-augmented adaptive-gain TD3 method (MAG-TD3) for FOC-based PMSM drives, with temporal error features embedded in the state perception. The proposed method is embedded in the conventional cascaded FOC framework. The RL agent generates voltage commands for current regulation while simultaneously adjusting the proportional gain of the outer speed PI loop. This coordinated design aims to improve the interaction between the electrical and mechanical dynamics within the cascaded control architecture. Additionally, the reward function is reformulated to incorporate torque-related regulation and gain-variation penalties. This enables the learning objective to more accurately reflect electromechanical coupling while promoting smooth adaptive-gain behavior. Furthermore, a lightweight recurrent memory module is introduced to encode the recent evolution of speed and q-axis current errors into compact temporal features. Consequently, the agent can better perceive transient error trends under reference changes and load disturbances, without requiring a substantial increase in input complexity.

The main contributions of this paper are summarized as follows:

1. A MAG-TD3 control framework incorporating temporal features is developed for FOC-based PMSM drives. By integrating the RL agent into the cascaded control structure, the proposed method enables coordinated learning of voltage commands for current regulation and adaptive gain tuning for speed control while preserving compatibility with conventional FOC implementation.
2. A learning objective considering electromechanical coupling is designed by restructuring the reward function to incorporate torque-related regulation and adaptive-gain penalty. This objective more effectively captures the coupling among q-axis current dynamics, electromagnetic torque generation, and rotor-speed response, allowing the learned policy to align more closely with the coupled electrical–mechanical behavior of the PMSM system.
3. A lightweight memory-augmented actor structure is introduced to incorporate short-term dynamic temporal features into the policy learning process. By enriching the state representation with temporal features, the proposed structure improves awareness of transient tracking trends under reference variations and dynamic disturbances, while concurrently avoiding an excessive increase in input complexity.

2. Problem Formulation

2.1. Mathematical Model of PMSM

The mathematical model of the three-phase PMSM in this paper is considered in the d–q rotating reference frame. To simplify the model and ensure its feasibility for real-time implementation feasibility, high-frequency effects such as magnetic saturation, core losses, and temperature variations were neglected. Under these assumptions, the stator voltage equations can be expressed as follows [3]:

$$v_d = R_s i_d + L_d \frac{di_d}{dt} - \omega_e L_q i_q \tag{1}$$

$$v_q = R_s i_q + L_q \frac{di_q}{dt} + \omega_e \left(L_d i_d + \psi_f \right) \tag{2}$$

where i_d, i_q, v_d, v_q, L_d, and L_q are the stator currents, voltages, and inductances in the d–q frame, respectively. R_s is the stator resistance, ψ_f is the permanent magnet flux linkage, and ω_e is the electrical angular velocity.

The electromagnetic torque T_e generated by the PMSM is given by

$$T_e = \frac{3}{2}p\left[\psi_f i_q + (L_d - L_q) i_d i_q\right] \tag{3}$$

For surface-mounted PMSM, the d-axis and q-axis inductances are generally equal: $L_d = L_q$. Accordingly, Equation (3) can be approximated as follows:

$$T_e = K_t i_q = \frac{3}{2}p\psi_f i_q \tag{4}$$

where K_t denotes the torque constant and p is the number of rotor pole pairs. Equation (4) indicates that i_q is the dominant variable for T_e generation, whereas i_d is typically set to zero for efficient FOC operation.

The proposed formulation does not impose $i_d = 0$ as a structural constraint. Instead, i_d is regulated by the tracking error in Equation (16). The actor generates the d-axis voltage action $\overline{v}_{d,t}$ in Equation (24), which regulates i_d through the d-axis dynamics in Equation (7). Hence, the framework can accommodate a prescribed non-zero i_d^{ref}, subject to the voltage and current constraints.

The mechanical dynamics of the motor are described by

$$J\frac{d\omega_m}{dt} = T_e - T_L - B\omega_m = T_e - T_L - B\frac{\omega_e}{p} \tag{5}$$

where ω_m is the mechanical angular velocity of the rotor, J is the total inertia, B is the viscous friction coefficient, and T_L is the external load torque. Substituting Equation (3) into Equation (5) yields the explicit electromechanical coupling relation.

$$J\frac{d\omega_m}{dt} = \frac{3}{2}p\left[\psi_f i_q + (L_d - L_q) i_d i_q\right] - T_L - B\omega_m \tag{6}$$

Equation (6) demonstrates that the electrical current dynamics and mechanical speed dynamics are coupled through T_e. Consequently, precise current regulation directly affects torque production and rotor speed response. This coupling renders coordinated control a more appropriate objective than an isolated current-tracking criterion.

For digital simulation and control implementation, the current dynamics are discretized with the sampling period T_s. The discrete-time current model in the dq-axis is obtained by applying the Euler-forward discretization method to Equations (1) and (2).

$$i_d(k+1) = i_d(k) + \frac{T_s}{L_d}\left[-R_s i_d(k) + \omega_e L_q i_q(k) + v_d\right] \tag{7}$$

$$i_q(k+1) = i_q(k) + \frac{T_s}{L_q}\left[-R_s i_q(k) - \omega_e\left(L_d i_d(k) + \psi_f\right) + v_q\right] \tag{8}$$

The PMSM is supplied by a three-phase two-level voltage-source inverter (VSI), and the switching states of the three bridge legs are denoted by S_a, S_b, and S_c, respectively. Consequently, the inverter is capable of generating eight distinct switching states. The corresponding stator voltage space vector generated by the inverter and the feasibility constraints in the dq-axis are given by

$$u_s = \frac{2}{3}V_{\text{dc}}\left(S_a + S_b \cdot e^{j\frac{2}{3}\pi} + S_c \cdot e^{-j\frac{2}{3}\pi}\right) \tag{9}$$

$$v_d^2 + v_d^2 \le V_{\text{lim}}^2,\ V_{\text{lim}} = V_{\text{dc}}/\sqrt{3} \tag{10}$$

where u_s denotes the inverter output voltage space vector, V_{dc} is the DC-link voltage, and V_{lim} is the maximum admissible magnitude of the reference voltage vector within the modulation region.

The overall implementation procedure of the proposed control strategy is illustrated in Figure 1. The PMSM drive is regulated within a cascaded FOC structure, where the learning agent generates the normalized d-q-axis voltage commands and the adaptive proportional gain factor for the speed loop. The performance variables and recursive memory features are then employed to construct the observation vector. The generated control actions are subsequently restricted by the admissible bounds prior to their implementation in the drive system.

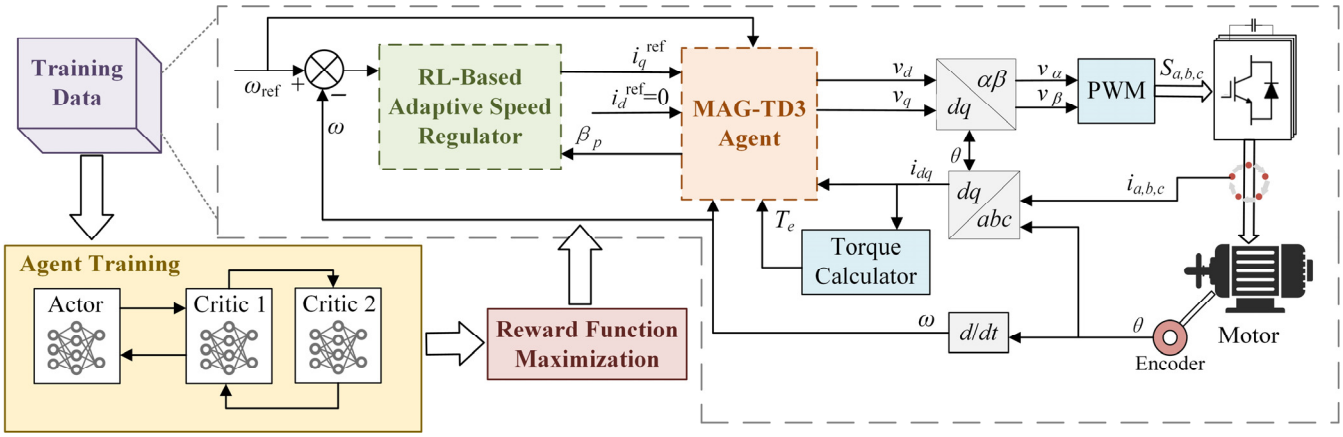

Figure 1. Overall flowchart of the RL-based control strategy for PMSM. Colored blocks represent functional modules, and dashed boxes indicate subsystem boundaries.

2.2. Gain-Adaptive Cascade FOC Structure

The proposed controller in the outer speed loop retains the standard PI-based structure and introduces a bounded online adjustment of its proportional path. This design maintains the outer-loop structure and standard FOC interfaces, while enabling adjustment of the speed-loop responsiveness according to operating conditions.

In the proposed cascaded architecture, the outer speed controller generates the torque-producing current reference i_q^{ref}, while the d-axis current reference is set to $i_d^{ref} = 0$ for the considered surface-mounted PMSM. The conventional inner d- and q-axis current PI regulators are replaced by the MAG-TD3 agent rather than being used in parallel with it. Specifically, the current tracking errors $e_d = i_d^{ref} - i_d$ and $e_q = i_q^{ref} - i_q$, together with the speed-related variables and temporal features, are included in the agent observation. The agent directly outputs the normalized voltage actions $\overline{v}_d$ and $\overline{v}_q$, as well as the adaptive gain action. After action scaling and constraint projection, the voltage commands v_d and v_q are converted into three-phase references through the inverse coordinate transformation and applied to the PWM inverter. The adaptive gain action modifies only the proportional gain of the outer speed PI loop.

The retained speed-loop controller can be mathematically represented as follows:

$$i_q^{\text{ref}} = \left(K_p^\omega + \frac{K_i^\omega}{s}\right)(\omega_{\text{ref}} - \omega_m) - B_a\omega_m \tag{11}$$

where ω_{ref} is the reference speed. The selection of gains is determined by the desired speed-loop bandwidth ω_b, in which $K_p^\omega = \omega_b J/K_t$, $K_i^\omega = \omega_b K_p^\omega$, and $B_a = (\omega_b J - B)/K_t$. In these definitions, the term $B_a\omega_m$ represents an active damping component.

The speed controller is updated at discrete times, and the adaptive signal acts on the proportional path. The q-axis current reference is reformulated as follows:

$$i_q^{\mathrm{ref}}(k) = K_{p,\mathrm{eff}}^{\omega}(k)(\omega_{\mathrm{ref}}(k) - \omega_m(k)) + x_\omega(k) \tag{12}$$

$$x_\omega(k) = x_\omega(k-1) + K_i^{\omega} T_\omega(\omega_{\mathrm{ref}}(k) - \omega_m(k)) \tag{13}$$

where $x_\omega(k)$ signifies the discrete integral state of the speed loop and T_ω denotes the speed-loop sampling period.

The effective proportional gain is defined as

$$K_{p,\mathrm{eff}}^{\omega}(k) = K_p^{\omega}\left[1 + \kappa_\beta \overline{\beta}_p(k)\right] \tag{14}$$

where $\overline{\beta}_p(k)$ denotes the bounded adaptive gain factor generated by the learning-based controller at control instant k, and the interval $[\beta_{\min}, \beta_{\max}]$ is predefined to ensure safe gain adaptation. The parameter κ_β is the gain adaptation coefficient.

Accordingly, the effective proportional gain is bounded by

$$K_p^{\omega}\left(1 + \kappa_\beta \beta_{\min}\right) \le K_{p,\mathrm{eff}}^{\omega}(k) \le K_p^{\omega}\left(1 + \kappa_\beta \beta_{\max}\right) \tag{15}$$

Equations (11)–(15) define the gain-adaptive speed-loop model used in the proposed control framework. Compared to the fixed-gain PI speed loop, the proposed structure introduces a constrained online adaptation of the proportional gain while maintaining the classic PI form. This mechanism enables the outer speed loop to respond more flexibly to reference variations and load disturbances.

2.3. Control Objective

The objective of the optimization process is to minimize the total control cost of the proposed PMSM drive system, while ensuring that the physical operating constraints are satisfied. The cost function consists of current-tracking cost, speed-regulation cost, torque-smoothness cost, voltage command cost, and adaptive-gain penalty. The current-tracking terms are introduced to maintain accurate d–q axis current regulation, the speed-regulation term is used to improve the dynamic tracking response, and the torque-smoothness term is included to reduce electromagnetic torque ripple. Concurrently, the voltage command cost and adaptive-gain penalty are incorporated to prevent excessively large voltage commands and overly aggressive adaptive-gain adjustments, respectively. The tracking error and the torque variation employed to quantify torque smoothness are defined as follows:

$$e_{dq}(k) = i_{dq}^{\mathrm{ref}}(k) - i_{dq}(k), e_\omega(k) = \omega_{\mathrm{ref}}(k) - \omega_m(k) \tag{16}$$

$$\Delta T_e(k) = T_e(k) - T_e(k-1) \tag{17}$$

To avoid scale imbalances among variables with different physical units, the normalized quantities are introduced as $\overline{e}_{dq}(k) = e_{dq}(k)/I_b$, $\overline{e}_\omega(k) = e_\omega(k)/\omega_b$, $\Delta\overline{T}_e(k) = \Delta T_e(k)/T_b$, $\overline{v}_{dq}(k) = v_{dq}(k)/V_b$, $\overline{\beta}_p(k) = \beta_p(k)/\beta_b$. Here, I_b, ω_b, T_b, V_b, and β_b denodenote the corresponding base values, which are selected according to the rated or admissible maximum values. The control sequence over a decision window of length N is as follows:

$$U_N(k) = \{u(k), u(k+1), \ldots, u(k+N-1)\},\ u(k) = \left[v_d(k), v_q(k), \beta_p(k)\right]^{\mathrm{T}} \tag{18}$$

The normalized voltage command vector $u(k+j)$ and the compact performance vector $z(k+j)$ are defined as follows:

$$\begin{cases} u_v(k+j) = \overline{v}_{dq}(k+j) = \left[\overline{v}_d(k+j), \overline{v}_q(k+j)\right]^{\mathrm{T}} \\ z(k+j) = \left[\overline{e}_{dq}^{\mathrm{T}}(k+j), \overline{e}_\omega(k+j), \Delta\overline{T}_e(k+j)\right]^{\mathrm{T}} \end{cases} \tag{19}$$

The optimization problem can be formulated as

$$\min_{U_N} J_C = \sum_{j=1}^{N}\left[\|z(k+j)\|_Q^2 + \|u(k+j)\|_R^2\right], \text{ s.t. Equations (9), (10) and (15)} \tag{20}$$

$$\|z(k+j)\|_Q^2 = z^{\mathrm{T}}(k+j)Qz(k+j),\ \|u(k+j)\|_R^2 = u^{\mathrm{T}}(k+j)Ru(k+j) \tag{21}$$

where $Q \succeq 0$ and $R \succeq 0$ denote positive semidefinite weighting matrices, which can be selected as diagonal matrices with nonnegative entries.

As illustrated in Equation (21), the first quadratic term penalizes the normalized compact performance vector and the second quadratic term limits excessive voltage utilization. The constraints delineated in Equations (10) and (15) ensure admissible voltage commands and restrict the adaptive-gain factor within prescribed limits.

3. Learning-Based Framework for PMSM Control

This section presents the learning-based control framework for the PMSM drive system. The reinforcement learning agent is embedded in the cascaded FOC structure, replacing the conventional d- and q-axis current PI regulators. The agent generates normalized d- and q-axis voltage actions while adjusting the proportional gain of the retained outer speed PI loop. The framework is formulated as a Markov decision process (MDP), and the TD3 algorithm is adopted to learn a continuous control policy, whose architecture is illustrated in Figure 2.

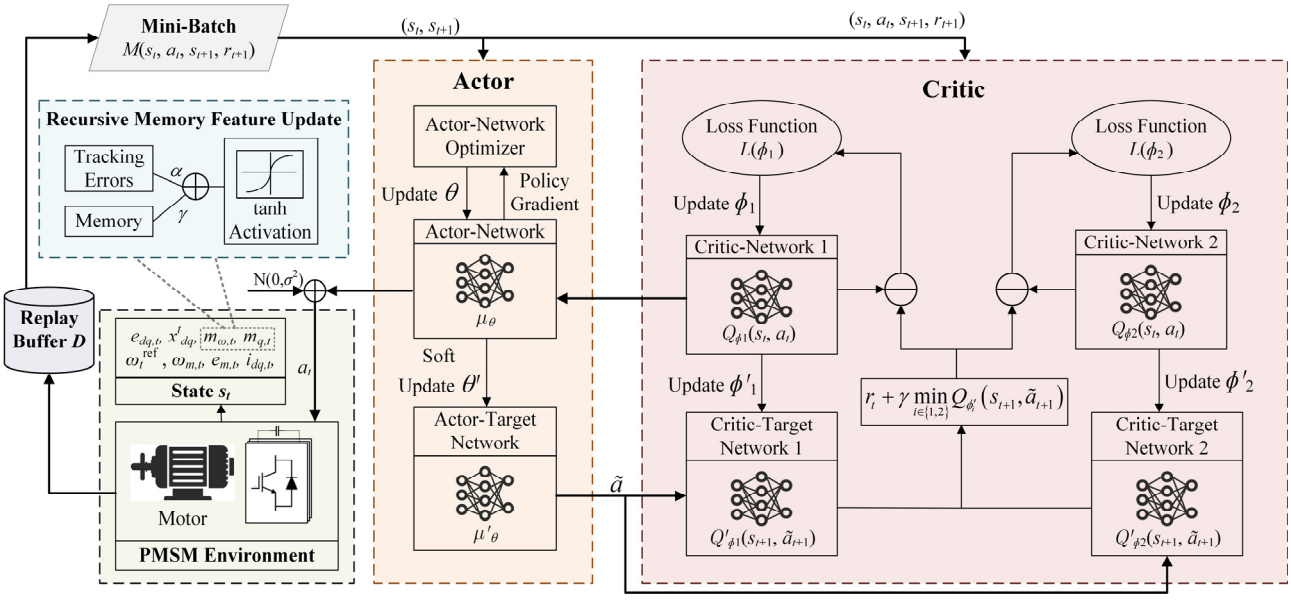

Figure 2. Framework of the proposed MAG-TD3 algorithm. Colored regions distinguish the actor, critic, memory-update, and PMSM-environment components; arrows show the data flow.

3.1. Preliminaries of DRL

Deep reinforcement learning (DRL) provides a data-driven framework for sequential decision-making problems. At each discrete control instant $t = 0, 1, 2, \ldots$, the agent observes

the current state $s_t \in \mathcal{S}$, selects an action $a_t \in \mathcal{A}$, receives a reward r_t, and then the environment evolves to the next state s_{t+1}. The environment under consideration consists of the inverter, motor, coordinate transformation modules, and the cascaded FOC controllers.

The PMSM control task is described as an MDP, defined by the tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P}$ is the transition probability, $r : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is the reward function, and $\gamma \in [0, 1]$ is the discount factor. The state transition is characterized by $s_{t+1} \sim \mathcal{P}(\cdot \mid s_t, a_t)$. The discount factor determines the relative importance of future rewards compared with immediate rewards. The policy learned by the agent is represented as $a_t = \mu_\theta(s_t)$, where μ_θ is a deterministic actor network parameterized by θ. The objective is to maximize the expected discounted return, expressed as $J(\mu_\theta) = \mathbb{E}_{\mu_\theta}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$. In the proposed control problem, the return is determined by current tracking accuracy, voltage command, torque smoothness, adaptive gain behavior, and the satisfaction of the safety constraint.

Unlike reward designs based solely on tracking errors, this work incorporates electromechanical coupling information into the learning process. The state includes recurrent memory features of the speed error e_ω and q-axis current error e_q, while the reward includes torque variation ΔT_e, adaptive gain penalty, and a safety penalty. These designs are detailed in the MDP formulation.

3.2. Formulation of System Dynamics as an MDP

State: To characterize the short-term dynamic trend of the PMSM drive system, the observation state of the agent is constructed using normalized speed variables, current feedback, tracking errors, integral states, and memory features.

$$s_t = \left[\overline{\omega}_t^{\text{ref}}, \overline{\omega}_{m,t}, \overline{e}_{\omega,t}, \overline{i}_{dq,t}, \overline{e}_{dq,t}, x_t^{dq}, m_{\omega,t}, m_{q,t}\right] \tag{22}$$

where $\overline{\omega}_t^{\text{ref}}$ and $\overline{\omega}_{m,t}$ denote the normalized reference and normalized measured mechanical speeds, respectively, and $\overline{i}_{dq,t}$ is the normalized d-q axis current feedback. Additionally, x_t^{dq} signifies the integral state of the current tracking error, which is introduced to provide accumulated error information and improve steady-state error suppression. The update rule is given by $x_t^\delta = x_{t-1}^\delta + T_s e_t^\delta, \delta \in \{d, q\}$.

To capture the short-term evolution of tracking errors, lightweight recursive memory features $m_{\omega,t}$ and $m_{q,t}$ are introduced. Inspired by the recurrent structure of recurrent neural networks (RNNs), each memory feature is updated using the present error, historical errors, and previous memory states. The tanh function is employed to ensure that the memory output remains bounded. The general form is expressed as follows:

$$m_{\delta,t} = \tanh\left(\sum_{j=0}^{n_e} \alpha_{\delta,j} \overline{e}_{\delta,t-j} + \sum_{j=1}^{n_m} \gamma_{\delta,j} m_{\delta,t-j}\right), \delta \in \{\omega, q\} \tag{23}$$

where $\alpha_{\delta,j}$ and $\gamma_{\delta,j}$ represent the weighting coefficients of the error terms and memory terms, respectively. n_e and n_m denote the orders of the historical error terms and historical memory terms. A finite-order recursive form is adopted to enhance dynamic representation capability while maintaining real-time computational efficiency.

Action: Based on the gain-adaptive cascaded FOC structure described in Section 2.2, the TD3 agent generates three normalized continuous actions to simultaneously regulate the d-q axis voltage control variables and the adaptive proportional gain factor of the speed loop. The action vector is defined as follows:

$$a_t = \left[\overline{v}_{d,t},\ \overline{v}_{q,t},\ a_t^\beta\right] \tag{24}$$

where $\overline{v}_{d,t}$ and $\overline{v}_{q,t}$ are utilized as the normalized *d*-*q* axis voltage commands, while a_t^{β} is linearly scaled to generate the adaptive speed-loop gain factor $\beta_{p,t}$ as follows:

$$\beta_{p,t} = \beta_{\min} + \frac{a_t^{\beta} + 1}{2}(\beta_{\max} - \beta_{\min}) \tag{25}$$

where the third action a_t^{β} is linearly mapped to the adaptive gain factor $\beta_{p,t}$ according to Equation (25). To guarantee safe control execution, the mapped factor is projected onto the predefined interval $[\beta_{\min}, \beta_{\max}]$, yielding $\overline{\beta}_{p,t}$. The bounded factor is then applied to Equation (14) to determine the effective proportional gain in the subsequent control update.

Accordingly, the agent regulates the current-loop voltage control variables through $\overline{v}_{d,t}$ and $\overline{v}_{q,t}$, and adjusts the speed-loop proportional gain online through a_t^{β}. This action design enables coordinated optimization of speed tracking, current regulation, and electromagnetic torque response. During action execution, the voltage magnitude and the adaptive gain are subject to the feasibility constraints given in Equations (10) and (15), respectively.

Reward: The immediate reward is defined as the negative value of a weighted operating cost based on the control variables defined in Section 2.3. This formulation is designed to enhance tracking performance while mitigating torque ripple, excessive voltage commands, and aggressive gain adaptation. The reward function is expressed as follows:

$$r_t = -\left(\lambda_e \left\|\overline{e}_{dq,t}\right\|^2 + \lambda_T \Delta \overline{T}_{e,t}^2 + \lambda_v \left\|\overline{v}_{dq,t}\right\|^2 + \lambda_\beta \overline{\beta}_{p,t}^2\right) - \rho_{s,t} \tag{26}$$

where $\lambda_e, \lambda_T, \lambda_v, \lambda_\beta \geq 0$ denote the corresponding weighting coefficients. The safety penalty $\rho_s(k)$ is introduced to enforce physical operating limits. It is set to zero when all constraints are satisfied; otherwise, a fixed penalty is imposed. Therefore, the proposed reward function jointly optimizes current tracking accuracy, torque smoothness, control effort, and gain adaptation while ensuring safe operation.

3.3. TD3 Algorithm

The TD3 algorithm is adopted to learn the continuous control policy for the PMSM drive system. It is suitable for the proposed framework since the control action is continuous. Compared with DDPG, TD3 improves the learning stability by incorporating clipped double Q-learning, target policy smoothing, and delayed policy updates. The TD3 agent consists of an actor network μ_θ, two critic networks Q_{ϕ_1} and Q_{ϕ_2}, and their corresponding target networks $\mu_{\theta'}$, $Q_{\phi_1'}$ and $Q_{\phi_2'}$, respectively. During the training process, the transition tuple (s_t, a_t, r_t, s_{t+1}) is stored in the replay buffer $\mathcal{D}$. For target policy smoothing, clipped Gaussian noise is added to the target action as

$$\widetilde{a}_{t+1} = \mathrm{clip}[\mu_{\theta'}(s_{t+1}) + \widetilde{\varepsilon}, a_{\min}, a_{\max}],\ \widetilde{\varepsilon} \sim \mathrm{clip}\left(N\left(0, \sigma^2\right), -c, c\right) \tag{27}$$

where $[a_{\min}, a_{\max}]$ denotes the admissible action range.

The Bellman target is determined by the smaller value estimated by the two target critics:

$$y_t = r_t + \gamma \min_{i \in \{1,2\}} Q_{\phi'_i}(s_{t+1}, \widetilde{a}_{t+1}) \tag{28}$$

The critic networks are updated by minimizing the Bellman error. The gradient of the critic loss is given by

$$\nabla_{\phi_i} L(\phi_i) = \mathbb{E}_{(s_t, a_t, r_t, s_{t+1}) \sim \mathcal{D}}\left[2\left(Q_{\phi_i}(s_t, a_t) - y_t\right)\nabla_{\phi_i} Q_{\phi_i}(s_t, a_t)\right],\ i \in \{1, 2\} \tag{29}$$

Based on this gradient, the parameters of the two critic networks are updated by gradient descent. The actor network is updated with a lower frequency than the critic networks. For the deterministic policy network in the actor, the values are optimized by the sampled policy gradient, which can be expressed as follows:

$$\nabla_\theta J \approx \mathbb{E}_{s_t \sim \mathcal{D}} \left[\nabla_a Q_\phi(s_t, a) \Big|_{a=\mu_\theta(s_t)} \nabla_\theta \mu_\theta(s_t) \right] \tag{30}$$

The values of the target network are gently updated, which enhances the stabilization of the learning process.

$$\phi_i' \leftarrow \tau\phi_i + (1-\tau)\phi_i',\ i \in \{1,2\},\ \theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{31}$$

where τ is the soft update coefficient.

Through the above update process, the learned TD3 policy generates voltage commands and the adaptive gain factors for coordinated PMSM control.

To promote stable learning and control execution, TD3 employs clipped double-Q learning, delayed policy updates, target policy smoothing, and soft target-network updates. The temporal features are constrained by the tanh(·) function in Equation (23), while the voltage commands and adaptive gain are constrained by Equations (10) and (15), respectively. In combination with the reward penalties on voltage utilization, torque variation, and gain adjustment, these mechanisms ensure the maintenance of bounded control actions and stable, closed-loop behavior.

4. Evaluation and Results

In this section, the effectiveness of the proposed RL-based control strategy is evaluated in a MATLAB/Simulink PMSM drive simulation environment (MATLAB R2025a). The validation focuses on three fundamental aspects: convergence behavior during training, the steady-state regulation performance, and the dynamic response to reference variations and load disturbances. To ensure a fair comparison, all controllers are tested using the same motor parameters, inverter constraints, sampling period, and operating conditions. The conventional PI controller and the standard TD3-based controller are selected as the baseline methods, and the PMSM parameters are summarized in Table 1. In addition to the speed-step test, a combined transient scenario involving motion reversal and load torque variation is also included to further evaluate the dynamic adaptability and disturbance rejection capability of the proposed controller.

Table 1. Main parameters of the system.

Parameter	Description	Value
R_s	Stator resistance	0.2930
L_d, L_q	dq-axis component of inductance	8.77×10^{-5} and 7.77×10^{-5}
ψ_f	Permanent flux	0.0046
p	Number of pole pairs	7
ω_b	Nominal mechanical velocity	$3476 \times \pi/30$
I_b, V_b	Nominal current and voltage	21.43 and 13.86
K_p^ω	Proportional gain of the speed loop	2.2992
K_i^ω	Integration coefficient of the speed loop	0.0304

4.1. Training Convergence

During the training stage, the TD3 agent was trained in a simulation environment implemented in MATLAB/Simulink. Subsequent to each control cycle, the agent received the reward calculated based on the system performance and updated its policy network

using the TD3 algorithm in Table 2. These parameters are selected to ensure stable learning and efficient convergence of the actor–critic networks.

Table 2. The hyperparameter settings for the training process.

Parameter	Value
Agent sample time	0.0002
Actor DNN architecture	[64, 32]
Critic DNN architecture	[64, 32]
Replay buffer length	100,000
Mini-batch size	512
Actor learning rate	0.001
Critic learning rate	0.001
Discount factor	0.995

The convergence behavior of the learning process is illustrated in Figure 3. During the initial training stage, the accumulated reward fluctuates because the agent explores a wide range of voltage commands and adaptive-gain factors. As training progresses, the reward gradually increases and then approaches a stable level, indicating that the agent has learned an effective policy for coordinating current regulation, speed tracking, and torque-related behavior.

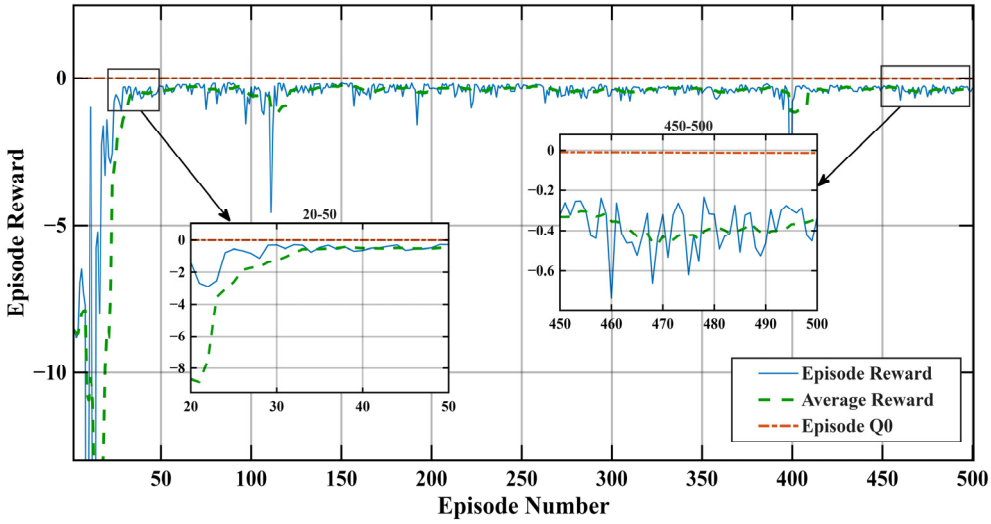

Figure 3. Training process of the proposed MAG-TD3 controller.

The proposed method demonstrates a stable convergence tendency, with enhancement primarily attributed to two factors. Firstly, the memory-enhanced observation provides additional information about the short-term evolution of key tracking errors, which helps reduce ambiguous state-action mapping during transient processes. Secondly, the adaptive-gain action enables the learning agent to regulate the speed-loop response with greater flexibility, rather than relying exclusively on voltage command adjustments.

4.2. Steady-State Performance Evaluation

To evaluate the steady-state performance of the proposed RL-based method, tests were conducted at 80% of the rated speed. Figure 4 compares the speed, *d*- and *q*-axis currents, and electromagnetic torque obtained with the PI, TD3, and proposed MAG-TD3 controllers. The speed and current responses are shown over the full simulation interval, while the torque inset enlarges the steady-state interval from 1.50 s to 1.52 s to make the relatively small torque ripple visible.

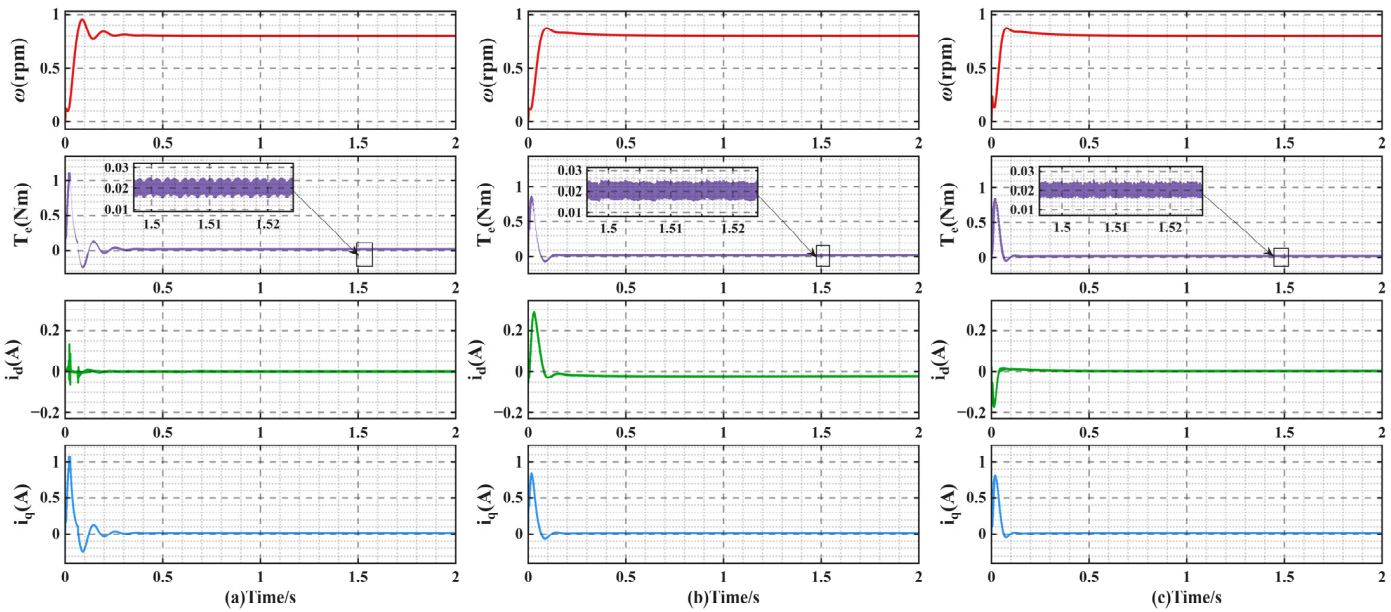

Figure 4. Comparison of speed, current, and electromagnetic torque responses at 80% of the rated speed: (**a**) conventional PI controller, (**b**) TD3-based controller, and (**c**) proposed MAG-TD3 controller. The torque inset shows the steady-state interval from 1.50 s to 1.52 s.

The conventional PI controller maintains stable operation under the nominal steady-state condition, but its performance is constrained by fixed control gains. The standard TD3 controller improves the tracking behavior through data-driven voltage command optimization; however, residual oscillations can still be observed in the current and torque responses. In contrast, the proposed MAG-TD3 controller achieves more accurate speed regulation and smoother current responses. The *d*-axis current remains close to its reference value, while the *q*-axis current supplies the required torque with reduced fluctuation. The enlarged torque waveforms further demonstrate a reduction in torque ripple, suggesting that the proposed controller enhances steady-state regulation quality while preserving the cascaded FOC structure.

4.3. Dynamic Performance

To further verify the dynamic adaptability of the proposed controller, a dynamic test was performed at the rotor speed increasing from 20% to 60%. During the transient process, the TD3 agent dynamically adjusted at the instant of speed variation to maintain optimal control performance. The performance variables and adaptive-gain behavior are demonstrated in Figures 5 and 6. To quantitatively evaluate the dynamic performance of the three control methods, the main transient response indices are summarized in Table 3. RMSE denotes root mean square error, MAE denotes mean absolute error, and RMS denotes root mean square value. These indices reflect the response speed, speed tracking accuracy, current smoothness, and torque fluctuation during the speed transition.

When the speed reference changes, all controllers ultimately track the command, but their transient behaviors differ. The PI controller responds rapidly but suffers from a larger overshoot and longer settling process. The standard TD3 controller enhances the transient response by leveraging voltage commands; however, it still generates observable current and torque oscillations during recovery.

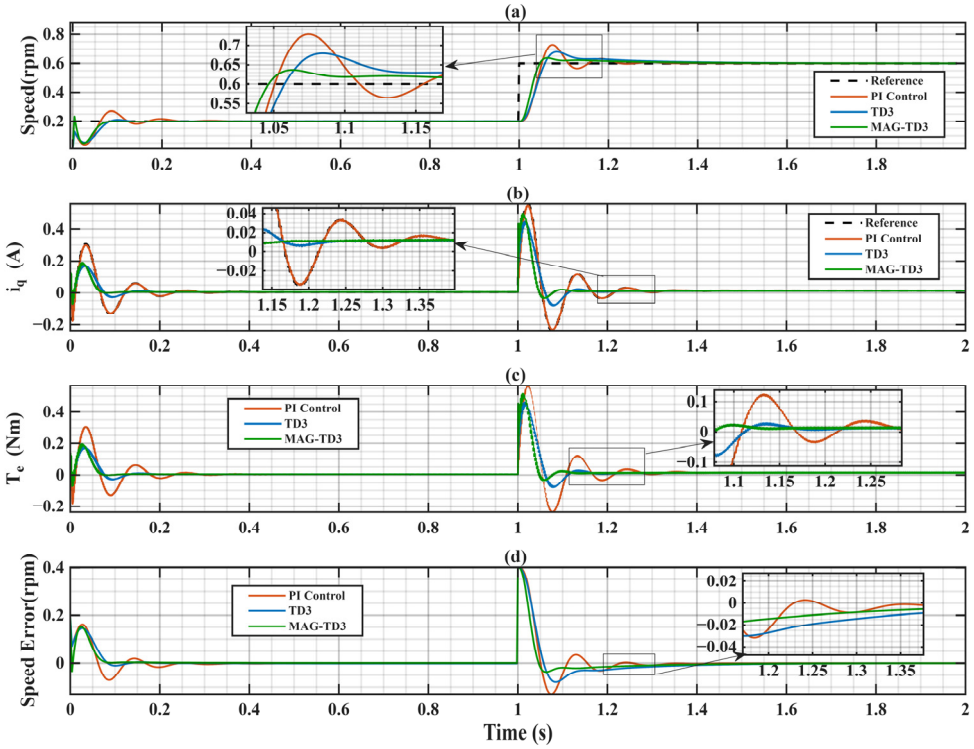

Figure 5. Dynamic performance from 20% to 60% rated rotor speed. (**a**) Rotor speed. (**b**) *q*-axis current. (**c**) Electromagnetic torque. (**d**) Speed tracking error.

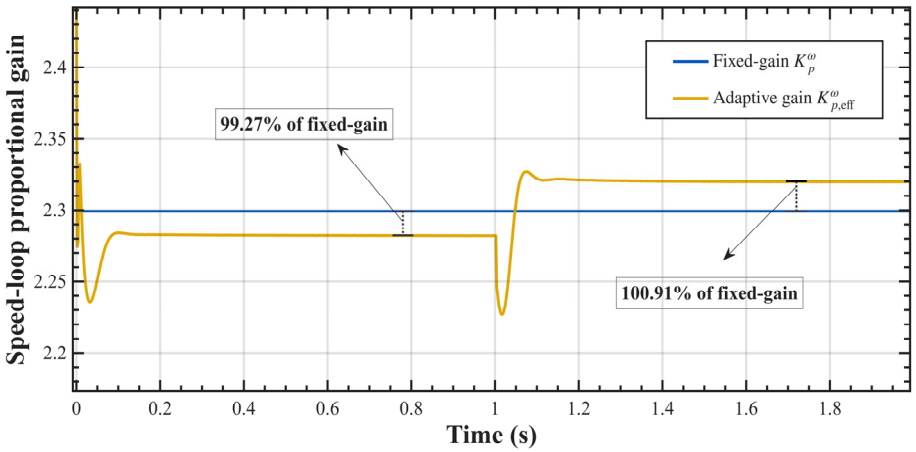

Figure 6. Evolution of the adaptive gain coefficient generated by the proposed controller.

Table 3. Dynamic performance comparison of different control methods.

Parameter	Conventional PI	TD3	Proposed MAG-TD3
Risetime (s)	0.0291	0.0359	0.0285
Settlingtime (s)	0.2350	0.3305	0.2125
Overshoot (%)	32.120	19.819	9.194
Speed RMSE (rpm)	0.0549	0.0531	0.0451
Speed MAE (rpm)	0.0330	0.0332	0.0238
I_q RMS (A)	0.0812	0.0584	0.0562
T_e RMS (Nm)	0.0838	0.0602	0.0581

In contrast, the proposed MAG-TD3 controller achieves faster speed recovery with smaller tracking errors and smoother current regulation. As demonstrated in Table 3, the proposed MAG-TD3 controller mitigates the overshoot to 9.194%, while attaining the

lowest speed RMSE and MAE of 0.0451 rpm and 0.0238 rpm, respectively. Furthermore, the RMS values of iq and Te are decreased to 0.0562 A and 0.0581 N·m, respectively, indicating smoother torque-producing current regulation and lower torque fluctuation. This enhancement is attributed not only to the adaptive voltage commands and gain adjustment, but also to the introduced lightweight memory features, which encode short-term historical tracking errors and previous memory states. By providing the agent with dynamic trend information beyond the instantaneous observation, the memory module improves the coordination between speed regulation, current dynamics, and torque generation. As shown in Figure 5, the q-axis current is adjusted promptly to generate the required torque, while the d-axis current maintains effective regulation around its reference.

The torque response further validates the effectiveness of the proposed reward design. Since both speed tracking accuracy and torque variation are considered, the controller avoids excessive compensation and suppresses unnecessary torque ripple. Consequently, the MAG-TD3 model successfully achieves a balanced compromise between rapid dynamic response and smooth torque production.

The fast transient response is supported by temporal features that provide information on recent error evolution, and by bounded adaptive gain, which adjusts the responsiveness of the speed loop during changes in operating conditions. As demonstrated in Table 3, the MAG-TD3 achieves a settling time of 0.2125 s with reduced overshoot and speed tracking errors, demonstrating its effective convergence performance during the transient process.

Figure 6 illustrates the evolution of the effective speed-loop proportional gain generated by MAG-TD3. In contrast to the fixed-gain PI scheme, the gain is adjusted according to the operating conditions. The adaptive factor is constrained within the predefined interval in Equation (15); consequently, the effective gain remains within the corresponding admissible range. These limits are established in the controller design to prevent excessive gain amplification, current oscillation, and excessive voltage demand.

As demonstrated in Figure 6, the effective gain remains relatively stable, indicating that the learned policy successfully avoids unnecessary variations in gain near steady state. During the reference transition, the gain is adjusted within the prescribed interval to improve the speed-loop response. While a wider interval may be considered, its selection depends on the operating constraints, the desired transient response, and the overall control objectives.

To further verify the dynamic adaptability of the proposed controller under more complex transient conditions, an additional test involving motion reversal and load torque variation was conducted. In this test, the reference speed changes from 0.2 rpm to -0.2 rpm at 1 s, while the load torque increases from 0.3 N·m to 0.6 N·m at 2 s. Therefore, this case simultaneously evaluates the speed reversal tracking capability and the disturbance rejection performance under load torque variation.

As shown in Figure 7, the conventional PI controller can eventually track the reversed speed reference, but it produces relatively large oscillations during the reversal process. Additionally, noticeable transient fluctuations appear in the speed and torque responses when the load torque increases. The standard TD3 controller improves the transient response compared with the PI controller, but residual speed deviation and electromagnetic torque fluctuation are still observed. In contrast, the proposed MAG-TD3 controller exhibits a smoother speed reversal process, smaller transient deviation, and more stable electromagnetic torque response.

These results demonstrate that the proposed MAG-TD3 controller is not limited to a single speed step. Due to its temporal error evolution features and bounded adaptive speed-loop gain regulation, the proposed controller improves transient tracking and disturbance rejection under complex dynamic conditions, including motion reversal and load torque variation.

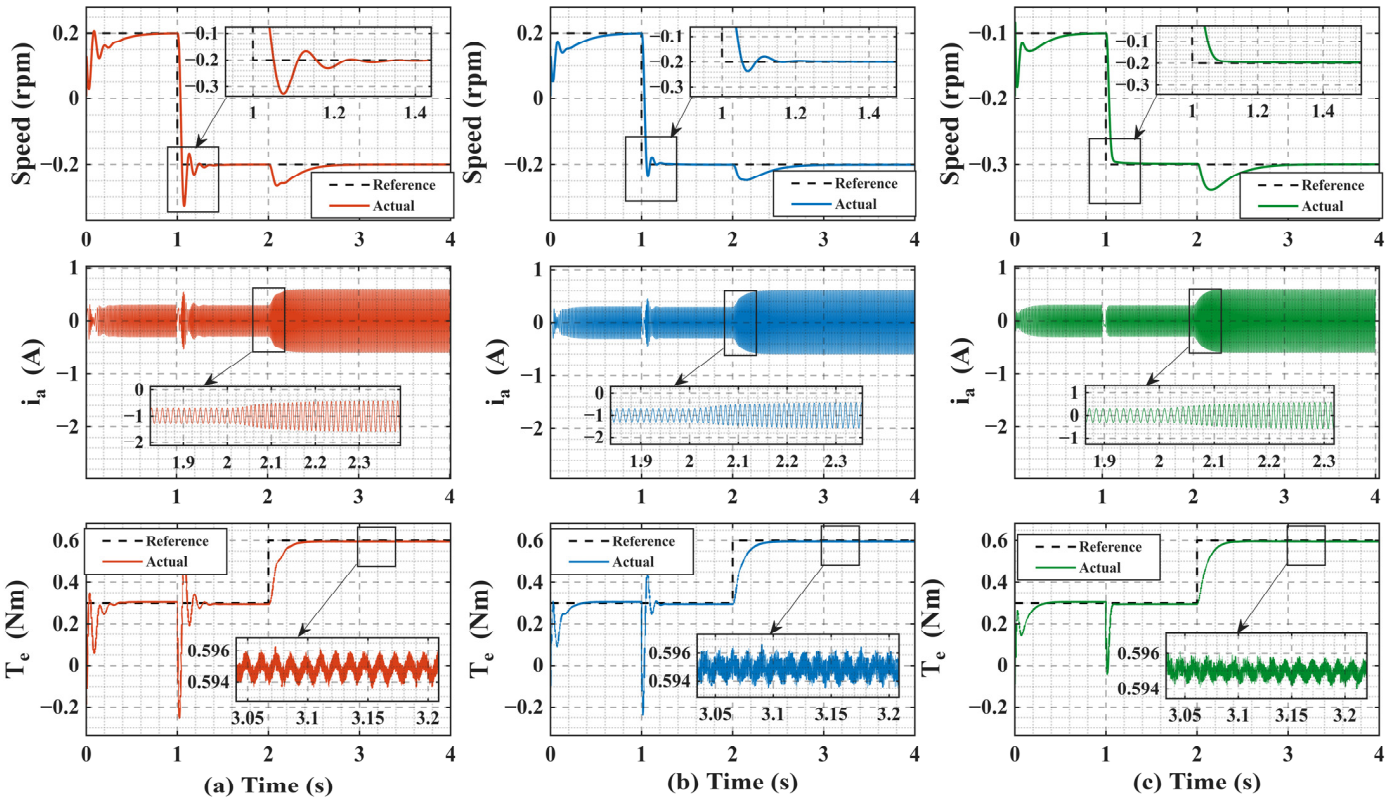

Figure 7. Dynamic performance under motion reversal and load torque variation. The reference speed changes from 0.2 rpm to −0.2 rpm at 1 s, and the load torque increases from 0.3 N·m to 0.6 N·m at 2 s. The speed, A-phase current, and electromagnetic torque responses are compared among: (**a**) conventional PI controller, (**b**) TD3-based controller, and (**c**) proposed MAG-TD3 controller.

5. Conclusions

This paper proposes a temporal-feature-enhanced reinforcement learning control approach for PMSM drives with adaptive speed-loop gain regulation. The proposed framework integrates the learning agent into a dual-loop architecture, enabling the controller to coordinate voltage command generation and online adjustment of the speed-loop gain. The temporal feature representation provides the agent with compact information on recent error evolution, while the multiobjective learning criterion guides the policy toward a balanced tradeoff among tracking performance, torque ripple suppression, and gain boundedness.

Simulation comparisons with conventional PI control and standard TD3 verify the advantages of the proposed scheme in terms of transient response, speed tracking accuracy, smoother torque behavior, and disturbance rejection under speed-step, motion reversal, and load torque variation conditions. These results demonstrate the effectiveness of temporal feature enhancement and adaptive gain regulation in improving the dynamic awareness of RL controllers and promoting coordinated regulation in PMSM drives. Future work will focus on experimentally validating and evaluating the robustness over a wider range of operating conditions and PMSM configurations. Additionally, further efforts will investigate theoretical analyses of convergence properties and real-time implementation on embedded motor-drive platforms.

Author Contributions: Conceptualization, investigation, H.Z.; data curation, supervision, J.C.; methodology, writing-original draft, writing-review & editing, software, validation, Z.W.; validation, formal analysis, A.S.; supervision, conceptualization, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Hongquan Zhang and Jingjun Cui were employed by the company COSMOPlat AIoT Technology. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviations

PMSM	Permanent magnet synchronous motor
FOC	Field-oriented control
PI	Proportional–integral
RL	Reinforcement learning
DRL	Deep reinforcement learning
DDPG	Deep deterministic policy gradient
TD3	Twin delayed deep deterministic policy gradient
MAG-TD3	Memory-augmented adaptive-gain TD3
MDP	Markov decision process
DNN	Deep neural network
RNN	Recurrent neural network
LSTM	Long short-term memory
VSI	Voltage-source inverter
RMSE	Root mean square error
MAE	Mean absolute error
RMS	Root mean square

Symbols

a_t	Action selected by the agent at time step t
a_t^{β}	Third action used to adjust the speed-loop proportional gain
$\mathcal{A}$	Action space
B	Viscous friction coefficient
B_a	Equivalent damping coefficient used in the speed-loop controller
e_d, e_q	d- and q-axis current tracking errors
e_ω	Rotor-speed tracking error
i_d, i_q	Measured d- and q-axis stator currents
i_d^{ref},i_q^{ref}	Reference d- and q-axis currents
J	Rotor moment of inertia; expected discounted return when used as $J(\mu_\theta)$
K_p^ω,K_i^ω	Proportional and integral gains of the outer speed PI controller
K_t	Torque constant
$m_{\omega,t}$,$m_{q,t}$	Recursive memory features for speed and q-axis current errors
n_e, n_m	Orders of historical error and memory terms, respectively
$\mathcal{P}$	State-transition probability function
r_t	Immediate reward at time step t
$\mathcal{S}$	State space
s_t	Observation state at time step t
T_e	Electromagnetic torque
ΔT_e	Electromagnetic torque variation
T_s	Control sampling period
u_d, u_q	Physical d- and q-axis voltage commands
v_d, v_q	Constrained d- and q-axis voltage commands applied to the modulation stage
$\overline{v}_d$,$\overline{v}_q$	Normalized d- and q-axis voltage actions generated by the agent
x_t^d,x_t^q	Integral states of the d- and q-axis current tracking errors
$\alpha_{\delta,j}$	Weighting coefficient of the historical error term in the memory update

γ	Discount factor in the MDP
$\gamma_{\delta,j}$	Weighting coefficient of the historical memory term in the memory update
δ	Axis index, $\delta \in \{d, q\}$
q	Parameter vector of the actor network
$\lambda_e, \lambda_T, \lambda_v, \lambda_\beta$	Weighting coefficients in the reward function
μ_θ	Deterministic policy parameterized by θ
ω_m	Measured mechanical rotor speed
ω_{ref}	Reference mechanical speed
ω_b	Desired speed-loop bandwidth
$\beta_{p,t}$	Adaptive factor for the outer speed-loop proportional gain
$\overline{\beta}_{p,t}$	Projected adaptive gain factor satisfying the prescribed bounds
$\beta_{\min}, \beta_{\max}$	Lower and upper bounds of the adaptive gain factor
$\rho_s(k)$	Safety penalty at control step k

References

1. Rodriguez, J.; Garcia, C.; Mora, A.; Flores-Bahamonde, F.; Acuna, P.; Novak, M.; Zhang, Y.; Tarisciotti, L.; Davari, S.A.; Zhang, Z.; et al. Latest advances of model predictive control in electrical drives—Part I: Basic concepts and advanced strategies. *IEEE Trans. Power Electron.* **2022**, *37*, 3927–3942. [CrossRef]
2. Rodriguez, J.; Garcia, C.; Mora, A.; Davari, S.A.; Rodas, J.; Valencia, D.F.; Elmorshedy, M.; Wang, F.; Zuo, K.; Tarisciotti, L.; et al. Latest advances of model predictive control in electrical drives—Part II: Applications and benchmarking with classical control methods. *IEEE Trans. Power Electron.* **2022**, *37*, 5047–5061. [CrossRef]
3. Nguyen, T.H.; Trinh, H.M.; Nguyen, T.T.; Jeon, J.W. Antiwindup composite adaptive speed control for permanent-magnet synchronous motors. *IEEE Trans. Power Electron.* **2026**, *41*, 9162–9179. [CrossRef]
4. Raj, R.; Agarwal, P.; Das, S. Adaptive fuzzy logic-based SMO utilizing higher order soft-sign function with ROA tuned PI gains for sensorless PMSM drive. *IEEE Trans. Ind. Electron.* **2026**, *73*, 9775–9786. [CrossRef]
5. Song, W.; Li, J.; Ma, C.; Xia, Y.; Yu, B. A simple tuning method of PI regulators in FOC for PMSM drives based on deadbeat predictive conception. *IEEE Trans. Transp. Electrif.* **2024**, *10*, 9852–9863. [CrossRef]
6. Zhang, L.; Tao, R.; Li, X. Model-free adaptive second-order sliding mode control for PMSM in electric traction systems based on terminal attractor. *IEEE Trans. Power Electron.* **2026**, *41*, 11973–11984. [CrossRef]
7. Wang, S.; Wang, Z.; Niu, X.; Zhu, Y.; Lu, W.; Liu, Y. An adaptive sliding mode control for PMSM servo systems beyond the inertia ratio limit. *IEEE Trans. Power Electron.* **2025**, *40*, 16352–16366. [CrossRef]
8. Luo, Y.; Luo, J.; Yang, K.; Yu, J. Adaptive moment estimation-based model predictive control for PMSM motor with low tracking error. *IEEE Trans. Ind. Electron.* **2026**, *73*, 230–241. [CrossRef]
9. Kumar, P.D.; Ramesh, T.; Ramakrishna, P.; Tangirala, I. Five-level inverter fed sensorless PMSM drive using modified MPTC with novel reduced search space algorithm. *IEEE Trans. Ind. Appl.* **2023**, *59*, 6826–6836. [CrossRef]
10. Mastanaiah, A.; Ramesh, T.; Naresh, S.V.K.; Bonthagorla, P.K. Deep reinforcement learning agent based speed controller for DTC-SVM of PMSM drive. *IET Power Electron.* **2025**, *18*, e70130. [CrossRef]
11. Liu, X.; Qiu, L.; Fang, Y.; Wang, K.; Li, Y.; Rodríguez, J. Combining data-driven and event-driven for online learning predictive control in power converters. *IEEE Trans. Power Electron.* **2025**, *40*, 563–573. [CrossRef]
12. Huang, Z.; Gong, J.; Xiao, X.; Gao, Y.; Xia, Y.; Wheeler, P.; Ji, B. Artificial intelligence and digital twin technologies for power converter control in transportation applications: A review. *IET Power Electron.* **2025**, *18*, e70013. [CrossRef]
13. Jakobeit, D.; Schenke, M.; Wallscheid, O. Universal direct torque controller for permanent magnet synchronous motors via meta-reinforcement learning. *IEEE Trans. Power Electron.* **2026**, *41*, 7394–7406. [CrossRef]
14. Farah, N.; Lei, G.; Zhu, J.; Guo, Y. Robust model-free reinforcement learning based current control of PMSM drives. *IEEE Trans. Transp. Electrif.* **2025**, *11*, 1061–1076. [CrossRef]
15. Zhao, J.; Cai, T.; Xiong, M.; Yang, C. Reinforcement learning and singular perturbation-based optimal speed synchronous control of a flexible coupling dual-PMSM system. *IEEE Trans. Ind. Inform.* **2025**, *21*, 9757–9766. [CrossRef]
16. Schenke, M.; Wallscheid, O. A deep Q-learning direct torque controller for permanent magnet synchronous motors. *IEEE Open J. Ind. Electron. Soc.* **2021**, *2*, 388–400. [CrossRef]
17. Bhattacharjee, S.; Halder, S.; Yan, Y.; Balamurali, A.; Iyer, L.V.; Kar, N.C. Real-time SIL validation of a novel PMSM control based on deep deterministic policy gradient scheme for electrified vehicles. *IEEE Trans. Power Electron.* **2022**, *37*, 9000–9011. [CrossRef]
18. Wang, Y.; Fang, S.; Hu, J. Active disturbance rejection control based on deep reinforcement learning of PMSM for more electric aircraft. *IEEE Trans. Power Electron.* **2023**, *38*, 406–416. [CrossRef]

19. Bui, M.X.; Nguyen, H.-N.; Nutkani, I.; Fernando, N.; Dutta, R.; Rahman, M.F. Reinforcement learning-based online estimation of the inductances of the permanent magnet synchronous machines. *IEEE Trans. Instrum. Meas.* **2025**, *74*, 3003212. [CrossRef]
20. Wang, Y.; Fang, S.; Hu, J.; Huang, D. A novel active disturbance rejection control of PMSM based on deep reinforcement learning for more electric aircraft. *IEEE Trans. Energy Convers.* **2023**, *38*, 1461–1470. [CrossRef]
21. Schenke, M.; Kirchgässner, W.; Wallscheid, O. Controller design for electrical drives by deep reinforcement learning: A proof of concept. *IEEE Trans. Ind. Inform.* **2020**, *16*, 4650–4658. [CrossRef]
22. Xie, H.; Ye, X.; Zhuang, Z.; Wei, Y.; Novak, M.; Wang, F. An improved reinforcement learning approach for cost function optimization in multiobjective FCS-MPC of PMSM drives. *IEEE Trans. Ind. Electron.* **2026**, *early access*. [CrossRef]
23. Chen, B.; Yang, J.; Li, H.; Yang, L.; Zhao, H. Fault detection and identification of PMSM current sensor based on attention mechanism and bidirectional long short-term memory network. *IEEE J. Emerg. Sel. Top. Power Electron.* **2025**, *13*, 6494–6506. [CrossRef]