An Improved Model-Free Active Disturbance Rejection Deadbeat Predictive Current Control Method of PMSM Based on Data-Driven

Yicheng Wang, Shuhua Fang, *Senior Member, IEEE*, and Demin Huang

***Abstract*—This article proposes an improved model-free active disturbance rejection deadbeat predictive current control(ADRDPCC) method for permanent magnet synchronous motor (PMSM) used in more electric aircraft (MEA) based on the data driven method, which is used to solve the parameter mismatch problem of deadbeat predictive current control (DPCC) and improve the performance of PMSM control system. DPCC model applied to the current loop of the MEA motor is established as the main control strategy of the system. The principle of active disturbance rejection control is combined with DPCC, and the ADRDPCC structure is formed to optimize the control strategy. A specific extended state observer (ESO) of ADRDPCC is designed to track the internal disturbance caused by parameter mismatch and the external disturbance in real time. DPCC is used as the control law of the ADRDPCC structure to predict the current and output the reference voltage based on the observation of ESO. A deep reinforcement learning model based on ADRDPCC is designed and trained based on the data-driven method. The trained model can compensate and optimize ADRDPCC based on the disturbance observed by ESO and the observed control state of PMSM. The simulated and experimental results show the superiority of the proposed method.**

***Index Terms*—Active disturbance rejection control, deadbeat predictive current control (DPCC), deep reinforcement learning (DRL), more electric aircraft (MEA), permanent magnet synchronous motor (PMSM).**

I. Introduction

MORE electric aircraft (MEA) has become the main trend of the modern aircraft industry due to the use of clean energy and the reduction of carbon emissions [1]. Permanent magnet synchronous motor (PMSM) has become one of the best choices for MEA motors due to its simple structure and high power density [2], [3]. Nowadays, feedforward flux weakening (FFW) control is becoming one of the popular control methods for MEA motors [4], [5].

Manuscript received 6 January 2023; revised 6 April 2023; accepted 22 May 2023. Date of publication 25 May 2023; date of current version 21 June 2023. This work was supported in part by the Aeronautical Science Foundation of China under Grant 20160769002 and in part by the Six Talent Peaks Project of Jiangsu Province under Grant GDZB-103. Recommended for publication by Associate Editor R. Kennel. *(Corresponding author: Shuhua Fang.)*

The authors are with the School of Electrical Engineering, Southeast University, Nanjing 210096, China (e-mail: 230218324@seu.edu.cn; shfang@seu.edu.cn; 220202916@seu.edu.cn).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2023.3280013.

Digital Object Identifier 10.1109/TPEL.2023.3280013

In MEA-PMSM-FFW control system, the performance of current control is the key factor to determine the system performance. Current predictive control is one of the common current control methods [6], which has good steady-state and dynamic performance and can control the current with high accuracy in a short time interval [7]. Predictive current control can be divided into two categories: model predictive current control (MPCC) and deadbeat predictive current control (DPCC). MPCC has the advantages of fast dynamic response, nonlinear constraint control, and real-time implementation [8]. A mathematical model self-tuning strategy is proposed in [9], which can correct the stator current model in real-time and improve the prediction accuracy of MPCC. The cost function of MPCC is designed in the form of PI [10], which is able to include the previous tracking error and the subsequent tracking error, and forces the prediction current with high robustness. However, the computational complexity of MPCC is high because all the basic voltages need to be enumerated. In a control cycle, only one voltage vector is selected, so that the current error cannot be further reduced [11]. MPCC is a model based control strategy, which is highly dependent on model parameters and is easily affected by load changes, skin effects, working temperature changes [12], [13].

In contrast, DPCC is used to reduce the control complexity and calculation burden, and the reference voltage vector with zero current error can be obtained with only one calculation [14]. This method predicts voltage reference based on discrete model, and uses space vector PWM to convert corresponding switch configuration [15]. A DPCC with thrust ripple equivalent compensation is used instead of a PI controller in order to improve the dynamic response of the current loop, which can compensate the cogging force [16]. In [17], a multistep DPCC is proposed, which has less computation time and better stability. DPCC has higher accuracy, but it is also troubled by model mismatch [18]. Some methods propose the model-free DPCC which can predict the action of the model based on the observed current difference [19], but the current difference has the problem of stagnation and hysteresis [20]. Although the stagnation can be solved, MPC optimum detection is neglected and the updated information is retrieved at the cost of an increase of the current ripple [21]. Employing observers in the control system is a good method to reduce the sensitivity of parameters [22], [23], but due to the lack of state prediction ability, it is difficult to have a better compensation effect on the control performance [15].

The data-driven method starts from the initial data or observation value, uses heuristic rules to find and establish the relationship between internal features, and then discovers the relevant laws. At present, scholars have applied the method in the field of electric drives. In [24], the practical real-time implementation of data-driven methods are verified, and data-driven paradigm is used as a systematic design tool for PMSM current controllers, and comparisons between the model-based and data-driven controls are provided to verify the advantages. The concepts of iterative learning control and data-driven are incorporated into the MPC framework, and the method has the effect of reducing parameter disturbance [25]. An intelligent gate drive control method based on data-driven is proposed, which can adjust the driver according to real-time operation conditions to improve the operating performance of the converter [26]. In [27], a data-driven adaptive PID control scheme is proposed, and simulation results show that the method can adjust the controller parameters of a quadrotor UAV online.

Machine learning (ML), which is a kind of data-driven artificial intelligence method, can describe unknown models by learning and exploring experiences and cases based on the development of computer technology [28]. Deep reinforcement learning (DRL), which is a kind of stochastic method based on deep neural network, can establish the model between observation value and demand value through maximum likelihood method [29]. At present, DRL has been successfully used in many practical electrical applications, such as the optimal decision-making of battery charging [30], parameter optimization of motor control [31] and so on. Because the characteristic of DRL is to predict the future state based on observation and experience [32], it is suitable to design a DRL model as the model-free improvement and compensation to DPCC in the case of parameter mismatch and external disturbance, which is one of the motivations of the proposed method.

Therefore, the existing unresolved issues are as follows.

1) The previous model-based methods to improve the disturbance rejection performance of DPCC are to make real-time parameters more accurate, such as parameter identification, which does not fundamentally solve the problem of DPCC being vulnerable to disturbance.
2) The previous DPCC using observers compensates the system with coefficient factors based on the observation results, which would be limited by the coefficient factors.
3) The previous model-free DPCC gives different response according to different states after a large number of experiments are summarized, which is close to the table lookup method and has the disadvantages of large calculation amount and inflexibility.

In response to the above issues, the novelty and contributions of this article are as follows.

1) An improved ADRDPCC based on the active disturbance rejection control structure is proposed, which improves the disturbance rejection ability of DPCC structurally, and has simple structure and strong robustness [33].
2) The proposed ADRDPCC improves the DPCC method with observers, designs a specific extended state observer (ESO) for the control environment, and breaks through the limitations of traditional coefficient factor compensation methods.
3) A model-free current prediction compensation mechanism is proposed and trained by the data-driven DRL method, and a gradient descent method is designed to make the DRL training model converge.

Under the premise of ensuring reliable operation, the training mechanism can compensate the control system based on the observations of ESO and the current state of the system flexibly and quickly. The proposed method combines the advantages of the model-based and the model-free DPCC, respectively, solving the problem that the controller is sensitivity to disturbances and parameter mismatches, and improving the control performance of MEA motor effectively.

The model of DPCC controller is built according to the theoretical analysis and is used as the current loop of the flux weakening control system in Section II. An ESO model is designed and the active disturbance rejection DPCC control framework is built in Section III-A. DRL model is constructed based on the design of the interface module, training mechanism and reward mechanism according to Markov decision process (MDP), and used to compensate the predicted current in the form of model-free method in Section III-B, C. The ADRDPCC control system is simulated and the experimental platform is built to verify the effectiveness of the proposed method in Sections IV and V. Finally, Section VI concludes this article.

II. Design of DPCC of Two-Step Delay

The goal of DPCC is to make the current at time k+1 equal to the given current at time k, that is, to achieve zero error tracking of the given current within a control cycle, which can be expressed as

$$i_{dq}(k+1) = i^*_{dq}(k) \tag{1}$$

where i_{dq} is the d–q axis current, and the superscript $*$ represents the given value. The equation of PMSM in d–q axis can be expressed as

$$\begin{cases} \frac{di_d}{dt} = -\frac{R_s i_d}{L_d} + \frac{L_q}{L_d}\omega_e i_q + \frac{u_d}{L_d} \\ \frac{di_q}{dt} = -\frac{L_d}{L_q}\omega_e i_d - \frac{R_s}{L_q} i_q + \frac{u_q}{L_q} - \frac{\omega_e \varphi_f}{L_q} \end{cases} \tag{2}$$

where u_{dq} is the d–q axis voltage, R_s is the stator resistance, L_{dq} is the d–q axis inductance, ω_e is the electrical angular velocity, and φ_f is the flux linkage. Discreting (2) with forward Euler method, one can get

$$\begin{cases} \frac{i_d(k+1)-i_d(k)}{T_s} = -\frac{R_{sz}}{L_{dz}} i_d(k) + \frac{L_{qz}}{L_{dz}}\omega_e(k) i_q(k) + \frac{u^*_d(k)}{L_{dz}} \\ \frac{i_q(k+1)-i_q(k)}{T_s} = -\frac{L_{dz}}{L_{qz}}\omega_e(k) i_d(k) - \frac{R_{sz}}{L_{qz}} i_q(k) \\ \qquad\qquad -\frac{\omega_e(k)\varphi_{fz}}{L_{qz}} + \frac{u^*_q(k)}{L_{qz}} \end{cases} \tag{3}$$

where T_s is the period from k to k+1. Since the internal parameters of the motor will change with time and environment, the subscript z is used to represent the time-varying parameter. The

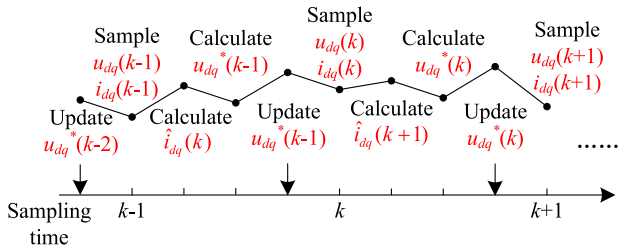


Fig. 1. Operation process of DPCC.

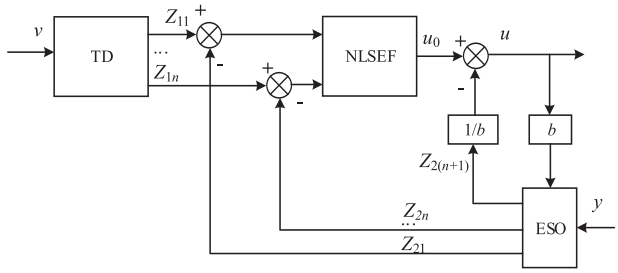


Fig. 2. Control framework of ADRC.

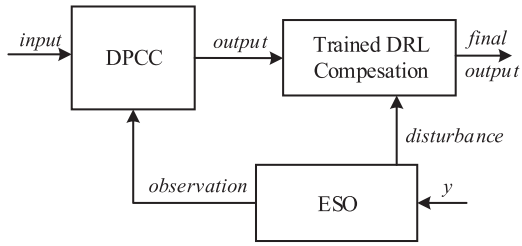


Fig. 3. Control principle of ADRDPCC.

corresponding voltage equation is

$$\begin{cases} u_d^*(k) = R_{sz} i_d(k) + L_{dz} \frac{i_d(k+1)-i_d(k)}{T_s} - \omega_e(k) L_{qz} i_q(k) \\ u_q^*(k) = R_{sz} i_q(k) + L_{qz} \frac{i_q(k+1)-i_q(k)}{T_s} \\ \qquad +\omega_e(k) L_{dz} i_d(k) + \omega_e(k) \varphi_{fz} \end{cases} . \tag{4}$$

Due to the unit calculation delay, the $u_{dq}{}^*(k)$ calculated at the time k will be applied to the motor control system at the time $k+1$. Therefore, the unit calculation delay needs to be compensated. The given voltage at the time $k+1$ is calculated at the time k to correct the control target, so that the current at the time $k+2$ is equal to the current at the time k, which is also called the two-step delayed DPCC. Therefore, the predicted current at time $k+1$ is

$$\begin{cases} \hat{i}_d(k+1) = [-\frac{R_{sz}}{L_{dz}} i_d(k) + \frac{L_{qz}}{L_{dz}} \omega_e(k+1) i_q(k) \\ \qquad + \frac{u_d^*(k-1)}{L_{dz}}] T_s + i_d(k) \\ \hat{i}_q(k+1) = [-\frac{L_{dz}}{L_{qz}} \omega_e(k+1) i_d(k) - \frac{R_{sz}}{L_{qz}} i_q(k) \\ \qquad - \frac{\omega_e(k+1)\varphi_{fz}}{L_{qz}} + \frac{u_q^*(k-1)}{L_{qz}}] T_s + i_q(k) \end{cases} \tag{5}$$

where the superscript ^ represents the estimated value. Since the interval of T_s is very small, $\omega_e(k+1) = \omega_e(k)$. By introducing (5) into (4), the two-step delayed DPCC can be obtained, and the operation process is as shown in Fig. 1.

III. Proposed ADRDPCC Strategy

A. Observer Design Based on ADRC

The principle of classical ADRC can be expressed as

$$\begin{cases} x^{(n)} = f(x, \dot{x}, \ldots, x^{(n-1)}, t) + \omega(t) + bu \\ y = x(t) \end{cases} \tag{6}$$

where $f(x, \ldots, x^{(n-1)}, t)$ is the control system, $\omega(t)$ is the disturbance, $x(t)$ is the output that can be measured, u is the control effect of the system, b is the control effect gain. Specifically, ADRC is composed of three parts to realize the function of (6): tracking differentiator (TD); ESO; and a control law such as nonlinear state error feedback (NLSEF). Fig. 2 shows the control framework of classical ADRC.

TD is used to arrange the transition process. The smoothing process provided by TD may lead to the effect of delayed tracking, so it can also be omitted [34]. ESO is the core component of ADRC, which can observe the output and its differential state. ESO can also estimate the total disturbance of the system under the condition of uncertain model, and the system can be compensated based on the disturbance. The control law is used to control the observed value of ESO and the input value. The design and selection of control law should be based on the requirements of control environment.

The improved ADRDPCC takes ADRC as the basis of the control framework. ESO, the core of ADRC, is used as the observer, and DPCC designed in Section II is used as the control law to improve the current control performance. The training DRL model compensates the output of DPCC based on the observation value of ESO. Fig. 3 shows the control principle of the proposed ADRDPCC method, which is an improved DPCC with a generalized ADRC structure.

The classic ESO formula can be written as

$$\begin{cases} \varepsilon_1 = z_1 - y \\ \dot{z}_1 = z_2 - \beta_1 \varepsilon_1 \\ \cdots\cdots \\ \dot{z}_{n+1} = z_{n+2} - \beta_{n+1} fal(\varepsilon_1, \alpha_n, \delta) + f_0(x_1, x_2, \ldots, x_{n+1}) \\ \qquad + bu \\ \dot{x}_{n+2}(t) = -\beta_{n+2} fal(\varepsilon_1, \alpha_{n+1}, \delta) \\ y = x_1 \end{cases} \tag{7}$$

where ε is the observer error, β is the gain coefficient, $f(x_1, x_2, \ldots, x_{n+1}, t)$ is the total disturbance of internal and external disturbances, $Z_1, \ldots, Z_n$ are the extended state of the control objective after analyzing the output y of the whole system, and fal() is nonlinear function

$$\text{fal}(x, a, \delta) = \begin{cases} \frac{x}{\delta^{1-a}}, & |x| \le \delta \\ \text{sign}(x), & |x| > \delta \end{cases} \tag{8}$$

where x is the error, a is the nonlinear factor, δ is the filter factor, and sign(x) is symbolic function. When $x \ge 0$, sign $= 1$, and when $x < 0$, sign $= -1$.

In order to make full use of ESO, it is necessary to design the observer in combination with the specific control system and control environment. The ESO design process of ADRDPCC

system is described as follows. Let $\Delta u_{dq} = u_{dq}$-$u_{dq}{}^*$, then (2) can be rewritten as

$$\begin{cases} \frac{di_d}{dt} = -\frac{R_s i_d}{L_d} + \frac{L_q}{L_d}\omega_e i_q + \frac{u_d^* + \Delta u_d}{L_d} \\ \frac{di_q}{dt} = -\frac{L_d}{L_q}\omega_e i_d - \frac{R_s}{L_q} i_q + \frac{u_q^* + \Delta u_q}{L_q} - \frac{\omega_e \varphi_f}{L_q} \end{cases}. \tag{9}$$

Since the internal parameters of the motor will change with the working condition and time, (9) can be rewritten as

$$\begin{cases} \frac{di_d}{dt} = \frac{1}{L_{dz}}(-R_{sz} i_d + L_{qz}\omega_e i_q + u_d^*) + \frac{\Delta u_d}{L_d} \\ \quad +(-\frac{R_s}{L_d} + \frac{R_{sz}}{L_{dz}})i_d + (\frac{1}{L_d} - \frac{1}{L_{dz}})u_d^* \\ = \frac{1}{L_{dz}}(-R_{sz} i_d + L_{qz}\omega_e i_q + u_d^*) + Z_d \\ \frac{di_q}{dt} = \frac{1}{L_{qz}}(-\omega_e L_{dz} i_d - R_{sz} i_q + u_q^* - \omega_e \varphi_{fz}) + \frac{\Delta u_q}{L_q} \\ \quad +(\frac{R_{sz}}{L_{qz}} - \frac{R_s}{L_q})i_q + (\frac{1}{L_{qz}} - \frac{1}{L_q})u_q^* - (\frac{\omega_e \varphi_f}{L_q} + \frac{\omega_e \varphi_{fz}}{L_{qz}}) \\ = \frac{1}{L_{qz}}(-\omega_e L_{dz} i_d - R_{sz} i_q + u_q^* - \omega_e \varphi_{fz}) + Z_q \\ Z_d = \frac{\Delta u_d}{L_d} + (-\frac{R_s}{L_d} + \frac{R_{sz}}{L_{dz}})i_d + (\frac{1}{L_d} - \frac{1}{L_{dz}})u_d^* \\ Z_q = \frac{\Delta u_q}{L_q} + (\frac{R_{sz}}{L_{qz}} - \frac{R_s}{L_q})i_q + (\frac{1}{L_{qz}} - \frac{1}{L_q})u_q^* \\ \quad -(\frac{\omega_e \varphi_f}{L_q} + \frac{\omega_e \varphi_{fz}}{L_{qz}}) \end{cases} \tag{10}$$

where Z_d and Z_q are the *d*–*q* axis current disturbance caused by parameters mismatch, respectively. Most previous DPCC only substitutes the changed parameters into (9), but the disturbance Z_{dq} are ignored. The analysis and compensation of Z_{dq} is consistent with the constructers of ADRC and ESO, which is one of the motivations of the method proposed in this article. ESO is used to observe the output and disturbance of the system, DPCC is used as the control law to control the observed value, and advanced DRL algorithm is used to compensate the proposed control system based on the disturbance observation value. On the basis of the framework of ADRC, the ADRDPCC algorithm proposed in this article is to improve the performance of the current loop. Therefore, the proposed method is not only an improvement of DPCC, but also can be seen as a special ADRC current controller.

The ESO model of ADRDPCC is constructed as

$$\begin{cases} e_d(k) = i_{do}(k) - \hat{i}_d(k) \\ \hat{i}_d(k+1) = \hat{i}_d(k) + T_s\{[-\frac{R_{sz}}{L_{dz}} i_{do}(k) + \frac{L_{qz}}{L_{dz}}\omega_e(k+1) i_{qo}(k) \\ \qquad + \frac{u_d^*(k-1)}{L_{dz}}] + \hat{Z}_d(k+1) + \beta_1 fal_1(e_d(k))\} \\ \hat{Z}_d(k+1) = \hat{Z}_d(k) + T_s \beta_2 fal_2(e_d(k)) \end{cases} \tag{11}$$

$$\begin{cases} e_q(k) = i_{qo}(k) - \hat{i}_q(k) \\ \hat{i}_q(k+1) = \hat{i}_q(k) + T_s\{[-\frac{R_{sz}}{L_{qz}} i_{qo}(k) - \frac{L_{dz}}{L_{qz}}\omega_e(k+1) i_{do}(k) \\ \qquad - \frac{\omega_e(k+1)\varphi_{fz}}{L_{qz}} + \frac{u_q^*(k-1)}{L_{qz}}] + \hat{Z}_q(k+1) \\ \qquad + \beta_3 fal_3(e_q(k))\} \\ \hat{Z}_q(k+1) = \hat{Z}_q(k) + T_s \beta_4 fal_4(e_q(k)) \end{cases} \tag{12}$$

where i_{dqo} is the observed value of the *d*–*q* axis current, $\beta_{1\text{-}4}$ are gain coefficients. Because DPCC is essentially a deterministic process of solving equations for expected values, and used to replace the traditional NLSEF as the control law of ADRC system, the stability analysis of this method is essentially the stability analysis of classical nonlinear ADRC, which has been fully proved by many scholars [35], [36], [37], [38].

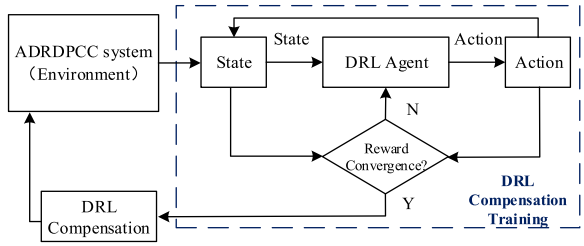


Fig. 4. Training principle of DRL compensation.

B. Compensation Design Based on MDP

Based on the observation and feedback of ESO, DPCC module is employed to output the original value of reference voltage, and the model-free DRL method is designed to compensate the reference voltage. In order to make model-free compensation mechanism effective, a compensation mechanism training scheme needs to be designed based on DRL method. The reward function should be designed to comprehensively evaluate the control effect of the motor after compensation. Each compensation for DPCC is regarded as an "Action" of the compensation mechanism, and the situation of the control system after compensation is regarded as the "State," and State at this time is evaluated according to the reward function. According to different evaluations, DRL will train the compensation mechanism to continuously adjust the compensation strategy until the end of training. DRL compensation mechanism after training can provide the most appropriate compensation value for the original voltage reference according to the different observation values of ESO, and obtain better control effect.

The process of setting Action, State, reward function is called MDP, in the form of which DRL is implemented. Specifically, MDP uses s_t to represent the current State and a_t to represent the current Action after decision. Each Action a_t taken in the current State s_t will generate a new State s_{t+1}. The following State-Action sequence τ will be obtained through continuous judgment and decision-making.

$$\tau \sim \{s_0, a_0, s_1, a_1 \ldots s_{t-1}, a_{t-1}, s_t\}. \tag{13}$$

The reward function can be designed according to different requirements, and the reward value r_t calculated according to the reward function is used to evaluate each decision and Action. The optimal decision strategy $a_t = \pi(s_t)$ can be obtained when the total reward value is maximum. Agent inside the DRL constantly interacts and explores with the control environment, updates and corrects Action a_t that should be taken under different States s_t according to the evaluation value r_t until r_t reaches the optimization and converges, so that the training of DRL compensation mechanism can be completed as shown in Fig. 4.

To be specific, the total reward value can be obtained while considering the impact of long-term returns

$$R_{tk} = \sum_{k=0} \gamma^k r_{t+k+1} \tag{14}$$

where γ is the discount coefficient with the size between (0, 1), which is used to prevent the unlimited expansion of the total

reward value. The total reward value is calculated in the form of expected value, and the reward expectation obtained according to the current state and strategy can be obtained

$$\begin{aligned} V_\pi(s_t) &= E_\pi\left[\sum_{k=0}^{\infty} R_{tk}|s_t\right] \\ &= E_\pi[r_{t+1}\cdot\gamma^0 + r_{t+2}\cdot\gamma^1 + \ldots|s_t] \\ &= E_\pi[r_{t+1}\cdot\gamma^0 + \gamma(r_{t+2} + r_{t+3}\cdot\gamma + \ldots)|s_t] \\ &= E_\pi[r_{t+1} + \gamma R_{tk+1}|s_t] \\ &= E_\pi[r_{t+1} + \gamma V_\pi(s_{t+1})|s_t]. \end{aligned} \tag{15}$$

Equation (15) is also called the state value function. Accordingly, the value-action value function which calculates the expected value according to the current state, action, and strategy can be obtained

$$\begin{aligned} Q_\pi(s_t, a_t) &= [r_s^a + \gamma V_\pi(s_1)|_{s_1}|s_t, a_t] \\ &= E_\pi\left[\sum_{k=0}^{\infty}\gamma^k r_{t+k+1}|s_t, a_t\right] \\ &= E_\pi[r_{t+1} + \gamma Q_\pi(s_{t+1}, a_{t+1})|s_t, a_t]. \end{aligned} \tag{16}$$

In order to optimize the output compensation, it is necessary to find the strategy that can maximize the state value function, which is the training goal of DRL compensation mechanism.

C. Training Process for ADRDPCC

The goal of DRL is to find the optimal policy π^* which can maximize the excepted value

$$\pi^* = \arg\max_{\pi} E_{\tau\sim\pi(\tau)}[r(\tau)]. \tag{17}$$

The optimal policy π^* is parameterized and represented by $J(\theta)$, where θ is the structure parameters of the policy. The expectation of total reward value is expanded, and one can get

$$J(\theta) = E_{\tau\sim\pi_\theta(\tau)}[r(\tau)] = \int_{\tau\sim\pi_\theta(\tau)} \pi_\theta(\tau) r(\tau) d\tau. \tag{18}$$

Differentiating (18), and the gradient of the optimal strategy can be obtained

$$\begin{aligned} \nabla_\theta J(\theta) &= \int_{\tau\sim\pi_\theta(\tau)} \nabla_\theta \pi_\theta(\tau) r(\tau) d\tau \\ &= \int_{\tau\sim\pi_\theta(\tau)} \pi_\theta(\tau)\nabla_\theta \log \pi_\theta(\tau) r(\tau) d\tau \\ &= E_{\tau\sim\pi_\theta(\tau)}[\nabla_\theta \log \pi_\theta(\tau) r(\tau)] \\ &= E_{\tau\sim\pi_\theta(\tau)}\left[\sum_{t=0}^{T}\nabla_\theta \log \pi_\theta(a_{i,t}|s_{i,t})\left(\sum_{t=0}^{T} r(s_{i,t}, a_{i,t})\right)\right] \\ &= \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{t=0}^{T}\nabla_\theta \log \pi_\theta(a_{i,t}|s_{i,t})\left(\sum_{t=0}^{T} r(s_{i,t}, a_{i,t})\right)\right]. \end{aligned} \tag{19}$$

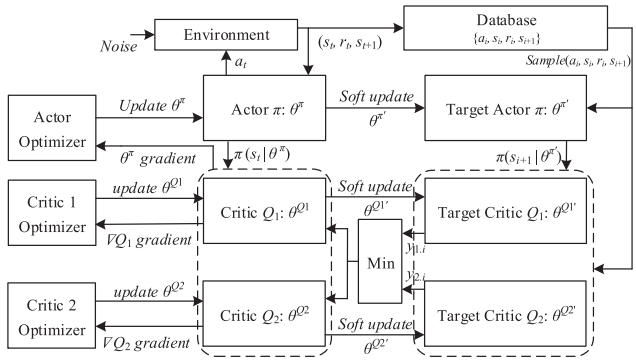


Fig. 5. Optimization framework of TD3 algorithm.

The gradient in (19) is calculated based on different network policy parameters θ. The optimal strategy can be found and the compensation mechanism can be optimized by continuously updating the parameters θ of the strategy and decreasing the gradient. The gradient descent method to optimize θ is designed as follows.

After each Action a_t, the corresponding r_t and s_{t+1} will be fed back, and the data set $\{s_t, a_t, r_t, s_{t+1}\}$ will be formed with a_t and s_t. The datasets will be uniformly stored to form the database in Fig. 5, which can be written as $\{(s_i, a_i, r_i, s_{i+1}) \mid i = 1,2,\ldots,N\}$, and Agent will randomly sample different numbers of data sets from the database for the DRL model training. Twin delayed deep deterministic strategy (TD3) algorithm with actor-critic structure shown in Fig. 5 is selected as the DRL algorithm to train the compensation mechanism and update θ. TD3 consists of actor which can implement the strategy and critic which can evaluate the strategy and optimize the actor network. The deterministic strategy gradient method is used to update θ, and two critics are used to mitigate the overestimation. Because TD3 algorithm has the disadvantage of weak exploration, noise is added to each action during training to expand the exploration range and prevent it from falling into the local optimum.

TD3 introduces target networks to avoid the divergence problem in continuous training, and uses delayed strategy to update the parameters after fully minimizing the time difference error. The update rate $\eta \le 1$, and the update formulas are

$$\begin{cases} \theta^{Q'} \leftarrow \eta\theta^{Q} + (1-\eta)\,\theta^{Q'} \\ \theta^{\pi'} \leftarrow \eta\theta^{\pi} + (1-\eta)\,\theta^{\pi'} \end{cases}. \tag{20}$$

The evaluation value y_i calculated from two critics, and the loss L between the target network and the original one, can be obtained by (21). Equation (19) is used to optimize θ^π which can maximize the evaluation value of Q, and θ^Q of critics are optimized by minimizing L

$$\begin{cases} y_i = r_i + \gamma \min\limits_{j=1,2} Q'\left(s_{i+1}, \pi'\left(s_{i+1}|\theta^{\pi'}\right)|\theta^{Q'j}\right) \\ L = \frac{1}{N}\sum\limits_{i}^{N}\left(\left(y_i - Q\left(s_i, a_i|\theta^{Qj}\right)\right)^2\right) \end{cases}. \tag{21}$$

When the training of DRL compensation mechanism is completed, the overall ADRDPCC control structure is formed as shown in Fig. 6.

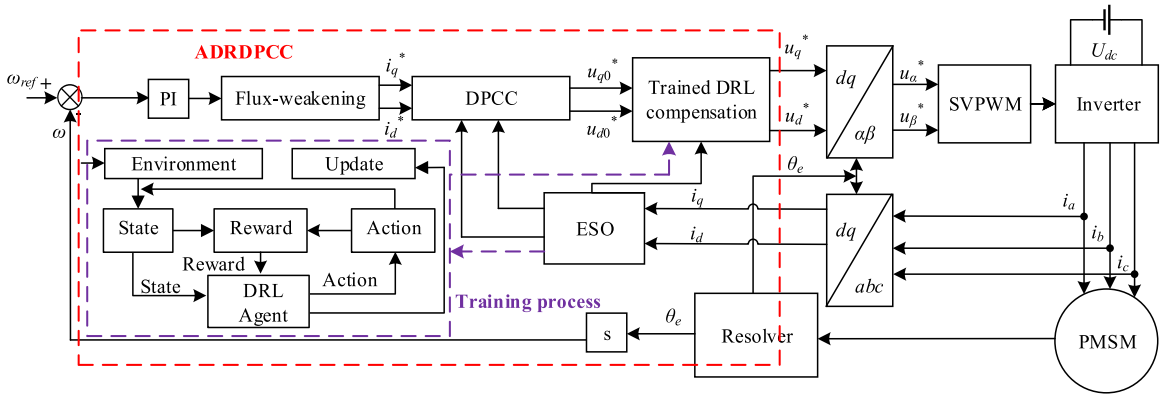


Fig. 6. ADRDPCC control system based on data driven of PMSM for MEA.

TABLE I
MAIN PARAMETERS OF PMSM

Symbol	Item	Value
R_s	Phase resistance	0.0713 Ω
L_q	q-axis inductance	605 μH
L_d	d-axis inductance	519.5 μH
Ψ_f	Rotor flux	0.0201 Wb
J	Rotational inertia	4.07x10^{-4} kg.m^2
p_n	Pole pairs	5

IV. SIMULATIONS OF ADRDPCC SYSTEM

A. Optimization Setting of ADRDPCC

Table I gives the main parameters of the motor used under simulation conditions. Because DRL training is autonomous and exploratory, and requires time and calculation, it is not suitable for online training. To ensure the accuracy and effectiveness of offline training, data and motor parameters for training are collected from the test of the experimental platform and the experimental motor. It is necessary to train the DRL compensation ability of the simulation model in various situations. Therefore, the simulation motor is designed to start to 1000 r/min and the start torque is set to $|0.5\sin(\pi t)|$ N·m. The control system is suddenly applied with a load torque of 1N·m at 3 s, and the speed is suddenly increased to 3000 r/min at 6 s. The effect of the periodic load disturbances, stepwise constante load disturbances, speed disturbances are studied in the simulation to fully demonstrate the advantages of the proposed method.

The given value of the d–q axis current, tracking error of the d–q axis current and speed curve are set as state of the training system. In order to improve the control performance of the proposed method, the current tracking error, speed tracking error and the influence of interference are taken as optimization objectives, and the design reward function is

$$R = \frac{1}{n}\sum_{i=1}^{n}\left(\begin{vmatrix} r_1 e_{os}/s_1 + r_2|e_l|/s_2 + r_3 e_{ids}/s_3 \\ +r_4 e_{iqs}/s_4 + r_5 t_{ss}/s_5 + r_6 t_{sl}/s_6 \end{vmatrix}\right)^i \quad (22)$$

where e_{os} is the error of the observed state and reference state, e_l is the speed error of the sudden load transition, and e_{idqs} is the error of the observed d–q axis current and d–q axis current, respectively. t_{ss}, and t_{sl} are the rising time of speed and the recovery time of the disturbance load, respectively, which could be calculated after the disturbances occur. $s_{1\text{-}6}$ are the standardization coefficients because of different dimensions between optimized objectives. $r_{1\text{-}6}$ are the weight coefficients which can be changed according to the different requirements of the application. The optimal results can be obtained after the value of R is converged to the smallest. Each interaction between the control system and DRL training model will generate and record a dataset. The DRL model will learn and be trained once every 150 interactions, and the parameters of DRL compensation mechanism will be updated.

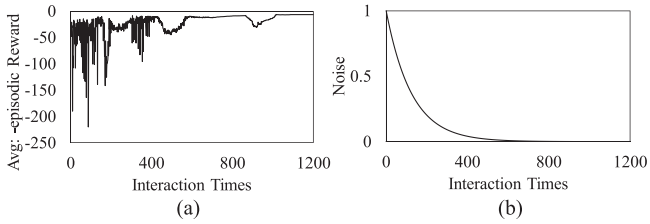


Fig. 7. Training process of ADRDPCC. (a) Convergence curve. (b) Noise curve.

B. Convergence of the DRL Compensation

The DRL calculation program is written in Python language, jointly trained by calling MATLAB, and implemented by the function of TensorFlow v2.10. The program is trained by a computer equipped with Inter(R) Core(TM) Inter(R) Core i7-10700KF, 32GB, and the procedure conforms to the real time coding specification. The Agent learns synchronously with the simulation. In the procedure of training Agent, the noise disturbance is applied to the output of Action to avoid the Agent being trapped in the local optimization. The noise disturbance is set to change exponentially, and gradually decreases with the round of interactions to demonstrate the real effect of the training, which enhances the robustness of the system.

Fig. 7 shows the training process and curves of ADRDPCC. Gaussian noise is added to each Action to explore a wider range as shown in Fig. 7(b), which also brings fluctuations, even when

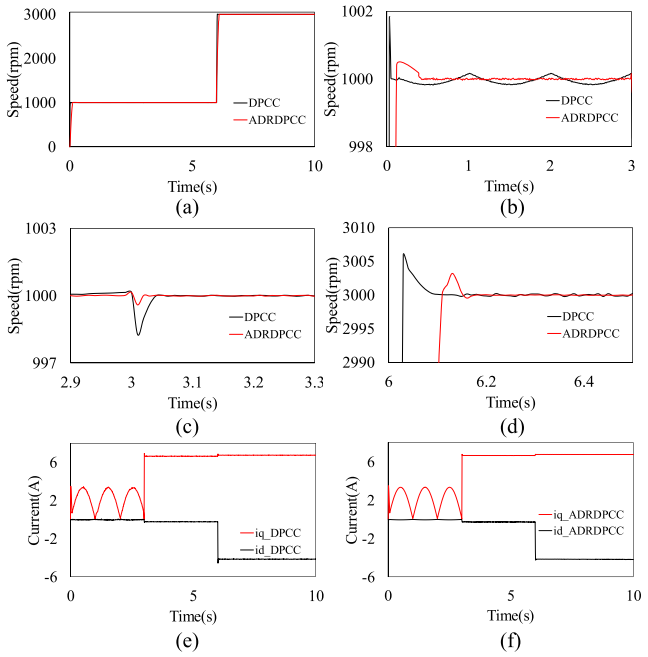


Fig. 8. Comparison of the speed simulation results. (a) Overview. (b) Start-up state. (c) Load transition state. (d) Speed transition state. (e) *d*–*q* axis current of DPCC. (f) *d*–*q* axis current of ADRDPCC.

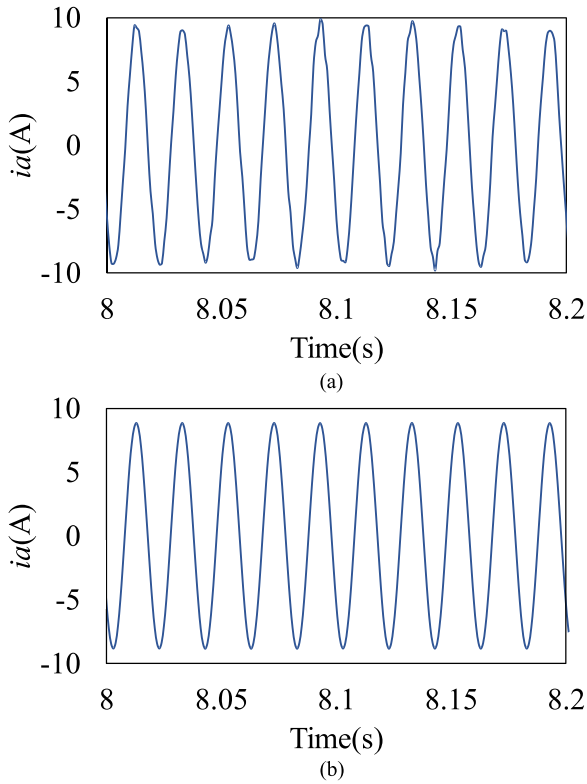


Fig. 9. Comparison of the current i_a simulation results. (a) Deadbeat predictive current control. (b) Active disturbance rejection deadbeat predictive current control.

the training is almost completed. Therefore, when the noise is reduced to near zero, DRL training is close to convergence and tends to be stable as shown in Fig. 7(a). The training converges close to 1000 rounds finally.

C. Comparison With DPCC

The trained DRL compensation mechanism is applied in the ADRDPCC control system, and the simulation comparison with the classic DPCC is shown in Figs. 8–10.

The control effect of the current loop can be verified from the speed tracking effect. It can be seen from Fig. 8(b) that the speed curve of ADRDPCC after training is smoother, and the speed of traditional DPCC is affected by persistent disturbances and fluctuate slightly in sine form. ADRDPCC has a slower response speed, which is due to the simpler calculation process of DPCC. However, Fig. 8(c) and (d) show that the proposed method has less overshoot and better disturbance rejection ability after training. The current waveform of the proposed method is more stable and has less jitter, which is shown in Fig. 8(e) and (f).

Figs. 9 and 10 show the analysis results of current i_a and total harmonic distortion (THD) under different control methods. THDs of the two control methods are 17.17% and 5.79%, respectively. Simulation results verify that the proposed method has more sinusoidal current and less harmonic distortion.

V. Experiments

In order to verify the effectiveness of the proposed control method, an experimental platform is built as shown in Fig. 11. The specific parameters of the motor are given in Table I. The experimental platform is "MEA RDS loading experimental

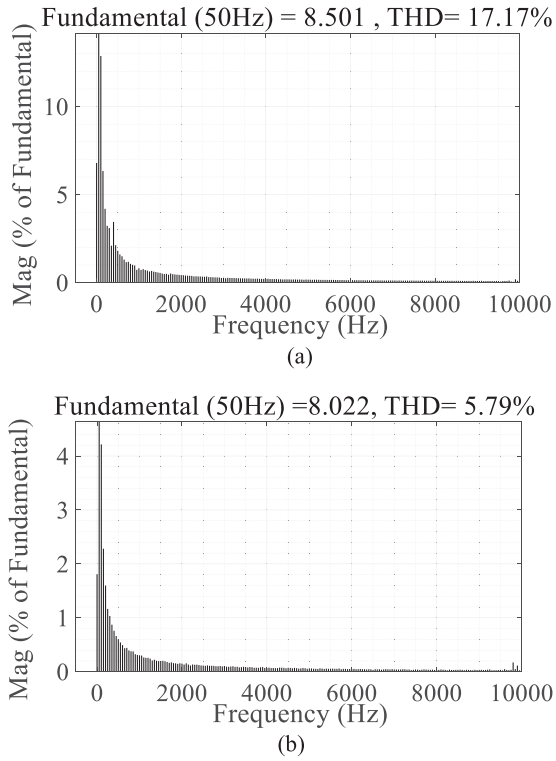


Fig. 10. Comparison of the simulation current harmonic spectrum. (a) Deadbeat predictive current control. (b) Active disturbance rejection deadbeat predictive current control.

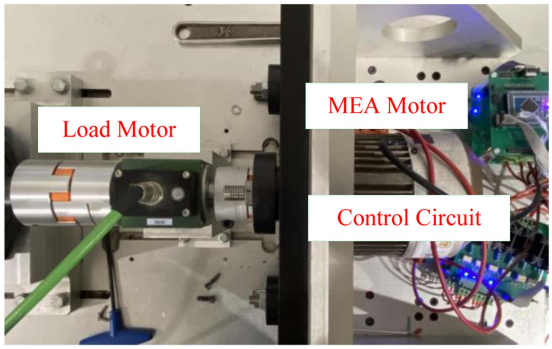


Fig. 11. Experimental platform.

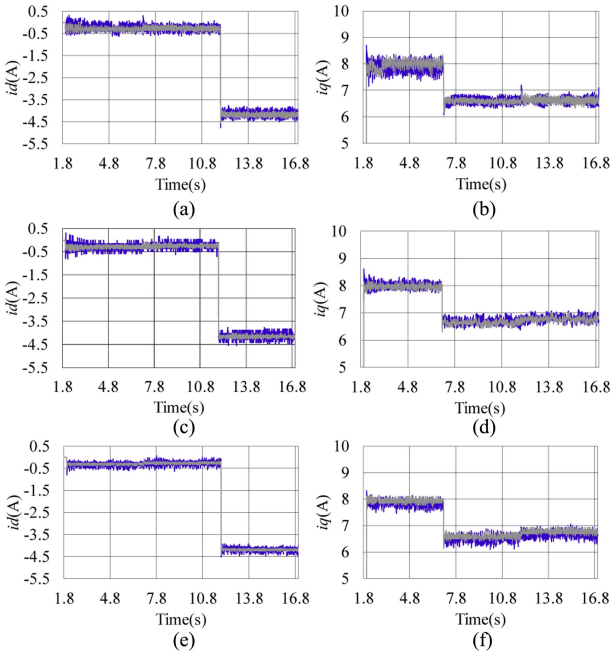


Fig. 12. d–q axis current under parameter mismatch $L_q = 2L_{q0}$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

platform," which plays the role of loading and control. All control methods are implemented in the form of C language in the micro controller unit (MCU) of STM32F407, in which the clock frequency is 168 MHz, and the sampling period is 50 μs. The algorithm execution time of the control scheme is 47 μs, which represents the computational complexity.

Parameter mismatch experiments are carried out, and the control effect under different parameters variations and different degrees of parameter variations are studied to verify the performance of DPCC, DPCC+ESO and ADRDPCC. In the experiments, considering that MEA motors rarely work in extremely low speed environment, starting at a low speed of 500 r/min is a suitable choice, and the initial torque of the motor are set to 1.2 N·m. The load torque of control system is suddenly changed to 1 N·m at 7 s, and a 3000 r/min speed is suddenly applied to the motor at 12 s. The results of current references and current responses with different parameter mismatch are shown in Figs. 12–17, where the blue line and the grey line

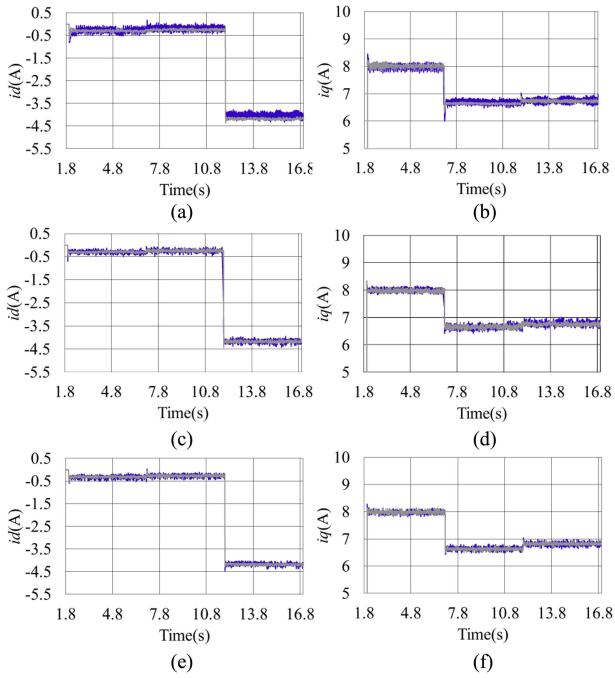


Fig. 13. d–q axis current under parameter mismatch $L_q = 0.5L_{q0}$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

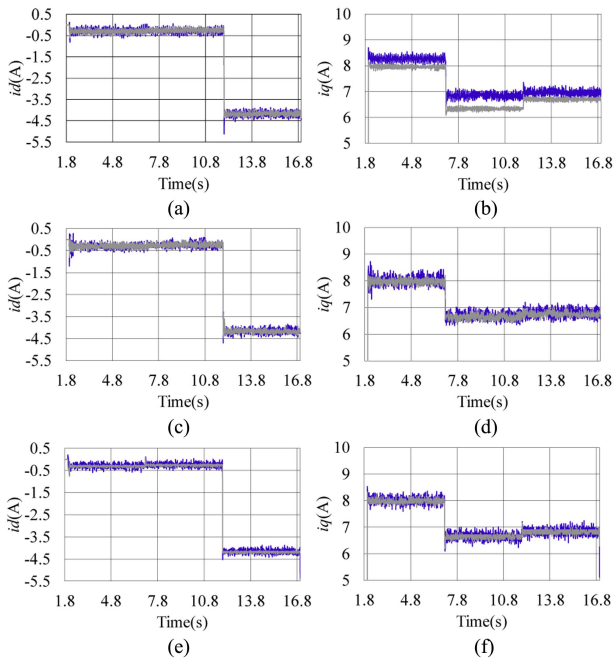


Fig. 14. d-q axis current under parameter mismatch $R = 10R_0$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

are the observation value and the reference value of the d–q axis, respectively. In the experiment, the effect of the speed disturbances, stepwise constant load disturbances and parameter mismatch disturbances are used to show the superiority of the proposed method. Each disturbance is subdivided to verify

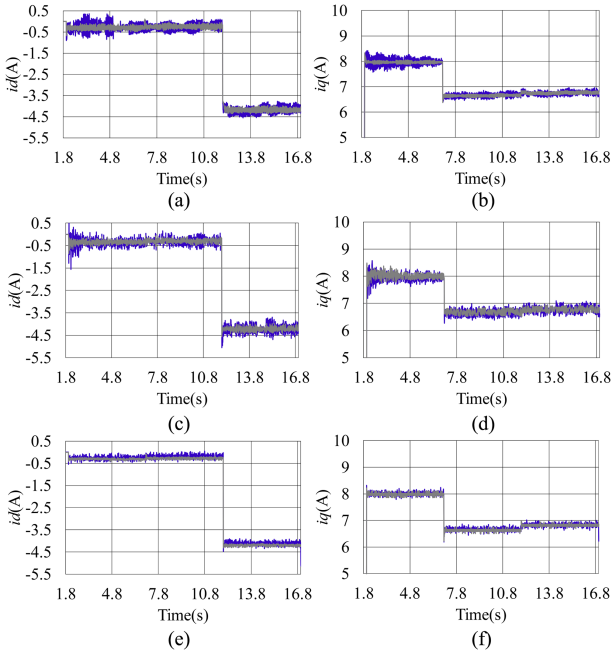


Fig. 15. d–q axis current under parameter mismatch $R = 0.1R_0$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

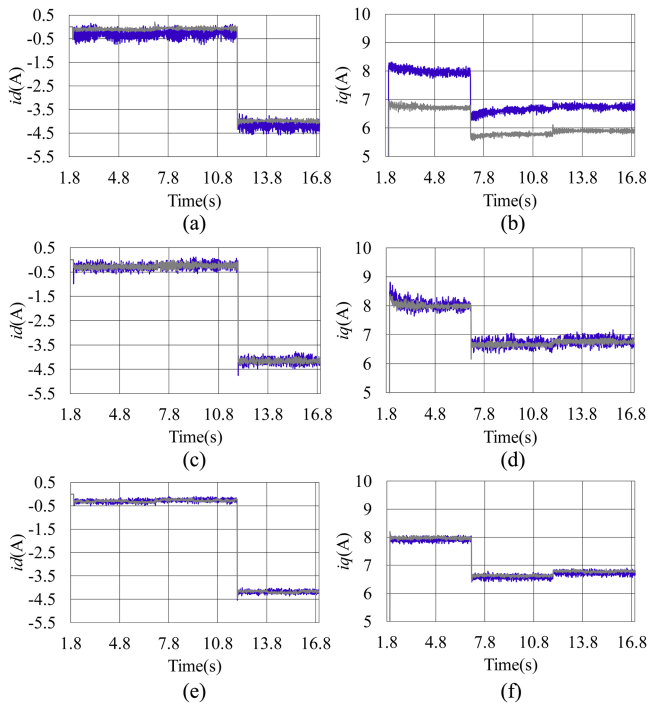


Fig. 16. d–q axis current under parameter mismatch $\psi = 5\psi_0$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

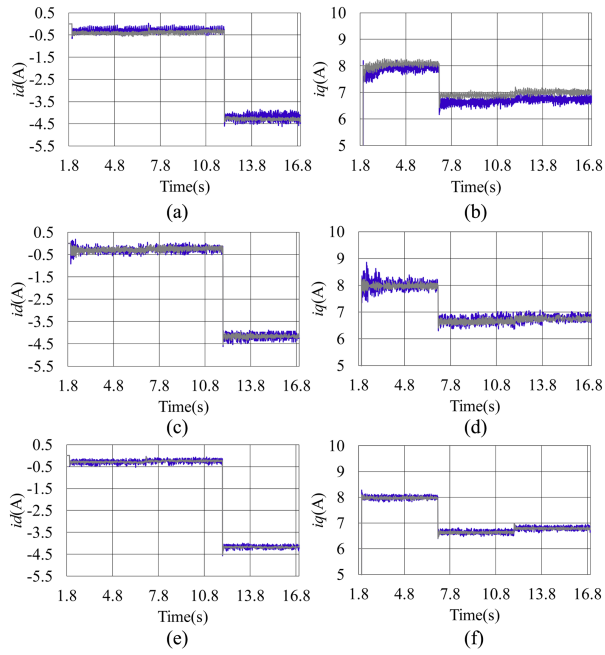


Fig. 17. d–q axis current under parameter mismatch $\psi = 0.2\psi_0$. (a) and (b) d–q axis current of DPCC. (c) and (d) d–q axis current of DPCC+ESO. (e) and (f) d–q axis current of ADRDPCC.

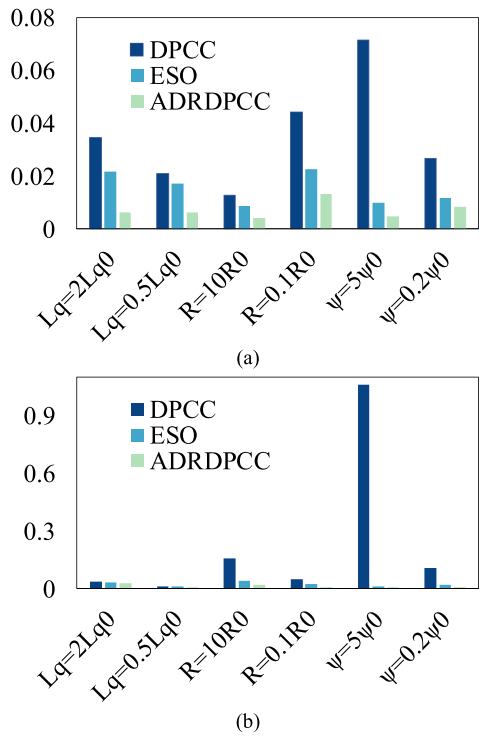


Fig. 18. Comparison of MSE. (a) MSE of d-axis current. (b) MSE of q-axis current.

the impact of varying degrees of each disturbance on control performance.

It can be seen from Figs. 12 and 13 that the performances of DPCC and DPCC+ESO are close to that of the proposed method when the inductance parameter is mismatched, but ADRDPCC has smaller current fluctuations.

Figs. 14 and 15 show that ADRDPCC can better track the reference current when $R = 10R_0$. DPCC+ESO has a better current tracking effect than DPCC, but the current jitter is large.

Though DPCC can track the reference current when $R = 0.1R_0$, but it is unstable compared with that of ADRDPCC.

Figs. 16 and 17 show that when flux linkage parameter is mismatched, DPCC cannot make the actual current track the reference current. The actual current of DPCC+ESO can track the reference current, but the current control is unstable and has a large jitter when ADRDPCC is almost unaffected. The experimental results show the effectiveness and superiority of the proposed method in respect of disturbance rejection and parameter mismatch.

Fig. 18 shows the comparison of mean square errors (MSEs) of experimental results in the form of histograms to present the quantitative discussion. It can be vaguely seen in Figs. 12–17 that in the case of parameter mismatch, the steady-state control effect of DPCC is relatively close to that of DPCC+ESO. In Fig. 18, it can be clearly seen that DPCC is more sensitive to disturbances, and the proposed method has better suppression effects in the face of parameter mismatches and disturbances compared to the other two methods.

VI. Conclusion

In this article, a model-free active disturbance rejection DPCC method based on data-drive is proposed. The proposed method is based on ADRC structure, using ESO as disturbance observer, and training DRL as model-free compensation mechanism to form ADRDPCC control method. Simulation results show that the proposed method can effectively reduce the current and speed tracking errors and improve the disturbance rejection ability according to the training objectives. Experimental results show that the proposed method can effectively suppress the influence of parameter mismatch on traditional DPCC methods. Simulation and experiments verify the effectiveness and superiority of the proposed method. In the future, more in-depth research will be conducted on a wider speed range control and optimization methods.

References

[1] E. Sayed et al., "Review of electric machines in more-/hybrid-/turbo-electric aircraft," *IEEE Trans. Transp. Electrific.*, vol. 7, no. 4, pp. 2976–3005, 2021.

[2] T. Zhao, S. Wu, and S. Cui, "Multiphase PMSM with asymmetric windings for more electric aircraft," *IEEE Trans. Transp. Electrific.*, vol. 6, no. 4, pp. 1592–1602, Dec. 2020.

[3] Z. Song, C. Liu, K. Feng, H. Zhao, and J. Yu, "Field prediction and validation of a slotless segmented-halbach permanent magnet synchronous machine for more electric aircraft," *IEEE Trans. Transp. Electrific.*, vol. 6, no. 4, pp. 1577–1591, Dec. 2020.

[4] M. A. A. Mohamed, M. Rashed, and S. Bozhko, "Enhanced flux weakening control of PMM-based starter-generator system in more electric aircraft," in *Proc. 10th Int. Conf. Power Electron., Mach. Drives*, 2021, pp. 249–254.

[5] J. Wang, J. Wu, C. Gan, and Q. Sun, "Comparative study of flux-weakening control methods for PMSM drive over wide speed range," in *Proc. 19rd Int. Conf. Elect. Mach. Syst.*, 2016, pp. 1–6.

[6] X. Yuan, S. Zhang, and C. Zhang, "Nonparametric predictive current control for PMSM," *IEEE Trans. Power Electron.*, vol. 35, no. 9, pp. 9332–9341, Sep. 2020.

[7] M. Yang, X. Lang, J. Long, and D. Xu, "A flux immunity robust predictive current control with incremental model and extended state observer for PMSM drive," *IEEE Trans. Power Electron.*, vol. 32, no. 12, pp. 9267–9279, Dec. 2017.

[8] W. Tu, G. Luo, R. Zhang, Z. Chen, and R. Kennel, "Finite-control-set model predictive current control for PMSM using grey prediction," in *Proc. IEEE Energy Convers. Congr. Expo.*, 2016, pp. 1–7.

[9] F. Wang, K. Zuo, P. Tao, and J. Rodríguez, "High performance model predictive control for PMSM by using stator current mathematical model self-regulation technique," *IEEE Trans. Power Electron.*, vol. 35, no. 12, pp. 13652–13662, Dec. 2020.

[10] X. Liu, L. Zhou, J. Wang, X. Gao, Z. Li, and Z. Zhang, "Robust predictive current control of permanent-magnet synchronous motors with newly designed cost function," *IEEE Trans. Power Electron.*, vol. 35, no. 10, pp. 10778–10788, Oct. 2020.

[11] X. Sun, T. Li, Z. Zhu, G. Lei, Y. Guo, and J. Zhu, "Speed sensorless model predictive current control based on finite position set for PMSHM drives," *IEEE Trans. Transp. Electrific.*, vol. 7, no. 4, pp. 2743–2752, Dec. 2021.

[12] O. Wallscheid and E. F. B. Ngoumtsa, "Investigation of disturbance observers for model predictive current control in electric drives," *IEEE Trans. Power Electron.*, vol. 35, no. 12, pp. 13563–13572, Dec. 2020.

[13] M. Bermudez, M. R. Arahal, M. J. Duran, and I. Gonzalez-Prieto, "Model predictive control of six-phase electric drives including ARX disturbance estimator," *IEEE Trans. Ind. Electron.*, vol. 68, no. 1, pp. 81–91, Jan. 2021.

[14] Y. Zhang, D. Xu, J. Liu, S. Gao, and W. Xu, "Performance improvement of model-predictive current control of permanent magnet synchronous motor drives," *IEEE Trans. Ind. Appl.*, vol. 53, no. 4, pp. 3683–3695, Aug. 2017.

[15] X. Zhang, B. Hou, and Y. Mei, "Deadbeat predictive current control of permanent-magnet synchronous motors with stator current and disturbance observer," *IEEE Trans. Power Electron.*, vol. 32, no. 5, pp. 3818–3834, May 2017.

[16] H. Cheng, S. Sun, X. Zhou, D. Shao, S. Mi, and Y. Hu, "Sensorless DPCC of PMLSM using SOGI-PLL-based high-order SMO with cogging force feedforward compensation," *IEEE Trans. Transp. Electrific.*, vol. 8, no. 1, pp. 1094–1104, Mar. 2022.

[17] Z. Liu et al., "A modified deadbeat predictive current control for improving dynamic performance of PMSM," *IEEE Trans. Power Electron.*, vol. 37, no. 12, pp. 14173–14185, Dec. 2022.

[18] X. Li, S. Zhang, C. Zhang, Y. Zhou, and C. Zhang, "An improved deadbeat predictive current control scheme for open-winding permanent magnet synchronous motors drives with disturbance observer," *IEEE Trans. Power Electron.*, vol. 36, no. 4, pp. 4622–4632, Apr. 2021.

[19] C.-K. Lin, T.-H. Liu, J.-T. Yu, L.-C. Fu, and C.-F. Hsiao, "Model-free predictive current control for interior permanent-magnet synchronous motor drives based on current difference detection technique," *IEEE Trans. Ind. Electron.*, vol. 61, no. 2, pp. 667–681, Feb. 2014.

[20] C.-K. Lin, J.-T. Yu, Y.-S. Lai, and H.-C. Yu, "Improved model-free predictive current control for synchronous reluctance motor drives," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3942–3953, Jun. 2016.

[21] P. Carlet, F. Tinazzi, S. Bolognani, and M. Zigliotto, "An effective model-free predictive current control for synchronous reluctance motor drives," *IEEE Trans. Ind. Appl.*, vol. 55, no. 4, pp. 3781–3790, Apr. 2019.

[22] J. Yang, J. Su, S. Li, and X. Yu, "High-order mismatched disturbance compensation for motion control systems via a continuous dynamic sliding mode approach," *IEEE Trans. Ind. Inform.*, vol. 10, no. 1, pp. 604–614, Feb. 2014.

[23] T. Türker, U. Buyukkeles, and A. F. Bakan, "A robust predictive current controller for PMSM drives," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3906–3914, Jan. 2016.

[24] P. G. Carlet, A. Favato, S. Bolognani, and F. Dörfler, "Data-driven continuous-set predictive current control for synchronous motor drives," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 6637–6646, Jun. 2022.

[25] W. Wu et al., "Data-driven iterative learning predictive control for power converters," *IEEE Trans. Power Electron.*, vol. 37, no. 12, pp. 14028–14033, Dec. 2022.

[26] D. Qin, L. Wang, S. Jin, and Z. Zhang, "Data-driven model-based smart control of intelligent gate drive for converter operational performance improvement," in *Proc. IEEE Appl. Power Electron. Conf. Expo.*, 2021, pp. 2405–2411.

[27] D. Nan, J. Li, Y. Weng, L. Lian, C. Yu, and S. Li, "Data-driven adaptive PID control of unknown quadrotor UAVs," in *Proc. IEEE 9th Data Driven Control Learn. Syst. Conf.*, 2020, pp. 953–958.

[28] J. Hu, Q. Wang, Y. Ye, and Y. Tang, "Toward online power system model identification: A deep reinforcement learning approach," *IEEE Trans. Power Syst.*, vol. 38, no. 3, pp. 2580–2593, May 2023, doi: 10.1109/TPWRS.2022.3180415.

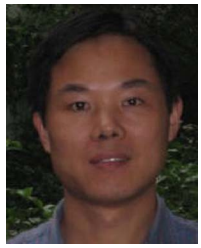
[29] L. Yan, X. Chen, Y. Chen, and J. Wen, "A cooperative charging control strategy for electric vehicles based on multi-agent deep reinforcement learning," *IEEE Trans. Ind. Inform.*, vol. 18, no. 12, pp. 8765–8775, Dec. 2022, doi: 10.1109/TII.2022.3152218.

[30] S. Park et al., "A deep reinforcement learning framework for fast charging of Li-ion batteries," *IEEE Trans. Transp. Electrific.*, vol. 8, no. 2, pp. 2770–2784, Jun. 2022.

[31] Y. Wang, S. Fang, and J. Hu, "Active disturbance rejection control based on deep reinforcement learning of PMSM for more electric aircraft," *IEEE Trans. Power Electron.*, vol. 38, no. 1, pp. 406–416, Jan. 2023.
[32] J. Hu, Q. Wang, and Y. Tang, "An online deep reinforcement learning based parameter identification method for HVDC system," in *Proc. IEEE Power Energy Soc. Gen. Meeting*, 2021, pp. 1–5.
[33] J. Han, "Active disturbance rejection control technique-the technique for estimating and compensating the uncertainties," 2008.
[34] Z.-Q. Gao, "Scaling and bandwidth-parameterization based controller tuning," in *Proc. Amer. Control Conf.*, 2003, pp. 4989–4996.
[35] B. Guo and Z. Zhao, "On convergence of nonlinear active disturbance rejection for SISO systems," in *Proc. 24th Chin. Control Decis. Conf.*, 2012, pp. 3507–3512.
[36] Z. Zhao and B. Guo, "A nonlinear extended state observer based on fractional power functions," *Automatica*, vol. 81, pp. 286–296, 2017.
[37] Z. Zhao and B. Guo, "On convergence of nonlinear active disturbance rejection control for MIMO systems," in *Proc. Chin. Control Conf.*, 2012, pp. 434–441.
[38] Z. Zhao and B. Guo, "A novel extended state observer for output tracking of MIMO systems with mismatched uncertainty," *IEEE Trans. Automat. Control*, vol. 63, no. 1, pp. 211–218, Jan. 2018.

Yicheng Wang received the B.Eng. degree in electrical engineering and automation from Southwest Jiaotong University, Chengdu, China in 2018, and the M.S. degree in electrical engineering in 2021 from Southeast University, Nanjing, China, where he is currently working toward the Ph.D. degree in electrical engineering.

His research interests include control strategies for permanent magnet machines and power electronics.



Shuhua Fang (Senior Member, IEEE) received the M.S. degree in electrical engineering from Shandong University of Science and Technology, Jinan, China, in 2004 and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2008.

From 1998 to 2001, he was an Associate Engineer with the Xuzhou Coal and Mine Machinery Factory, where his research activities were primarily in the area of the design, analysis and control of electrical apparatus for coal and mine. Since 2008, he has been with Southeast University, where he is currently a full Professor with the School of Electrical Engineering. From 2013 to 2014, he was a Visiting Professor with the Wisconsin Electric Machine and Power Electronics Consortium, University of Wisconsin-Madison. His research interests include design, simulation and control for both permanent magnet actuator and intelligent apparatus.

Demin Huang received the B.Eng. degree in electrical engineering and automation from Nanjing Institute of Technology, Nanjing, China in 2020. He is currently working toward the M.S. degree in electrical engineering at Southeast University, Nanjing, China.

His research interests include the control strategies of permanent magnet machines and power electronics.