Letters

Learning-Based Neural Dynamic Surface Predictive Control for MMC

Xing Liu, *Senior Member, IEEE*, Lin Qiu, *Senior Member, IEEE*, Jose Rodríguez, *Life Fellow, IEEE*, Kui Wang, *Senior Member, IEEE*, Yongdong Li, *Senior Member, IEEE*, and Youtong Fang, *Senior Member, IEEE*

***Abstract*—Reinforcement learning technique was developed recently as an interesting topic in designing adaptive optimal controllers. This technique explicitly provided a feasible solution to circumvent the "curse of dimensionality" and requiring a system model inherent in the classical dynamic programming algorithm. By virtue of this property, in our work, by introducing this technique into a predictor-based online adaptive neural dynamic surface predictive control architecture, we concentrate on a novel robust predictive control framework subject to system uncertainties. To be specific, in this presented framework, an adaptive dynamic programming control strategy utilizing a critic neural network point of view is developed to learn the optimal control policy. Our modification is able to facilitate the alleviation of performance deterioration caused by system uncertainties and enable the smooth and fast learning, while keeping the merits of the finite control-set model predictive control. Finally, the interest and applicability of the proposed control methodology are verified by performance evaluation.**

***Index Terms*—Dynamic surface control, finite control-set model predictive control, neural network, reinforcement learning.**

I. Introduction

RECENTLY, reinforcement learning (RL) has received considerable research interests [1]. Its main characteristic is to improve control policy by properly evaluating feedbacks from environments, avoiding directly analytically solving the Hamilton–Jacobi–Bellman (HJB) equations. In this technique, it is desirable to design a controller, which not only optimizes some predefined performance criteria but also guarantees an adequate level of performance. To this end, many highlighted results have been published [2]. This success has accelerated the evolution of RL techniques in real-world systems. However, although widely accepted from the industry perspective, parametric uncertainties and external unknown disturbances are inevitable issues in practical. These limitations lead to additional difficulties to design RL-based optimal controllers.

A. Related Work and Motivation

To circumvent these limitations, some advances have been successfully achieved. First, a RL-based online optimal control strategy in hybrid ac–dc microgrids is developed in [3], where two neural networks (NNs) referred to as actor NN and critic NN are deployed. The actor NN aims to execute control behaviors, and the critic NN aims to appraise control performance and make feedback to actor. By employing this strategy, the unknown system dynamics can be estimated, and thus, the performance degradation caused by the uncertainties can be alleviated. Afterward, the authors in [4] presented a novel RL-based optimal tracking control, wherein unmeasurable disturbances are taken into account. Thus, the convergence of the control parameters to the optimal solution can be ensured. Subsequently, the authors in [5] proposed an adaptive observation-based efficient RL scheme for uncertain systems. In the described approach, a new concurrent learning adaptive extended observer is explored, where the system state and parameter can be estimated simultaneously. Consequently, the desired control performance can be achieved. Despite the significant research progress, to our knowledge, it seems that no attempt has been made to enhance the robustness and tracking performance of system by deploying the RL approach in combination with the neural dynamic surface predictive control (NDSPC) architecture for power converter systems. This consideration constitutes the main motivation of this letter.

B. Main Contribution

Inspired by the observations made previously, in view of a different perspective, this letter goes one step further and focuses

Manuscript received 1 June 2022; revised 6 July 2022 and 26 July 2022; accepted 15 August 2022. Date of publication 22 August 2022; date of current version 10 October 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 51807177 and Grant 51827810, and in part by China Postdoctoral Science Foundation under Grant 2020M681855, and in part by the National Key Research and Development Program of China under Grant 2019YFB1504603, and in part by Natural Science Foundation of Zhejiang Province under Grant LY21E070004 and Grant LY22E070003. This work of Jose Rodriguez was supported by ANID through projects under Grant FB0008, Grant 1221293, and Grant 1210208. *(Corresponding author: Lin Qiu.)*

Xing Liu and Youtong Fang are with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: xingldl@zju.edu.cn; youtong@zju.edu.cn).

Lin Qiu is with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China, and also with the Zhejiang University-University of Illinois at Urbana-Champaign Institute, Hangzhou 310027, China (e-mail: qiu_lin@zju.edu.cn).

Kui Wang and Yongdong Li are with the State Key Laboratory of Power System, Department of Electrical Engineering, Tsinghua University, Beijing 100084, China (e-mail: wangkui@tsinghua.edu.cn; liyd@tsinghua.edu.cn).

Jose Rodríguez is with the Faculty of Engineering, Universidad San Sebastian Santiago, Santiago de Chile 8420524, Chile (e-mail: jose.rodriguezp@uss.cl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2022.3200857.

Digital Object Identifier 10.1109/TPEL.2022.3200857

on investigating a novel methodology called learning-based NDSPC solution, which is performed by incorporating a RL-based online adaptive NN dynamic surface control (DSC) technique into the finite control-set model predictive control (FCS-MPC) architecture. The objectives of this letter are to enhance the robustness and to handle the optimal power tracking control tasks for power converter systems. In the suggested proposal, an adaptive dynamic programming (ADP) control utilizing a critic NN point of view is developed to learn the optimal control policy. Meanwhile, the critic NN is introduced into the proposed design to online approximate the optimal performance index function of the HJB equation. The key features of our modification are that this proposal not only yields satisfactory robust performance, but also ensures an acceptable level of systematic control objectives tracking and performance optimization.

To sufficiently clear the main contribution of this letter, we contribute two main points to the relevant literature.

1) We first attempt to propose a learning-based NDSPC framework such that unknown nonlinear system dynamics and external disturbances can be simultaneously addressed for power converter systems. Different from existing FCS-MPC studies, it contributes to better representation capability of system uncertainties, and further enhances the generalization capability of the proposed design, and thus it can be more widely applied to other scenarios.
2) Another key contribution of this letter is the conjunction of RL technique and FCS-MPC architecture to cope effectively with the sophisticated online optimization problem subject to robustness characteristics, achieving the satisfactory tracking control behaviors. Moreover, it is noticeable that the presented proposal can be regarded as a proof of concept and can open a very attractive area for future research, which inspires more researchers to devote effort to this field.

Finally, we show the merits of the proposed control methodology by means of a numerical example, where we compare it to the state-of-the-art FCS-MPC solutions for modular multilevel converter (MMC).

II. Proposed Learning-Based Neural Dynamic Surface Predictive Control Solution

In this section, we focus on a new composite robust predictive control solution to identify the system dynamics and lumped uncertainties for power converter systems. It can be performed by integrating a RL-based adaptive neural dynamic surface technique into FCS-MPC framework subject to robustness characteristics, while guaranteeing adaptability to different system dynamics, model variations, and environment changes, as it will be elaborated as follows. Generally aiming at medium voltage applications, the configuration of an MMC, which is connected to the main ac power supply via input filter inductance and resistance, is addressed and shown in Fig. 1.

A. Mathematical Model of MMC

The upper and lower arm voltage of phase-x ($x \in \{a, b, c\}$) with respect to the dc-link mid-point can be described by the following form [6], [7]:

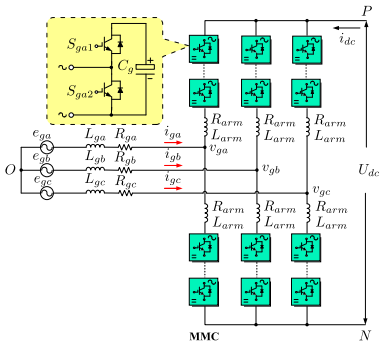

Fig. 1. Schematic diagram of the MMC topology.

$$\begin{cases} \frac{U_{dc}}{2} = & v_{gux} + L_{arm}\frac{di_{ux}}{dt} + i_{ux}R_{arm} - i_{gx}R_g - L_g\frac{di_{gx}}{dt} \\ & +e_{gx} \\ -\frac{U_{dc}}{2} = & -v_{glx} - L_{arm}\frac{di_{lx}}{dt} - i_{lx}R_{arm} - i_{gx}R_g - L_g\frac{di_{gx}}{dt} \\ & +e_{gx} \end{cases} \tag{1}$$

where U_{dc} denotes dc-link voltage. R_{arm} and L_{arm} denote the arm resistance and inductance, respectively. R_g and L_g denote the input filter resistance and inductance, respectively. e_{gx} and i_{gx} denote the input voltage and current of phase-x, respectively. i_{ux} and i_{lx} denote the upper and lower arm current of phase-x, respectively. v_{gux} and v_{glx} denote the upper and lower arm voltage of phase-x, respectively.

According to (1), the input current dynamic of phase-x can be illustrated as

$$\frac{di_{gx}}{dt} = \frac{1}{L_z}(e_{gx} - v_{gx}) - \frac{R_z}{L_z}i_{gx} \tag{2}$$

where $R_z = R_g + 0.5R_{arm}$, and $L_z = L_g + 0.5L_{arm}$. $v_{gx} = \frac{v_{glx} - v_{gux}}{2}$.

Next, we address the dynamic model of MMC in terms of new system control inputs u_P and u_Q as

$$\begin{cases} \frac{dU_{dc}}{dt} = \frac{P_g}{C_{dc}U_{dc}} - \frac{U_{dc}}{C_{dc}R_L} \\ \frac{dP_g}{dt} = -\frac{3}{2L_z}u_P - w_gQ_g - \frac{R_z}{L_z}P_g + \frac{3}{2L_z}({e_{g\alpha}}^2 + {e_{g\beta}}^2) \\ \frac{dQ_g}{dt} = \frac{3}{2L_z}u_Q + w_gP_g - \frac{R_z}{L_z}Q_g \end{cases} \tag{3}$$

where

$$\begin{cases} u_P = e_{g\alpha}v_{g\alpha} + e_{g\beta}v_{g\beta} \\ u_Q = -e_{g\beta}v_{g\alpha} + e_{g\alpha}v_{g\beta} \end{cases} \tag{4}$$

and where P_g and Q_g denote the grid-side input active and reactive powers, respectively. ω_g denotes the grid angular frequency. C_{dc} and R_L denote dc-link capacitor and load resistance, respectively.

We note that the original control inputs $v_{g\alpha}$ and $v_{g\beta}$ can be decoupled by u_P and u_Q. Thus, based on the above analysis, (3) is reconstructed as

$$\begin{cases} \dot{x}_1 = b_1x_2 + f_1(x_1) \\ \dot{x}_2 = b_2u_P + f_2(x_2) \\ \dot{x}_3 = b_3u_Q + f_3(x_3) \end{cases} \tag{5}$$

where

$$\begin{cases} f_1(x_1) = -\frac{U_{dc}}{C_{dc}R_L} \\ f_2(x_2) = -w_g Q_g - \frac{R_z}{L_z}P_g + \frac{3}{2L_z}({e_{g\alpha}}^2 + {e_{g\beta}}^2) \\ f_3(x_3) = w_g P_g - \frac{R_z}{L_z}Q_g \end{cases} \tag{6}$$

and where $x_1 = U_{dc}$, $x_2 = P_g$, and $x_3 = Q_g$. $b_1 = \frac{1}{C_{dc}U_{dc}}$, $b_2 = -\frac{3}{2L_z}$, and $b_3 = \frac{3}{2L_z}$. $f_1(x_1)$, $f_2(x_2)$, $f_3(x_3)$, b_1, b_2, and b_3 denote the unknown nonlinear system dynamics [3].

B. Predictor-Based NN-DSC Design

First of all, the dynamic surface error is defined as $e_1 = x_1 - x_d$, whose time derivative along (5) yields

$$\dot{e}_1 = b_1 x_2 + f_1(x_1) - \dot{x}_d \tag{7}$$

where x_d denotes the desired trajectory.

It is worth mentioning that the unknown nonlinear function $f_i(x_i)$ $(i = 1, 2, 3)$ can be approximated by utilizing a NN together with an activation function $\Phi_i(x_i)$. Then, it can be described by the following form [8]:

$$f_1(x_1) = \mathrm{W}_1^{\mathrm{T}}\Phi_1(x_1) + \varepsilon_1 \tag{8}$$

where W_1 denotes the optimal weight, satisfying $\|\mathrm{W}_1\|_{\mathrm{F}} \leq \bar{\mathrm{W}}_1$ with $\bar{\mathrm{W}}_1 \in \Re$ being positive constants. ε_1 denotes the minimum approximation error. $|\varepsilon_1| \leq \bar{\varepsilon}_1$ with $\bar{\varepsilon}_1 \in \Re$ being positive constants.

To stabilize e_1, we formulate a virtual control law u_D as

$$u_D = \frac{1}{b_1}\left(-\gamma_1 e_1 - \hat{\mathrm{W}}_1^{\mathrm{T}}\Phi_1(x_1) + \dot{x}_d\right) \tag{9}$$

where γ_1 is a positive constant.

Meanwhile, a neural predictor is here deployed so as to design an update law for $\hat{\mathrm{W}}_1$ as demonstrated below [9], [10]

$$\dot{\hat{x}}_1 = \hat{\mathrm{W}}_1^{\mathrm{T}}\Phi_1(x_1) + b_1 x_2 - k_1\tilde{x}_1 \tag{10}$$

where $\tilde{x}_1 = \hat{x}_1 - x_1$. $k_1 \in \Re$ is a positive constant. $\hat{x}_1$ denotes the observed value of x_1. $\hat{\mathrm{W}}_1$ is the estimate of W_1 that is updated

$$\dot{\hat{\mathrm{W}}}_1 = -\Gamma_1[\Phi_1(x_1)\tilde{x}_1 + k_{w1}\hat{\mathrm{W}}_1] \tag{11}$$

where $\Gamma_1 \in \Re$ and $k_{w1} \in \Re$ are positive constants.

Alternatively, to avoid repeatedly differentiating, we introduce a first-order filter α_{1d}, and let u_D pass through it with time constant τ_1, i.e.,

$$\tau_1\dot{\alpha}_{1d} + \alpha_{1d} = u_D,\ \alpha_{1d}(0) = u_D(0). \tag{12}$$

By defining the output error of this filter as $e_2 = x_2 - \alpha_{1\,d}$, its time derivative yields

$$\dot{e}_2 = b_2 u_P + f_2(x_2) - \frac{u_D - \alpha_{1d}}{\tau_1}. \tag{13}$$

In view of the approximation characteristics of the NN, the nonlinear function $f_2(x_2)$ can be approximated as

$$f_2(x_2) = \mathrm{W}_2^{\mathrm{T}}\Phi_2(x_2) + \varepsilon_2 \tag{14}$$

where W_2 denotes the optimal weight matrix, satisfying $\|\mathrm{W}_2\|_{\mathrm{F}} \leq \bar{\mathrm{W}}_2$ with $\bar{\mathrm{W}}_2 \in \Re$ being positive constants. ε_2 denotes the minimum approximation error. $|\varepsilon_2| \leq \bar{\varepsilon}_2$ with $\bar{\varepsilon}_2 \in \Re$ being positive constants.

Next, a practical control law u_P is presented so as to stabilize e_2, and one has

$$u_P = \frac{1}{b_2}\left(-\gamma_2 e_2 - b_1 e_1 - \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + \frac{u_D - \alpha_{1d}}{\tau_1}\right) \tag{15}$$

where γ_2 is a positive constant.

Following a similar way, to design an update law for $\hat{\mathrm{W}}_2$, a neural predictor is addressed as

$$\dot{\hat{x}}_2 = \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + b_2 u_P - k_2\tilde{x}_2 \tag{16}$$

where $k_2 \in \Re$ is a positive constant. $\tilde{x}_2 = \hat{x}_2 - x_2$. $\hat{x}_2$ denotes the observed value. $\hat{\mathrm{W}}_2$ is the estimate of W_2 that is updated

$$\dot{\hat{\mathrm{W}}}_2 = -\Gamma_2[\Phi_2(x_2)\tilde{x}_2 + k_{w2}\hat{\mathrm{W}}_2] \tag{17}$$

where $\Gamma_2 \in \Re$ and $k_{w2} \in \Re$ are positive constants.

Let $e_3 = x_3$, its time derivative can be expressed as follows:

$$\dot{e}_3 = b_3 u_Q + f_3(x_3). \tag{18}$$

Similarly, applying a NN to approximate the nonlinear function $f_3(x_3)$ yields

$$f_3(x_3) = \mathrm{W}_3^{\mathrm{T}}\Phi_3(x_3) + \varepsilon_3 \tag{19}$$

where W_3 denotes the optimal weight matrix, satisfying $\|\mathrm{W}_3\|_{\mathrm{F}} \leq \bar{\mathrm{W}}_3$ with $\bar{\mathrm{W}}_3 \in \Re$ being positive constants. ε_3 denotes the minimum approximation error. $|\varepsilon_3| \leq \bar{\varepsilon}_3$ with $\bar{\varepsilon}_3 \in \Re$ being positive constants.

To stabilize e_3, we choose a practical control law u_Q as

$$u_Q = \frac{1}{b_3}\left(-\gamma_3 e_3 - \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3)\right) \tag{20}$$

where γ_3 is a positive constant.

To design an update law for $\hat{\mathrm{W}}_3$, a predictor is described by the following form:

$$\dot{\hat{x}}_3 = \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3) + b_3 u_Q - k_3\tilde{x}_3 \tag{21}$$

where $k_3 \in \Re$ is a positive constant. $\tilde{x}_3 = \hat{x}_3 - x_3$. $\hat{x}_3$ denotes the observed value. Let $\hat{\mathrm{W}}_3$ be an estimate of W_3 that is updated

$$\dot{\hat{\mathrm{W}}}_3 = -\Gamma_3\left[\Phi_3(x_3)\tilde{x}_3 + k_{w3}\hat{\mathrm{W}}_3\right] \tag{22}$$

where $\Gamma_3 \in \Re$ and $k_{w3} \in \Re$ are positive constants.

C. Learning-Based Robust Optimal FCS-MPC Design

To formalize the optimal control problem considered in this letter, the time derivative of the dynamic surface error can be reorganized by the following expression:

$$\dot{\boldsymbol{e}} = \boldsymbol{f} + \boldsymbol{b}\boldsymbol{u}_g \tag{23}$$

where $\boldsymbol{e} = [e_2, e_3]^{\mathrm{T}}$, $\boldsymbol{f} = [f_2(x_2) - \frac{u_D - \alpha_{1d}}{\tau_1}, f_3(x_3)]^{\mathrm{T}}$, $\boldsymbol{b} = [b_2, b_3]^{\mathrm{T}}$, and $\boldsymbol{u}_g = [u_P, u_Q]^{\mathrm{T}}$.

The control objective is to design a nonlinear robust optimal controller under uncertain environment which ensures that the

system state tracks the specified trajectory with desired accuracy, while minimizing the following performance index function [1]:

$$\mathrm{V}(\boldsymbol{e}) = \int_t^{\infty} r(\boldsymbol{e}(\tau), \boldsymbol{u}_g(\tau))d\tau \tag{24}$$

where $r(\boldsymbol{e}, \boldsymbol{u}_g) = \boldsymbol{e}^{\mathrm{T}}Q\boldsymbol{e} + \boldsymbol{u}_g^{\mathrm{T}}R\boldsymbol{u}_g$ is the utility function, and Q and R are positive-definite diagonal constant matrices.

Next, associated with (23) and (24), Hamiltonian function is generated as

$$H(\boldsymbol{e}, \boldsymbol{u}_g, \nabla\mathrm{V}(\boldsymbol{e})) = \nabla\mathrm{V}^{\mathrm{T}}(\boldsymbol{f} + \boldsymbol{b}\boldsymbol{u}_g) + \boldsymbol{e}^{\mathrm{T}}Q\boldsymbol{e} + \boldsymbol{u}_g^{\mathrm{T}}R\boldsymbol{u}_g \tag{25}$$

where $\nabla\mathrm{V}(\boldsymbol{e}) = \partial\mathrm{V}(\boldsymbol{e})/\partial\boldsymbol{e}$ is the gradient of $\mathrm{V}(\boldsymbol{e})$ with respect to $\boldsymbol{e}$.

The optimal performance index function is yielded as

$$\mathrm{V}^*(\boldsymbol{e}) = \min_{\boldsymbol{u}_g \in \Psi(\Omega)} \left(\int_t^{\infty} r(\boldsymbol{e}(\tau), \boldsymbol{u}_g(\tau))d\tau \right). \tag{26}$$

Then, (26) can be obtained by solving the HJB equation as

$$\min_{\boldsymbol{u}_g \in \Psi(\Omega)} [H(\boldsymbol{e}, \boldsymbol{u}_g, \nabla\mathrm{V}^*(\boldsymbol{e}))] = 0. \tag{27}$$

Using the fact $\partial H(\boldsymbol{e}, \boldsymbol{u}_g, \nabla\mathrm{V}^*(\boldsymbol{e}))/\partial\boldsymbol{u}_g = 0$, the optimal control policy $\boldsymbol{u}_g^*$ can be derived. We have that

$$\boldsymbol{u}_g^* = -\frac{1}{2}R^{-1}\boldsymbol{b}^{\mathrm{T}}\nabla\mathrm{V}^*(\boldsymbol{e}) \tag{28}$$

where $\nabla\mathrm{V}^*(\boldsymbol{e}) = \partial\mathrm{V}^*(\boldsymbol{e})/\partial\boldsymbol{e}$.

Remark 1: Based on the above observations, it should be pointed out that, to obtain the ideally optimal control solution (28), one has to solve the HJB equation (27) for the optimal performance index function (26). However, due to the complex nonlinearities of the HJB (27), analytically working out its solution is extremely difficult and even impossible [1], [2], [3]. Hence, in the following, we will take into account a RL-based online optimal control solution to derive the optimal controller, which is implemented by employing a critic NN.

As described above, NN has the potential to be universal approximators and can be integrated with RL for approximating functions. Thus, the optimal value function can be represented by a single-layer NN in the following equation:

$$\mathrm{V}(\boldsymbol{e}) = \mathrm{W}_c^{\mathrm{T}}\Phi_c(\boldsymbol{e}) + \varepsilon_c \tag{29}$$

where W_c denotes the optimal weight, satisfying $\|\mathrm{W}_c\|_F \leq \bar{\mathrm{W}}_c$ with $\bar{\mathrm{W}}_c \in \Re$ being positive constants. ε_c denotes the critic NN approximation error. $\|\varepsilon_c\| \leq \bar{\varepsilon}_c$ with $\bar{\varepsilon}_c \in \Re$ being positive constants.

After that, the derivative of the value function $\mathrm{V}(\boldsymbol{e})$ with respect to $\boldsymbol{e}$ is

$$\nabla\mathrm{V}(\boldsymbol{e}) = \nabla\Phi_c^{\mathrm{T}}\mathrm{W}_c + \nabla\varepsilon_c \tag{30}$$

where $\nabla\Phi_c = \partial\Phi_c(\boldsymbol{e})/\partial\boldsymbol{e}$ denotes the activation function gradient, and $\nabla\varepsilon_c = \partial\varepsilon_c(\boldsymbol{e})/\partial\boldsymbol{e}$ denotes the NN reconstruction error gradient.

Let $\hat{\mathrm{W}}_c$ be an estimate of W_c, a critic NN is here utilized to approximate the value function, and thus, we have the estimate

TABLE I
LEARNING-BASED NDSPC SOLUTION

Predictor-Based NN-DSC Design
Error dynamics
$\dot{e}_1 = b_1x_2 + f_1(x_1) - \dot{x}_d$
$\dot{e}_2 = b_2u_P + f_2(x_2) - \frac{u_D - \alpha_{1d}}{\tau_1}$
$\dot{e}_3 = b_3u_Q + f_3(x_3)$
Auxiliary first-order filter
$\tau_1\dot{\alpha}_{1d} + \alpha_{1d} = u_D,\ \alpha_{1d}(0) = u_D(0)$
Neural predictor
$\dot{\hat{x}}_1 = \hat{\mathrm{W}}_1^{\mathrm{T}}\Phi_1(x_1) + b_1x_2 - k_1\tilde{x}_1$
$\dot{\hat{x}}_2 = \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + b_2u_P - k_2\tilde{x}_2$
$\dot{\hat{x}}_3 = \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3) + b_3u_Q - k_3\tilde{x}_3$
NN update law
$\dot{\hat{\mathrm{W}}}_1 = -\Gamma_1[\Phi_1(x_1)\tilde{x}_1 + k_{w1}\hat{\mathrm{W}}_1]$
$\dot{\hat{\mathrm{W}}}_2 = -\Gamma_2[\Phi_2(x_2)\tilde{x}_2 + k_{w2}\hat{\mathrm{W}}_2]$
$\dot{\hat{\mathrm{W}}}_3 = -\Gamma_3[\Phi_3(x_3)\tilde{x}_3 + k_{w3}\hat{\mathrm{W}}_3]$
NN control law
$u_D = \frac{1}{b_1}(-\gamma_1e_1 - \hat{\mathrm{W}}_1^{\mathrm{T}}\Phi_1(x_1) + \dot{x}_d)$
$u_P = \frac{1}{b_2}(-\gamma_2e_2 - b_1e_1 - \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + \frac{u_D - \alpha_{1d}}{\tau_1})$
$u_Q = \frac{1}{b_3}(-\gamma_3e_3 - \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3))$
Learning-Based Robust Optimal FCS-MPC Design
Optimal control policy
$\boldsymbol{u}_g^* = -\frac{1}{2}R^{-1}\boldsymbol{b}^{\mathrm{T}}\nabla\mathrm{V}^*(\boldsymbol{e})$
Approximated optimal control policy
$\hat{\boldsymbol{u}}_g^* = -\frac{1}{2}R^{-1}\boldsymbol{b}^{\mathrm{T}}\nabla\Phi_c^{\mathrm{T}}\hat{\mathrm{W}}_c$
Critic NN weight update law
$\dot{\hat{\mathrm{W}}}_c = -\Gamma_c\sigma_c\left(\sigma_c^{\mathrm{T}}\hat{\mathrm{W}}_c + \boldsymbol{e}^{\mathrm{T}}Q\boldsymbol{e} + \boldsymbol{u}_g^{\mathrm{T}}R\boldsymbol{u}_g\right)$
Practical control law
$u_P^* = \frac{1}{b_2}(-\gamma_2e_2 - b_1e_1 - \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + \frac{u_D - \alpha_{1d}}{\tau_1}) + \hat{u}_{gP}^*$
$u_Q^* = \frac{1}{b_3}(-\gamma_3e_3 - \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3)) + \hat{u}_{gQ}^*$
Cost function
$C_{Fnew} = \left\|\boldsymbol{v}_{gnew}^*(k+1) - \boldsymbol{v}_g(k+1)\right\|$

of $\mathrm{V}(\boldsymbol{e})$ as

$$\hat{\mathrm{V}}(\boldsymbol{e}) = \hat{\mathrm{W}}_c^{\mathrm{T}}\Phi_c(\boldsymbol{e}). \tag{31}$$

Hence, the estimated value function gradient can be expressed as follows:

$$\nabla\hat{\mathrm{V}}(\boldsymbol{e}) = \nabla\Phi_c^{\mathrm{T}}\hat{\mathrm{W}}_c. \tag{32}$$

In view of the optimal control law in (28) and (32), the approximated optimal control policy yields:

$$\hat{\boldsymbol{u}}_g^* = -\frac{1}{2}R^{-1}\boldsymbol{b}^{\mathrm{T}}\nabla\Phi_c^{\mathrm{T}}\hat{\mathrm{W}}_c \tag{33}$$

where the critic NN weight update law can be represented by

$$\dot{\hat{\mathrm{W}}}_c = -\Gamma_c\sigma_c\left(\sigma_c^{\mathrm{T}}\hat{\mathrm{W}}_c + \boldsymbol{e}^{\mathrm{T}}Q\boldsymbol{e} + \boldsymbol{u}_g^{\mathrm{T}}R\boldsymbol{u}_g\right) \tag{34}$$

and where $\Gamma_c > 0$ is the critic NN learning rate. $\sigma_c = \sigma/(\sigma^{\mathrm{T}}\sigma + 1)$, $\sigma = \nabla\Phi_c^{\mathrm{T}}[\boldsymbol{f} + \boldsymbol{b}\boldsymbol{u}_g]$. According to the definition of σ_c, there exists a positive constant $\sigma_{c\mathrm{M}} > 1$ such that $\|\sigma_c\| \leq \sigma_{c\mathrm{M}}$.

Based on above analyses, the practical control law with respect to the approximated optimal control policy can be reconstructed as

$$\begin{cases} u_P^* = \frac{1}{b_2}\left(-\gamma_2e_2 - b_1e_1 - \hat{\mathrm{W}}_2^{\mathrm{T}}\Phi_2(x_2) + \frac{u_D - \alpha_{1d}}{\tau_1}\right) + \hat{u}_{gP}^* \\ u_Q^* = \frac{1}{b_3}(-\gamma_3e_3 - \hat{\mathrm{W}}_3^{\mathrm{T}}\Phi_3(x_3)) + \hat{u}_{gQ}^* \end{cases} \tag{35}$$

TABLE II
COMPARISON RESULTS BETWEEN SIX DIFFERENT CONTROL SCHEMES

Control Method	Execution Time (μs)			THD (%)	Average SF (Hz)	Tracking Error (%)
	Optimization	Sorting	Sum	Match/Mismatch	Match/Mismatch	Match/Mismatch
Level-based FCS-MPC [15]	25.395	19.974	45.369	1.31/3.15	1465/1134	0.817/1.860
Direct level-based MPC [16]	9.779	19.113	28.892	2.82/4.11	1091/997	1.823/2.571
Fast FCS-MPC [11]	15.193	19.446	34.639	1.61/4.94	1463/1139	0.875/2.908
PNN-based FCS-MPC [9]	31.583	18.825	50.408	1.14/2.66	1319/1394	0.643/1.391
ESO-based FCS-MPC [17]	24.505	18.549	43.054	1.46/2.97	1335/1318	0.841/1.832
Proposed learning-based NDSPC	37.534	17.781	55.315	1.10/1.33	1400/1578	0.683/0.806

where $\hat{u}^*_{gP}$ and $\hat{u}^*_{gQ}$ denotes the approximate optimal controllers given in (33), respectively.

Finally, we can rewrite the original system control inputs as

$$\begin{bmatrix} v^*_{g\alpha} \\ v^*_{g\beta} \end{bmatrix} = \begin{bmatrix} e_{g\alpha} & e_{g\beta} \\ -e_{g\beta} & e_{g\alpha} \end{bmatrix}^{-1} \begin{bmatrix} u^*_P \\ u^*_Q \end{bmatrix}. \tag{36}$$

In addition to the limitation regarding the system uncertainties, another interesting research theme is the selection of weighting factors. According to well-established knowledge, its selection is cumbersome work. Hence, to avoid this issue, an appealing alternative solution with potential to this long-standing research issue is to construct a predefined cost function without weighting factors [9]. In accordance with this, this expression (36) allows one to obtain this suggested cost function, and it yields

$$C_{Fnew} = \left| \boldsymbol{v}^*_{gnew}(k+1) - \boldsymbol{v}_g(k+1) \right|. \tag{37}$$

To illustrate our control solution clearly, the proposed learning-based NDSPC framework for power converter systems is summarized in Table I. After that, the conventional submodule (SM) capacitor voltage balancing technique in [11] is here incorporated into this proposal for the real-time implementation. Due to the page limit, for more information about the stability analysis, the readers can refer to [8], [12], [13].

III. PERFORMANCE EVALUATION

To evaluate the performance of this proposal, in this section, numerous simulation and experimental tests have been carried out using a laboratory prototype. For implementation, the real MMC prototype using Silicon Carbide Metal-Oxide-Semiconductor Field-Effect Transistors (SiC-MOSFETs) C3M0065100 K is controlled by deploying a NI PXI platform. Meanwhile, the circuit parameters are given as follows. The dc-link reference voltage is 500 V. The input filter resistance and inductance are $0.01\,\Omega$ and 10 mH, respectively. The arm filter resistance and inductance are $0.2\,\Omega$ and 5 mH, respectively. The phase voltage peak is 200 V. The load resistance is $20\,\Omega$. The number of SMs per arm is 4. The sampling/control period is $100\,\mu s$. Furthermore, to compensate the time delay of control command, a conventional delay compensation control technique in [14] is adopted in this work.

A. Simulation Result Analysis

In order to demonstrate the merits of the proposed methodology, the control performance of system is verified, as shown in Fig. 2. For a fair comparison, the same parameters with six different control solutions are chosen in our work. At first, as depicted in Fig. 2(I), the measured dc-link voltage can track its reference with good accuracy under parameter mismatch condition. Next, it can be observed from Fig. 2(II) that the control performance of the three-phase grid-side input currents in steady-state can be obtained by the six different control approaches. It is noteworthy that, under parameter mismatch condition, the total harmonic distortion (THD) values of the grid-side input currents with six different control schemes are 3.15%, 4.11%, 4.94%, 2.66%, 2.97%, and 1.33%, respectively. In this sense, the robust control performance in this proposal can be improved by up to 57.78%, 67.64%, 73.08%, 50%, and 55.22%, respectively.

Afterward, Fig. 2(III) shows the three-phase grid-side input active and reactive powers in a steady state. Furthermore, the transient characteristics of the grid-side input powers can be illustrated in Fig. 2(IV), where the input powers quickly converge to their desired references. Thus, as expected, it can be seen that the accurate tracking ability and lower power ripples by this proposal can be achieved in comparison with other existing studies. Finally, as appreciated in the figures, numerical simulation results demonstrate that the proposed learning-based NDSPC solution is capable of generating sinusoidal waveforms of the three-phase grid-side input currents, and the satisfactory robust and tracking performance can be obtained in an acceptable range. In the following, the experimental result analysis will be presented.

B. Experimental Result Analysis

To fully illustrate the interest and applicability of the proposal, the experimental results with the proposed learning-based NDSPC solution are evaluated in Fig. 3. First, the grid-side input powers and voltage/current waveforms can be depicted in Fig. 3(a). Meanwhile, Fig. 3(b) shows the experimental waveforms of the dc-link voltage and three-phase grid-side currents in a steady state. Next, the transient response of estimated input powers can be presented in Fig. 3(c)–(d), where the input powers quickly converge to their desired references. Thus, by deploying this proposal, the accurate tracking control performance can be obtained. Finally, the control performance of the dc-link voltage and SM capacitor voltages as well as circulating current is shown in Fig. 3(e). In general, as can be clearly appreciated in Fig. 3, it is further demonstrated that the desired control behaviors can be successfully fulfilled. Next, the robustness analysis will be discussed in detail.

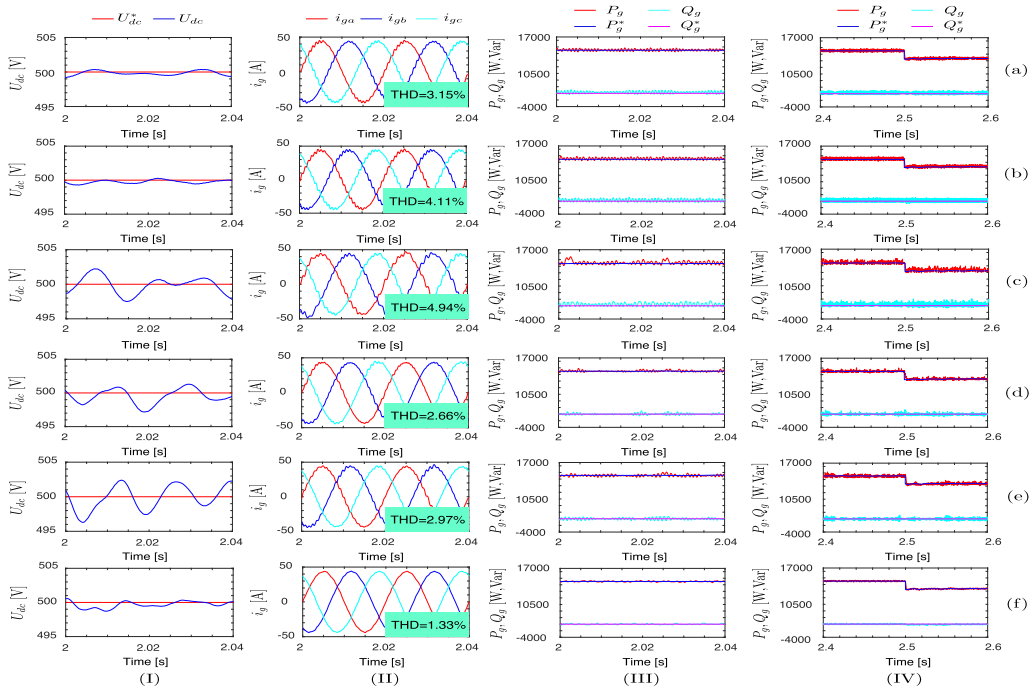

Fig. 2. The performance evaluation with six different control approaches under the parameter mismatch ($L_g = 8$ mH). (a) Level-based FCS-MPC method [15]. (b) Direct level-based MPC method [16]. (c) Fast FCS-MPC method [11]. (d) PNN-based FCS-MPC method [9]. (e) ESO-based FCS-MPC method [17]. (f) Proposed learning-based NDSPC solution.

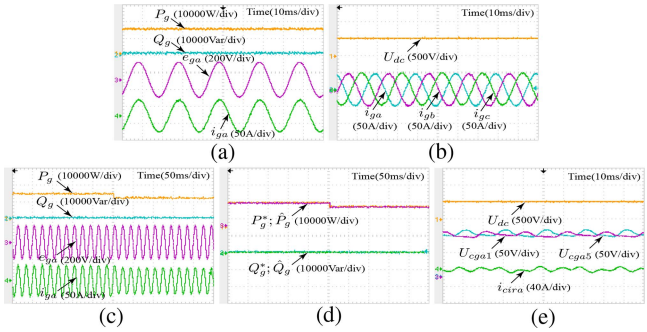

Fig. 3. Experimental results with the proposed learning-based NDSPC solution. (a) Active and reactive powers; grid-side voltage and current. (b) DC-link voltage and grid-side currents. (c) Active and reactive powers; grid-side voltage and current. (d) Estimated active and reactive powers as well as their references. (e) DC-link voltage and SM capacitor voltages as well as circulating current.

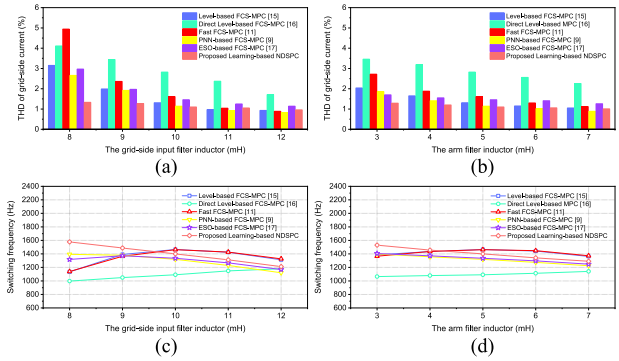

Fig. 4. Evolution of the THD and SF with six different control solutions under input/arm filter inductance mismatch. (a) THD under input filter inductance mismatch. (b) THD under arm filter inductance mismatch. (c) SF under input filter inductance mismatch. (d) SF under arm filter inductance mismatch.

C. Robustness Analysis

To further validate the robust performance of the proposed solution, a thorough analysis of the effect of parameter mismatch in term of THD and average switching frequency (SF) is evaluated. As shown in Fig. 4, with the decrease of load inductance from 10 to 8 mH, the standard model-based predictive control approaches becomes inaccurate, leading to an obvious degradation of performance. On the other hand, compared with the state-of-the-art FCS-MPC schemes, the performance of this proposal is almost unchanged under the parameter mismatch condition, allowing for strong robustness against the parametric variation.

Next, to highlight major research gap of the proposed solution with other existing MPC studies, a comparative analysis is carried out with respect to four aspects as given in Table II, including the execution time, THD, average SF, and tracking error. As illustrated in this table, different from standard model-based predictive control schemes that need an explicit physical or mathematical model of the system, the proposed solution can mitigate performance degradation caused by the parametric uncertainties. Alternatively, the execution time and average SF as well as tracking error with six different control solutions can be depicted at the same sampling/control period. Finally, our test proves the applicability of the proposal, and these results illustrate that our proposal can tolerate a wider range of uncertainties and it works as expected.

IV. Conclusion and Future Work

This letter investigates the possibility of utilizing a novel robust predictive control methodology for solving the ongoing research challenges in FCS-MPC, i.e., parametric uncertainties and external unknown disturbances. To be specific, it can be realized by integrating a RL-based online adaptive NN-DSC technique into the FCS-MPC framework. We showed that this framework can compensate the uncertainties in system models and enable the smooth and fast learning, and thus was able to enhance the robustness and tracking performance. Finally, the experimental results were provided to characterize the obtained benefits by utilizing an MMC as a case example, and it is concluded that the proposed solution can fulfill control tasks.

Admittedly, this suggested proposal is still in an early stage of development. Continued research on this solution will be required so as to achieve its full potential and verify its universality. Despite these improvements, unavoidable discrepancies between system models and real-world dynamics may lead to degradation of system performance including instability. Consequently, the robust performance of FCS-MPC in power converters is still an open topic of research that requires further attention in order to implement MPC in industrial applications. For potential future work, one topic of interest is to extend the results of this letter to various nonlinear control systems by introducing an actor-critic RL technique into FCS-MPC framework under actuator and sensor attack scenarios, which not only makes the closed-loop system asymptotically stable but also guarantees an desirable level of robust performance.

References

[1] D. R. Liu, S. Xue, B. Zhao, B. Luo, and Q. L. Wei, "Adaptive dynamic programming for control: A survey and recent advances," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 1, pp. 142–160, Jan. 2021.

[2] D. Wang, H. B. He, and D. R. Liu, "Adaptive critic nonlinear robust control: A survey," *IEEE Trans. Cybern.*, vol. 47, no. 10, pp. 3429–3451, Oct. 2017.

[3] J. J. Duan, Z. H. Yi, D. Shi, C. Lin, X. Lu, and Z. W. Wang, "Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid AC-DC microgrids," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5355–5364, Sep. 2019.

[4] S. A. A. Rizvi, A. J. Pertzborn, and Z. L. Lin, "Reinforcement learning based optimal tracking control under unmeasurable disturbances with application to HVAC systems," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, to be published, doi: 10.1109/TNNLS.2021.3085358.

[5] M. P. Ran and L. H. Xie, "Adaptive observation-based efficient reinforcement learning for uncertain systems," *IEEE Trans. Neural Netw. Learn. Syst.*, 2021, to be published, doi: 10.1109/TNNLS.2021.3070852.

[6] X. Liu et al., "Event-triggered neural predictor-based FCS-MPC for MMC," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 6433–6440, Jun. 2022.

[7] X. Liu, L. Qiu, Y. T. Fang, K. Wang, Y. D. Li, and J. Rodríguez, "A fuzzy approximation for FCS-MPC in power converters," *IEEE Trans. Power Electron.*, vol. 37, no. 8, pp. 9153–9163, Aug. 2022.

[8] D. Wang and J. Huang, "Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form," *IEEE Trans. Neural Netw.*, vol. 16, no. 1, pp. 195–202, Jan. 2005.

[9] X. Liu et al., "Predictor-based neural network finite-set predictive control for modular multilevel converter," *IEEE Trans. Ind. Electron.*, vol. 68, no. 11, pp. 11621–11627, Nov. 2021.

[10] X. Liu et al., "Data-driven neural predictors-based robust MPC for power converters," *IEEE Trans. Power Electron.*, vol. 37, no. 10, pp. 11650–11661, Oct. 2022.

[11] Z. Gong, P. Dai, X. B. Yuan, X. J. Wu, and G. S. Guo, "Design and experimental evaluation of fast model predictive control for modular multilevel converters," *IEEE Trans. Ind. Electron.*, vol. 63, no. 6, pp. 3845–3856, Jun. 2016.

[12] H. Y. Dong, X. W. Zhao, Q. L. Hu, H. Y. Yang, and P. Y. Qi, "Learning-based attitude tracking control with high-performance parameter estimation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 58, no. 3, pp. 2218–2230, Jun. 2022.

[13] G. X. Wen, C. L. P. Chen, S. Z. S. Ge, H. L. Yang, and X. G. Liu, "Optimized adaptive nonlinear tracking control using actor-critic reinforcement learning strategy," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 4969–4977,Sep. 2019.

[14] P. Cortés, J. Rodríguez, C. Silva, and A. Flores, "Delay compensation in model predictive current control of a three-phase inverter," *IEEE Trans. Ind. Electron.*, vol. 59, no. 2, pp. 1323–1325, Feb. 2012.

[15] J. Moon, J. Gwon, J. Park, D. Kang, and J. Kim, "Model predictive control with a reduced number of considered states in a modular multilevel converter for HVDC system," *IEEE Trans. Power Del.*, vol. 30, no. 2, pp. 608–617, Apr. 2015.

[16] X. Liu et al., "A fast finite-level-state model predictive control strategy for sensorless modular multilevel monverter," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 3, pp. 3570–3581, Jun. 2021.

[17] Z. F. Song, Y. J. Tian, Z. Yan, and Z. Chen, "Direct power control for three-phase two-level voltage-source rectifiers based on extended-state observation," *IEEE Trans. Ind. Electron.*, vol. 63, no. 7, pp. 4593–4603, Jul. 2016.