Multi-Agent Soft Actor-Critic Aided Active Disturbance Rejection Control of DC Solid-State Transformer

Yu Zeng, *Member, IEEE*, Gaowen Liang, *Member, IEEE*, Qingxiang Liu, *Member, IEEE*, Ezequiel Rodriguez, *Member, IEEE*, Josep Pou, *Fellow, IEEE*, Huamin Jie, *Student Member, IEEE*, Xiong Liu, *Senior Member, IEEE,* Xin Zhang, *Senior Member, IEEE,* Janardhana Kotturu, *Member, IEEE,* Amit Gupta, *Fellow, IEEE*

***Abstract*—The dc solid-state transformer (dcSST) plays a vital role in interconnecting diverse dc sources and loads in dc microgrids. However, output voltage regulation and submodule power balance control have always been two essential control challenges of the dcSST. To address these challenges, this paper proposes a multi-agent soft actor-critic-based active disturbance rejection control (MASAC-ADRC) method. The ADRC method is used for uncertainty estimation and compensation in modular dcSST. Specifically, the controller gains of the ADRC method are optimized adaptively by the MASAC method, thus enhancing the adaptability to changing conditions from the environment. Compared with the existing controller gain optimization methods, the proposed method does not rely on predetermined datasets. Instead, it provides a tailored strategy, employing a neural network to map optimal ADRC parameters from the measured states. Through the incorporation of diverse environmental scenarios, encompassing variant input voltage and output power, the MASAC-ADRC method achieves superior dynamic performance. Experimental validation underscores the efficacy of the proposed algorithm, showcasing its superiority over alternative approaches. The proposed method yields enhancements exceeding 50% in dynamic performance metrics such as overshoot, settling time, and mean square error when compared to existing methods.**

***Index Terms*—Active disturbance rejection control (ADRC), dc solid state transformer (dcSST), controller gain optimization, multi-agent soft actor-critic (MASAC).**

I. Introduction

DC microgrids have gained prominence with the rapid advancement of power electronic technologies and the large-scale integration of renewable energy systems (RESs). Battery energy storage systems (BESSs) can be a promising solution to buffer the power oscillations caused by the RESs [1]-[2]. The dual active bridge (DAB)-based dc solid-state transformer (dcSST) is a critical bidirectional dc/dc converter to interconnect RESs, BESSs, and loads with different voltage and current levels, increasing power transmission capability, and increasing redundant operation capability [3]-[4]. Fig. 1 depicts the circuit representation of the dcSST with *N* submodules (SMs) connected in series at the input and in parallel at the output.

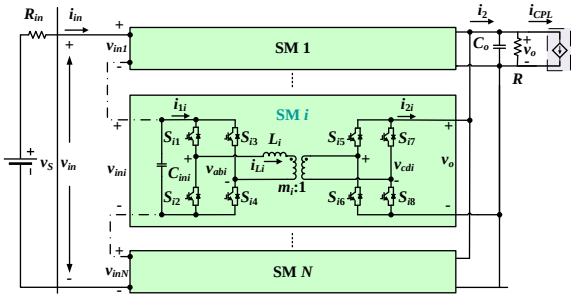


Fig. 1. dcSST in a typical dc microgrid.

The control of dcSST targets two essential objectives. The first one is to regulate the output voltage [5], [6]. The second objective is to balance power between the different SMs [7]. Proportional integral (PI)-based methods, including master-slave-based power balancing control strategies [8], [9], reshaped impedance scheme [10], input-oriented power-sharing control [11], and robust adjustable dc-link voltage control [12], have been widely utilized for dcSSTs. These approaches have shown optimal performance in simultaneously achieving power balance and regulating output voltage. However, The PI-based methods above cannot achieve fast dynamic performance [13], [14]. To improve the dynamic performance, nonlinear model predictive control (MPC) methods are proposed, which are able to balance the power distribution among different SMs [15]. A hybrid PI-MPC control scheme is introduced to decrease the current stress and improve the efficiency of the modular DAB converter [16]. In [17], the MPC controller combined with the nonlinear inductor is used to increase the power transfer capability in dc system. However, the MPC method is vulnerable to parameter variations in the system.

To control the dcSST in dc microgrids, estimating and compensating for dynamic uncertainties and disturbances is crucial. Active disturbance rejection control (ADRC) offers rapid performance, robust rejection ability, and minimal dependence on exact models. The ADRC controller consolidates uncertainties and disturbances for feedforward estimation and compensation [18], [19]. Adjusting the controller gains of the ADRC is a pivotal stage as it exerts a direct and substantial influence on the performance of the system. Given the time-consuming nature of manual tuning,

Manuscript received October 10, 2023; revised January 18, 2024, and April 4, 2024; accepted May 6, 2024. This work was supported in part by the National Research Foundation (NRF) of Singapore, Rolls-Royce Singapore Pte. Ltd., and in part by Nanyang Technological University, Singapore. (*Corresponding author: Qingxiang Liu*).

Yu Zeng, Qingxiang Liu, Josep Pou, and Huamin Jie are with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: zeng0119@e.ntu.edu.sg; qingxian001@e.ntu.edu.sg; josep.pou@ieee.org; jieh0002@e.ntu.edu.sg).

Gaowen Liang and Ezequiel Rodriguez are with the Energy Research Institute, Nanyang Technological University, Singapore 639798 (e-mail: gaowen001@e.ntu.edu.sg; ezequiel001@e.ntu.edu.sg).

Xiong Liu is with the Energy Electricity Research Center, International Energy College, Jinan University, Zhuhai 519070, China. (e-mail: liushawn123@ieee.org).

Xin Zhang is with the College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China (email: zhangxin_ieee@zju.edu.cn).

Janardhana Kotturu, and Amit Kumar Gupta are with the Rolls-Royce Singapore Private Limited, Singapore 638673 (e-mail: janardhana.kotturu@rolls-royce-electrical.com; amit.gupta@ieee.org).

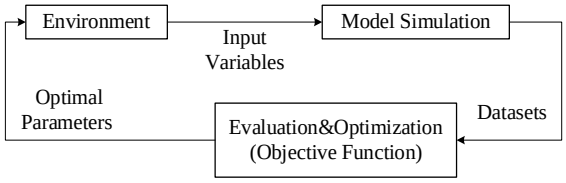


Fig. 2. Previous parameters optimization algorithms (PSO, ACO, and ANN).

there is a significant demand for an accurate automatic tuning algorithm. Tuning methods inspired by biological behaviors, such as ant colony optimization (ACO) [20], and particle swarm optimization (PSO) [21], offer promising solutions for determining controller gains. However, they suffer from premature convergence and local optimum problems.

In [22], an artificial-neural-network-based ADRC (ANN-ADRC) method is proposed to estimate the uncertainties and disturbances adaptively under all operating points to improve dynamic performance. However, a large amount of high-quality accurate training data is needed for these supervised learning-based algorithms [22]-[23].

As shown in Fig. 2, the existing ACO, PSO, and ANN-based algorithms require a predetermined dataset and output the constant control gains for specific operating conditions. These algorithms generally yield static control laws with poor adaptability and robustness to varying microgrid environments such as variant input voltages and output power [24]. Besides, the learning process is tedious and time-consuming. The optimal dynamic performance, such as rising time and overshoot, is difficult to achieve for power converters across various operating points [25].

To solve the issues above, deep reinforcement learning (DRL) methods are attractive candidates to optimize the controller gains of the modular converter combining the advantages of robust optimization ability and less dependence on models [26]-[27]. To incorporate environmental scenarios and provide a tailored strategy for environmental adaptive controller parameters, the multi-agent deep reinforcement learning (MADRL) method is proposed in this paper to collaboratively tune the ADRC parameters of multiple SMs of the dcSST in real-time. Due to high sampling efficiency and little influence by hyperparameters, multi-agent soft actor-critic (MASAC) is used in this paper [28]-[29]. Compared with the existing ACO, PSO, and ANN-based ADRC methods above, the proposed MASAC-ADRC algorithm provides significant advantages in terms of adaptability, and generalization, particularly in complex and dynamic environments such as dc microgrids. The detailed comparisons and discussions are summarized in Section IV.

The contributions of this paper include:

1)A MASAC-ADRC method is proposed for the dcSST for the first time. This method incorporates prior physical knowledge and environmental scenarios of the dcSST into the training of neural networks, providing physical boundary and termination signals of the dcSST, exhibiting faster convergence.

2)*The proposed MASAC-ADRC algorithm learns from interactions with the environment rather than relying on predetermined datasets, yielding inherent adaptability.* The MASAC-ADRC method responds effectively to changing conditions and manages uncertainties more efficiently compared to traditional parameter optimization techniques.

3)*The proposed method offers a tailored strategy (a neural network) rather than specific optimal parameters.* The trained neural network effectively serves as a rapid surrogate model to map the optimal ADRC parameters from varying environment information quickly and accurately.

4)*This proposed method can play a guiding role in designing environmental adaptive controller parameters*. Adaptive optimal control parameters can be easily set to the modular converters according to the application scenarios of different users according to the proposed method.

The remainder of this paper is structured as follows. Section II details the architecture of the ADRC controller design. In Section III, the MASAC-based adaptive parameter tuning for the ADRC controller is introduced. Section IV discusses the hardware design and validates the proposed methodology through hardware experiments. Finally, Section V concludes the paper.

II. ADRC Controller Design for the DCSST

This section begins with a comprehensive review of the mathematical model for the dcSST. Subsequently, an ADRC controller is developed utilizing the model above. To guarantee the robustness and stability of the ADRC controller, detailed computations and derivations are performed to determine the stability constraints.

A. Mathematical Model for dcSST

Fig. 1 shows the schematic of the dcSST in the dc microgrid. The CPLs can be modeled as:

$$i_{CPL} = \frac{p_{CPL}}{v_o} \tag{1}$$

where p_{CPL} is the power of the CPL and v_o is the output voltage.

As shown in Fig. 1, the system comprises two full-bridge circuits accompanied by a high-frequency transformer in each DAB SM i, (i=1, 2... N) [30]-[31]. The duty ratio of the square voltage waveforms at the primary and secondary sides of the high-frequency transformer, v_{abi} and v_{cdi} respectively, are fixed as 50% in each SM. Specifically, phase-shifted square wave modulation is considered, given its simplicity and fast dynamic response. For each DAB SM i, C_{ini} and C_o are the input and the output capacitances of SM i of the dcSST, respectively.

For SM i of the dcSST, the voltage drop across the inductor L_i corresponds to the difference between the voltage on the primary side, v_{abi}, and the voltage on the secondary side, v_{cdi}, which generates current i_{Li}. The transmission power of SM i corresponds to:

$$p_i = \frac{m_i d_i (1-d_i) v_{ini} v_o}{2 f_s L_i} \tag{2}$$

where f_s is the switching frequency of the DAB, v_o refers to the output voltage across capacitor C_o. m_i, L_i and v_{ini} are the transformer turns ratio, the ac-link inductance, and the input voltage across capacitor C_{ini}, respectively. d_i is the phase shift ratio between v_{abi} and v_{cdi}. Note that $0 \le d_i \le 1$.

The averaged and decoupled model of the dcSST can be given by (3) [9], [32]:

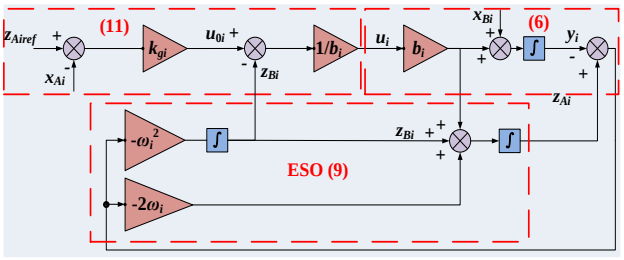


Fig. 3. Block diagram of the ADRC for the SM *i* of the dcSST.

$$\begin{cases} \dfrac{dv_{ini}}{dt} = \dfrac{g_{in}u_i}{R_{in}C_{ini}} - \dfrac{v_{ini}}{R_{in}C_{ini}}, & i = 1,2,...,N-1 \\ \vdots \\ \dfrac{dv_o}{dt} = \dfrac{g_o u_N}{C_o} - \dfrac{v_o}{RC_o} - \dfrac{p_{CPL}}{C_o v_o} \end{cases} \tag{3}$$

where g_{in} and g_o are given in (4), while virtual control variables u_i are given in (5) [9].

$$\begin{cases} g_{in} = \dfrac{V_o}{V_{in}} g_o \\ g_o = \dfrac{V_o(1-2D_i)}{(1-D_i)D_i R} \\ V_{in} = V_s - \dfrac{m_i R_{in} V_o D_i (1-D_i)}{2 f_s L_i} \end{cases} \tag{4}$$

where V_s and R_{in} are the Thevenin equivalent of the source voltage and input resistance of the source, respectively.

Assuming N=3, u_i is given as [9]:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \tag{5}$$

B. ADRC Controller Design

The core concept of the ADRC involves employing an extended state observer (ESO) to both estimate and compensate for the total disturbances f_i. The state-space representation can be articulated as [19]:

$$\begin{cases} \dfrac{dx_{Ai}}{dt} = x_{Bi} + b_i u_i \\ \dfrac{dx_{Bi}}{dt} = \dfrac{df_i}{dt} \\ y_i = x_{Ai} \end{cases} \tag{6}$$

Considering x_{Ai}= {v_{in1}, v_{in2}, v_o}, x_{Bi}={f_1, f_2, f_3}, and comparing (6) and (3), the generalized disturbance f_i corresponds to (7):

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} -\dfrac{1}{R_{in}C_{in1}} v_{in1} \\ -\dfrac{1}{R_{in}C_{in2}} v_{in2} \\ -\dfrac{1}{RC_o} v_o - \dfrac{1}{C_o v_o} p_{CPL} \end{bmatrix} \tag{7}$$

b_i corresponds to (8):

$$\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix} = \begin{bmatrix} \dfrac{g_{in}}{R_{in}C_{in1}} \\ \dfrac{g_{in}}{R_{in}C_{in2}} \\ \dfrac{g_o}{C_o} \end{bmatrix} \tag{8}$$

Fig. 3 illustrates the ADRC scheme for the dcSST. For the three decoupled first-order systems, the ESO is required to estimate only one state variable and one lumped disturbance variable. Using the bandwidth parametrization method for the observer gains [22], the corresponding ESO is constructed as:

$$\begin{bmatrix} \dfrac{dz_{Ai}}{dt} \\ \dfrac{dz_{Bi}}{dt} \end{bmatrix} = \begin{bmatrix} -l_{Ai} & 1 \\ -l_{Bi} & 0 \end{bmatrix} \begin{bmatrix} z_{Ai} \\ z_{Bi} \end{bmatrix} + \begin{bmatrix} b_i & l_{Ai} \\ 0 & l_{Bi} \end{bmatrix} \begin{bmatrix} u_i \\ y_i \end{bmatrix}, \quad \omega_i > 0 \tag{9}$$

where z_{Ai} and z_{Bi} provide the estimates of x_{Ai} and x_{Bi}, respectively, and l_{Ai} and l_{Bi} can be represented as:

$$\begin{cases} l_{Ai} = 2\omega_i \\ l_{Bi} = \omega_i^2 \end{cases}, \quad \omega_i > 0 \tag{10}$$

where ω_i is the observer gain.

The control law is:

$$\begin{cases} u_i = (u_{0i} - z_{Bi}) / b_i \\ u_{0i} = k_{gi}(z_{Airef} - z_{Ai}) \approx k_{gi}(z_{Airef} - x_{Ai}) \end{cases} \tag{11}$$

where k_{gi} represents the feedback controller gain, and z_{Airef} represents the reference value {v_{in1ref}, v_{in2ref}, v_{oref}}.

C. Stability Constraints of ESO

The system's stability is contingent upon the time step h. Equation (6) can be transformed into its discrete form as follows:

$$\begin{cases} x_{Ai}(n+1) = x_{Ai}(n) + h x_{Bi}(n) + h b_i u_i(n) \\ x_{Bi}(n+1) = x_{Bi}(n) \end{cases} \tag{12}$$

By utilizing a zero-order holder and assuming a sufficiently small h, the expression for x_{Ai} in the z domain is:

$$\begin{cases} z_{Ai}(n+1) = z_{Ai}(n) + h l_{Ai}(x_{Ai}(n) - z_{Ai}(n)) + h z_{Bi}(n) + h b_i u_i(n) \\ z_{Bi}(n+1) = z_{Bi}(n) + h l_{Bi}(x_{Ai}(n) - z_{Ai}(n)) \end{cases} \tag{13}$$

Let $e_{Ai}(n)$=$x_{Ai}(n)$-$z_{Ai}(n)$, $e_{Bi}(n)$=$x_{Bi}(n)$-$z_{Bi}(n)$ be the observer errors, then from (12) and (13), we can obtain:

$$\begin{bmatrix} e_{Ai}(n+1) \\ e_{Bi}(n+1) \end{bmatrix} = \begin{bmatrix} 1 - h l_{Ai} & h \\ -h l_{Bi} & 1 \end{bmatrix} \begin{bmatrix} e_{Ai}(n) \\ e_{Bi}(n) \end{bmatrix} = A \begin{bmatrix} e_{Ai}(n) \\ e_{Bi}(n) \end{bmatrix} \tag{14}$$

The characteristic equation of A is:

$$z^2 - (2 - l_{Ai}h)z + 1 - l_{Ai}h + l_{Bi}h^2 = 0 \tag{15}$$

According to the Routh stability criterion and taking into account that ADRC gain k_{gi} should be bounded between 1/5 and 1/2 of w_i for sufficient ESO bandwidth, the following set of inequalities are derived:

$$\begin{cases} (\omega_i h - 2)^2 > 0 \\ 4\omega_i h - 2\omega_i^2 h^2 > 0 \\ \omega_i^2 h^2 > 0 \\ \dfrac{1}{5}\omega_i \le k_{gi} \le \dfrac{1}{2}\omega_i \end{cases} \tag{16}$$

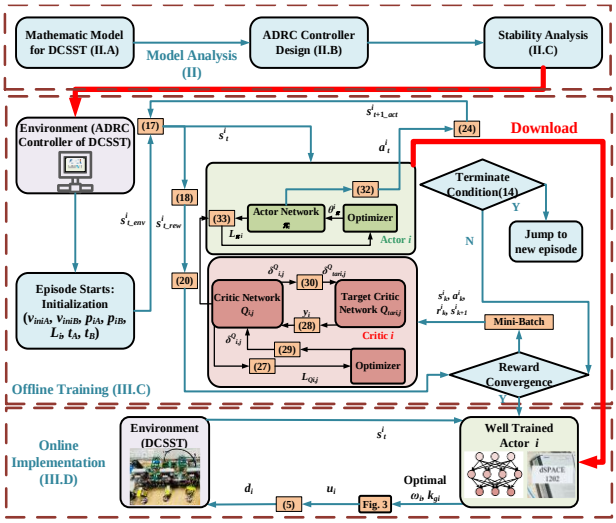


Fig. 4. MASAC-ADRC parameters tuning algorithms.

In the next section, inequalities (16) are taken into account in the training of the multiple agents. This prior physical information improves the training speed and enhances the controller performance.

III. MASAC-Based ADRC Controller

This section starts with an overview of the utilization of the proposed MASAC methods for the adaptive tuning of ADRC controllers. Subsequently, a comprehensive exposition of Markov Games is provided to address the challenges of parameter tuning for modular converters. Following this, the process of centralized training for the MASAC method is elucidated. Finally, the decentralized implementation stage is introduced.

A. Overview of the proposed MASAC-ADRC method

Fig. 4 shows an overview of the proposed MASAC-ADRC method. The inputs, or environment, are initialized at the start of every episode, and actor networks are optimized to output the best actions. Inequalities (16) are considered in the termination condition. The controller parameters $\{\omega_i, k_{gi}\}$ that need to be optimized are referred to as action states $s^i_{t\_act}$, while the actions a^i_t are the corrections added to ADRC parameters, and are denoted as $\{\Delta\omega_i, \Delta k_{gi}\}$. At the beginning of every episode, the DRL Agent *i* receives an environmental state $s^i_{t\_env}$ for initialization. According to the reward state $s^i_{t\_rew}$, the Agent *i* receives a reward r^i_t based on the reward function, which will be defined next. The relevant hyperparameters are summarized in Table I.

At time step *t*, each Agent *i* receives a local state s^i_t from the environment. Upon receiving the current state s^i_t, the Agent *i* will subsequently select an action a^i_t to be performed next. Upon acting a^i_t, the Agent *i* will receive a reward function r^i_t. Following the state transition distribution, the Agent *i* can observe the new state s^i_{t+1}.

The converter model outlined in Section II acts as the environment for the MASAC. Within this framework, *N* agents work collaboratively to fine-tune the controller gain combinations $\{\omega_i, k_{gi}\}$ to ensure power balance and voltage stability.

TABLE I
HYPERPARAMETERS FOR THE PROPOSED MASAC ALGORITHM

Parameter	Value
Maximum episode number, *M*	2000
Learning rate for the critic, l_r^c	0.0001
Learning rate for the actor, l_r^a	0.001
Discount factor, γ	0.98
Target smooth factor, τ	1e-3
Mini batch size, *B*	64
Temperature parameter, α	1.5e-3
Replay buffer size, N_B	1e6
Average window length, *L*	10

As illustrated in Fig. 4, the overall design methodology contains model analysis, offline training, and online implementation. In model analysis stage, a mathematical model of the DCSST is built; Then, the ADRC controller is designed to estimate and compensate for the uncertainties and disturbances of the DCSST; The next step is the stability analysis, the stability constraints of the controller gains $\{\omega_i, k_{gi}\}$ are determined.

The second stage is the offline training stage, all agents collaboratively adjust their actions based on the real-time state s^i_t. The deep neural networks for all agents are trained simultaneously, and the parameters of all agents are updated to maximize the reward function, which will be elaborated on later. Throughout each learning iteration, agents explore a range of actions.

Upon completion of the training process, the actor neural networks for all agents are deployed in a decentralized manner in online implementation stage.

As indicated in Fig. 4, these finely tuned actor neural networks are downloaded to the microcontroller dSPACE 1202. Within the microcontroller, every agent generates the optimal phase shift ratio combinations $\{d_1, d_2, d_3\}$ for the different SMs based on varying states. The overarching goal for all agents is to regulate the output voltage, and equitably distribute the power among different SMs.

B. Formulation of Markov Games in MASAC

Markov Games are an effective solution to solve control gain parameters optimization problems in this work. In Markov Games, the notations *S*, *A*, and *R* commonly represent the state, the action, and the reward space, respectively. The descriptions of these spaces are provided as follows:

1) State Space S: $s^i_t \in S$ represents the states of all agents. For Agent *i*, s^i_t is

$$s^i_t = \left\{ s^i_{t\_env}, s^i_{t\_act}, s^i_{t\_rew} \right\} \tag{17}$$

where $s^i_{t\_env}$, $s^i_{t\_act}$, $s^i_{t\_rew}$ can be expressed as:

$$\begin{cases} s^i_{t\_env} = \{v_{ini}, p_i, L_i\} \\ s^i_{t\_act} = \{\omega_i, k_{g_i}\} \\ s^i_{t\_rew} = \{e_{vin\_i}, \int e_{vin\_i} dt, e_{vo\_i}, \int e_{vo\_i} dt\} \end{cases} \tag{18}$$

where e_{vin_i} and e_{vo_i} are the input and output voltage errors in the *i-th* SM, respectively, i.e.,

$$\begin{cases} e_{vin\_i} = | v_{ini} - v_{iniref} | \\ e_{vo\_i} = | v_{oi} - v_{oiref} | \end{cases} \tag{19}$$

TABLE II
SYSTEM PARAMETERS OF THE DCSST

Nominal Conditions			
Nominal source voltage, V_S	150 V	Nominal transmission power, P	150 W
Reference input voltage of SM i, v_{iniref}	v_{in}/N	Reference output voltage, V_{oref}	50 V
Turn ratios, m_i	1:1	Ac-link inductance, L_i	100 μH
Number of SMs, N	3	Switching frequency, f_s	20 kHz
Boundary Conditions			
Maximum input voltage, V_{inmax}	165 V	Minimum input voltage, V_{inmin}	135 V
Maximum transmission power, P_{max}	300 W	Minimum transmission power, P_{min}	0
Maximum ac-link inductor, L_{max}	110 μH	Minimum ac-link inductor, L_{min}	90 μH
Modulation Variables Range			
$0 \le d_i \le 1$			
Operating Condition Ranges			
$P_{min} \le p_{iA} \le P_{max}$		$L_{min} \le L_i \le L_{max}$	$P_{min} \le p_{iB} \le P_{max}$
$V_{inmin} \le v_{iniA} \le V_{inmax}$			$V_{inmin} \le v_{iniB} \le V_{inmax}$

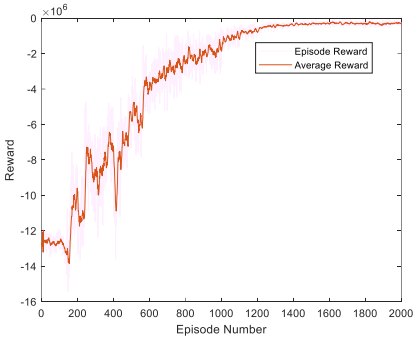


Fig. 5. Episode reward and average reward for Agent 3 using the proposed MASAC controller.

2) Action Space A: $a_t^i \in A$ comprises the actions of all agents. The next state s_{t+1}^i is determined by the current state s_t^i and the current action a_t^i: $\{\Delta\omega_i, \Delta k_{gi}\}$.

3) Reward Space R: $r_t^i \in R$ is the instant reward for agent i. Performance indicators such as overshoot and settling time are analyzed quantitatively.

When the value r_t^i is the largest, the optimization result is the best. r_t^i can be expressed as:

$$r_t^i = r_t^{vin\_i} + r_t^{vo\_i} + r_t^{d\_i} \tag{20}$$

The hybrid reward comprises the output voltage control reward $r_t^{vo\_i}$, the local input voltage error reward $r_t^{vin\_i}$, and the dynamic performance reward $r_t^{d\_i}$. These rewards are defined by equations (21), (22), and (23), respectively,

$$r_t^{vo\_i} = -0.1e_{vo\_i}^2 - 10\left(IF\left(\left|e_{vo\_i}\right| > 1\right)\right) - \left(IF\left(\left|e_{vo\_i}\right| > 0.1\right)\right) \tag{21}$$

$$r_t^{vin\_i} = -0.1e_{vin\_i}^2 - 10\left(IF\left(\left|e_{vin\_i}\right| > 1\right)\right) - \left(IF\left(\left|e_{vin\_i}\right| > 0.1\right)\right) \tag{22}$$

$$r_t^{d\_i} = -(t_v^i + t_s^i + \sigma^i / 100) \tag{23}$$

where the logical function IF outputs "1" if the comparison is "True", and "0" otherwise. t_v^i, t_s^i, and σ^i are the settling time of v_{ini} when v_{ini} changes from v_{iniA} to v_{iniB}, the settling time of v_{ini} when p_i changes from p_{iA} to p_{iB}, and the voltage overshoot of v_o when p_i changes from p_{iA} to p_{iB}, respectively.

4) State transition

The state transition process is:

$$\begin{cases} a_t^i = \pi(s_{t\_act}^i) \\ s_{t+1\_act}^i = a_t^i + s_{t\_act}^i \end{cases} \tag{24}$$

where every agent maintains a replay buffer. For every time step t, the transition experience $\{s_t^i, a_t^i, r_t^i, s_{t+1}^i\}$ is stored in the replay buffer of Agent i.

5) Termination signal

a) Protection Termination. For Agent i, the termination triggers when p_i and v_{ini} are outside the safe ranges $[P_{min}, P_{max}]$ and $[V_{inmin}, V_{inmax}]$, respectively, with P_{min}, P_{max}, V_{inmin}, V_{inmax} prescribed safe limits for the processed power and input voltage of the SM i given in Table II. This over-power and over-voltage protection will guarantee the system is safe.

b) Stability Termination. Similarly, the termination will trigger when stability regions in (16) are violated. The termination signal is designed to end an episode if the agent deviates significantly from its intended goal.

C. Centralized Training of the MASAC Algorithm

In this paper, the MASAC algorithm is proposed by modelling each SM as an agent. Table I presents the hyperparameters. The discount factor is set firstly. A high discount factor will make the agent focus on short-term rewards, while a low discount factor will make the agent focus on long-term rewards. Based on large simulation comparisons, the discount factor γ is set as 0.98. Then the target smooth factor τ is set. The target smooth factor, also known as the exploration factor or epsilon-greedy factor, influences the exploration and exploitation trade-off in deep reinforcement learning. Based on large simulations, the target smooth factor is set as 1e-3. Then we set the learning rate l_r^a and l_r^c. A high learning rate will make the agent learn quickly, but it may overshoot the optimal policy. A low learning rate will make the agent learn slowly, but it may converge to a sub-optimal policy. As the SAC agent is based on actor-critic architecture, the learning rate for the actor l_r^a and the learning rate for the critic l_r^c can be set as 0.001 and 0.0001, respectively.

Finally, other hyperparameters such as the replay batch size N_B, mini-batch size B, episode numbers M, temperature parameter α, and average window length L are simply adjusted via monitoring the performance of the agent.

For this actor-critic based Agent i, the Critic i comprises four approximation functions: j-th critic function of Agent i Q_i=$\{Q_{i,j}(s_t^i, a_t^i|\delta_{i,j}^Q)$ parameterized by $\delta_{i,j}^Q$, and j-th target critic function of Agent i $Q_{tari,j}(s_t^i, a_t^i|\delta_{tari,j}^Q)\}$ parameterized by $\delta_{tari,j}^Q (j=1,2)$. Meanwhile, Actor i comprises one approximation function $\pi_i(s_t^i|\theta_i^\pi)$ parameterized by θ_i^π. The proposed offline training algorithm can be summarized in three main steps: 1) Environment interaction and data collection. 2)policy evaluation, and 3) policy improvement.

1) Environment interaction and data collection

In this step, different scenarios $\{v_{iniA}, p_{iA}, v_{iniB}, p_{iB}, L_i, t_A, t_B\}$ are randomly chosen as the current training environment. The input voltage v_{ini} changes from v_{iniA} to v_{iniB} at t_A, and the output power reference changes from p_{iA} to p_{iB} at t_B. t_A is set as 0.3s and t_B is set as 0.5s. At the beginning of each training episode, various input parameters $s_{t\_env}^i = \{v_{ini}, p_i, L_i\}$ are initialized. The

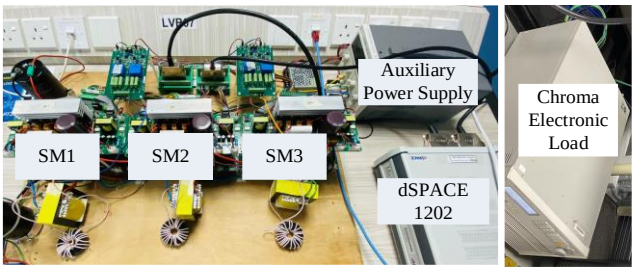


Fig. 6. Overview of the experimental setup of the dcSST.

range of these input parameters, {v_{ini}, p_i, L_i} is detailed in Table II. The states s_t^i of each agent are updated according to (24). The stochasticity and uncertainty can aid MASAC agents in identifying optimal controller gains under diverse input conditions.

2) Policy evaluation

As shown in red block in Fig. 4, the collected data is used to update the agents' value functions $Q_{i,j}$. These functions estimate the expected reward for a given state-action pair and state. The agent typically uses a neural network to approximate these functions, and the weights and biases of the network are updated using gradient descent with a loss function derived from the Bellman equation.

Based on the soft Bellman iteration, $Q_{i,j}$ is given:

$$Q_{i,j}(s_t^i, a_t^i) = r_t^i + \gamma E_{s_{t+1}^i, a_{t+1}^i \sim \pi_{\theta_i^\pi}}[Q_{i,j}(s_{t+1}^i, a_{t+1}^i)] + \gamma \cdot \alpha \cdot H\left(\pi(a_{t+1}^i \mid s_{t+1}^i)\right) \quad (25)$$

where α and γ are the temperature parameter, and the discount factor, respectively, as in Table I. The entropy $H(\cdot)$ can be given:

$$H\left(\pi(a_t^i \mid s_t^i)\right) = E_{s_t^i, a_t^i \sim \pi_{\theta_i^\pi}}[-\log\left(\pi_{\theta_i^\pi}(a_t^i \mid s_t^i)\right)] \quad (26)$$

where E[.] is the mathematical expectation, and $\pi_{\theta_i^\pi}(a_t^i|s_t^i)$ is the actor function of Agent i.

In neural network training, optimization typically involves computing the gradient for a randomly selected mini-batch {s_k^i, a_k^i, r_k^i, s_{k+1}^i} (k=1, 2…, B), drawn from the replay buffer. By utilizing a mini-batch, the parameters of each critic can be updated by minimizing the loss function $L_{Qi,j}$ of the Critic i,

$$L_{Q_{i,j}} = \frac{1}{B}\sum_{k=1}^{B}[y_k^i - Q_{i,j}(s_k^i, a_k^i \mid \delta_{i,j}^Q)]^2, j = 1,2 \quad (27)$$

where y_k^i represents the value function target of the Critic i, and it corresponds to,

$$y_k^i = r_k^i + \gamma E_{s_{k+1}^i, a_{k+1}^i \sim \pi_{\theta_i^\pi}}\left[\min_{j=1,2}\left(Q_{tari,j}(s_{k+1}^i, a_{k+1}^i \mid \delta_{tari,j}^Q)\right) + \alpha H\left(\pi(a_{k+1}^i \mid s_{k+1}^i)\right)\right] \quad (28)$$

Based on the gradient rule to achieve optimization, $\delta_{i,j}^Q$ can be given:

$$\delta_{i,j}^Q \leftarrow \delta_{i,j}^Q + lr^c \nabla_{\delta_{i,j}^Q} L_{Q_{i,j}}(\delta_{i,j}^Q), j = 1,2 \quad (29)$$

where lr^c is the learning rate for critic networks. Using target smooth factor τ, $\delta_{tari,j}^Q$ can be given:

$$\delta_{tari,j}^Q \leftarrow \tau\delta_{i,j}^Q + (1-\tau)\delta_{tari,j}^Q, j = 1,2 \quad (30)$$

3) Policy improvement

As shown in the green block in Fig. 4, the agent updates its policy (the actor) using the learned value functions (the critic). The goal is to optimize the policy so that it maximizes the expected return while maintaining sufficient exploration, which is encouraged by the entropy term. The updated policy is derived using the Kullback-Leibler divergence approach:

$$\pi_i = \arg\min_{\pi_i \in \Pi} D_{KL}\left(\pi_{\theta_i^\pi}(a_t^i \mid s_t^i) \,\|\, \frac{\exp(\frac{1}{\alpha}Q_{i,j}(s_t^i, a_t^i \mid \delta_{i,j}^Q))}{Z(s_t^i)}\right) \quad (31)$$

where $Z(s_t^i)$, $D_{KL}(\cdot)$, and Π are the logarithm partition function, the feasible arrays, and the Kullback-Leibler divergence, respectively.

The action a_t^i is reparametrize using the mean $\pi_{\theta_i^\pi}^{\mu}(a_{t+1}^i|s_{t+1}^i)$, and the standard deviation $\pi_{\theta_i^\pi}^{\sigma}(a_{t+1}^i|s_{t+1}^i)$ of a Gaussian distribution. Defining τ_t as an input noise vector, a_t^i is given by:

$$a_t^i = \pi_{\theta_i^\pi}^{\mu}(a_{t+1}^i \mid s_{t+1}^i) + \tau_t \cdot \pi_{\theta_i^\pi}^{\sigma}(a_{t+1}^i \mid s_{t+1}^i) \quad (32)$$

The policy loss function $L_{\pi i}$ for Actor i is given by:

$$L_{\pi_i} = E_{s_t \sim D, a_t \sim \pi_{\theta_i^\pi}}[\log \pi_{\theta_i^\pi}(a_k^i \mid s_k^i) - \frac{1}{\alpha}Q_{i,j}(s_k^i, a_k^i \mid \delta_{i,j}^Q)] \quad (33)$$

The parameters of actor neural networks are optimized using the gradient rule as in (31), i.e.,

$$\theta_i^\pi \leftarrow \theta_i^\pi + lr^a \nabla_{\theta_i^\pi} L_{\pi i}(\theta_i^\pi) \quad (34)$$

where lr^a represents the learning rate for actor networks.

Table I enumerates the primary parameters of the training process. The training process will terminate once the maximum number of episodes, M, is reached. The evolution of the episode rewards and the average rewards for Agent 3 during the training progress is shown in Fig. 5. The training time is 48.12 h. According to Fig. 5, the whole training process can be divided into three stages, namely the exploration stage (about 0 to 580 episodes), the learning stage (about 580 to 1030 episodes), and the converging stage (about 1030 to 2000 episodes).

D. Decentralized Implementation Stage

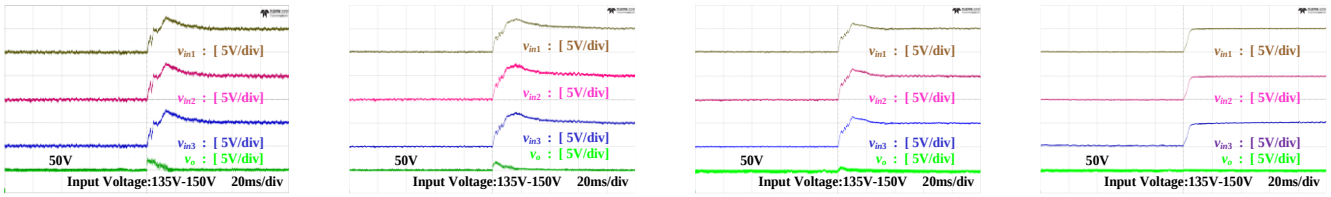
After determining the optimal parameters for the deep neural network during training, each agent is executed on a hardware prototype in a decentralized manner. It is important to emphasize that all agents are deployed to produce optimal actions in different environments. Utilizing only local information, each agent oversees an SM of the dcSST to ensure power balance and regulate output voltage.

IV. Hardware Experiments

To further validate and assess the controller's performance in a real environment, the proposed MASAC algorithm was implemented in a dcSST hardware prototype, as depicted in Fig. 6. The dSPACE DS1202, equipped with an internal Xilinx FPGA, serves as the digital control board, generating gate signals for all the SMs of the dcSST. The circuit parameters of the dcSST are presented in Table II. Different comparison experiments including input voltage variations, transmission power variations, power balance variations, and parameter mismatches, are implemented to validate the performance of the proposed MASAC controller. Mean square errors (MSEs) of the voltage are used to evaluate the disturbance rejection performance, where a lower MSE means a better disturbance rejection performance.

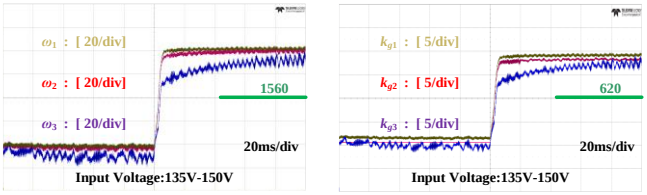
$$\begin{cases} MSE(v_{ini}) = \frac{1}{T}\sum_{t=1}^{T} e_{vin\_i}^2(t) \\ MSE(v_o) = \frac{1}{T}\sum_{t=1}^{T} e_{vo}^2(t) \end{cases} \quad (35)$$

A. Scenario A: Comparison Experiments of Input Voltage Variations with other ADRC Methods



(a) (b) (c) (d)

Fig. 7. Experimental voltage waveforms when the input voltage increases from 135 V to 150 V (a) ACO-ADRC, (b) PSO-ADRC, (c) ANN-ADRC, and (d) MASAC-ADRC.



(a) (b)

Fig. 8. Controller Parameters $\{\omega_i, k_{gi}\}$ of the proposed MASAC-ADRC controller when the input voltage increases from 135 V to 150 V (a) ω_i, (b) k_{gi}.

TABLE III
COMPARISON OF THE DIFFERENT EXPERIMENTAL METRICS OBTAINED DURING THE TRANSIENT SCENARIO A (FIG. 7) FOR DIFFERENT ADRC-BASED CONTROL SCHEMES

References	[20]	[21]	[22]	This paper
Control schemes	ACO-ADRC	PSO-ADRC	ANN-ADRC	MASAC-ADRC
Controller parameters	w_1:1386.7 w_2:1401.2 w_3:1365.3 k_{g1}:523.4 k_{g2}:518.3 k_{g3}:526.2	w_1:1435.3 w_2:1398.6 w_3:1468.5 k_{g1}:555.4 k_{g2}:576.9 k_{g3}:568.7	w_1:1492.3 w_2:1511.5 w_3:1476.6 k_{g1}:593.6 k_{g2}:601.5 k_{g3}:578.6	w_1:[1522.3,1601.8] w_2:[1509.8,1602.5 w_3:[1498.6,1599.2] k_{g1}: [609.2,630.1] k_{g2}: [608.9,629.8] k_{g3}: [607.5,631.1]
Dynamic performance	Low	Medium	Medium	High
Overshoot of v_o, σ_{vo}	6.094%	4.236%	2.114%	0.643%
Settling time of v_{in1}, t_{vin1}(ms)	34.974	31.045	11.253	6.455
MSE of v_{in1}, MSE_{vin1}	0.462	0.485	0.248	0.189

To thoroughly assess the benefits of the proposed controller, benchmark comparisons are made against the ACO-ADRC [20], PSO-ADRC [21], and ANN-ADRC controllers [22]. The three benchmark controllers above use the constant controller parameters $\{\omega_i, k_{gi}\}$, which have been optimized using the nominal operating conditions and parameters listed in the first three rows of Table II, and whose values are shown in Table III.

The controller parameters $\{\omega_i, k_{gi}\}$ of the proposed MASAC-ADRC method are shown in Fig. 8. From Fig. 8, we can observe that the proposed method offers a tailored strategy (a neural network) to generate adaptive parameters.

In Scenario A, the initial input voltage is set at 135 V, followed by an increase to 150 V. A comparison of the performance of various control algorithms is depicted in Fig. 7, with performance indicators summarized in Table III. Compared to the overshoot in output voltage, σ_{vo}, obtained with static control laws ACO-ADRC [20], PSO-ADRC [21], and ANN-ADRC [22], the proposed MASAC with dynamic control gains shows improvements of 89.4%, 84.8%, and 69.6%, respectively.

The experimental results demonstrate that the output voltage can be well regulated, and the power balance among the three SMs can be effectively achieved for the proposed MASAC-ADRC controller during the voltage increase. These comparative results highlight that the proposed controller can achieve optimal dynamic and static performances.

TABLE IV
COMPARATIVE EXPERIMENTAL RESULTS BETWEEN DIFFERENT ADRC-BASED SCHEMES UNDER POWER TRANSITION IN FIG. 9

References	[20]	[21]	[22]	This paper
Control schemes	ACO-ADRC	PSO-ADRC	ANN-ADRC	MASAC-ADRC
Overshoot of v_o, σ_{vo}	6.989%	5.313%	3.989%	2.613%
Settling time of v_{in2}, t_{vin2}(ms)	17.535	21.687	14.236	1.351
Mean square error of v_{in2}, MSE_{vin2}	0.412	0.389	0.215	0.069

TABLE V
COMPARISON EXPERIMENTS BETWEEN DIFFERENT SCHEMES UNDER PARAMETERS MISMATCH AND POWER BALANCE VARIATIONS

References	[9]	[15]	[22]	This paper
Control schemes	PI	MPC	ANN-ADRC	MASAC-ADRC
Execution time	1.6 μs	10.2 μs	7.3 μs	7.8 μs
Overshoot of v_o, σ_{vo}	9.4 %	3.2%	3.1%	2.8%
Settling time of v_{in1}, t_{vin1}(ms)	96.967	66.676	59.168	8.364
Mean square error of v_{in1}, MSE_{vin1}	2.363	1.304	1.002	0.317

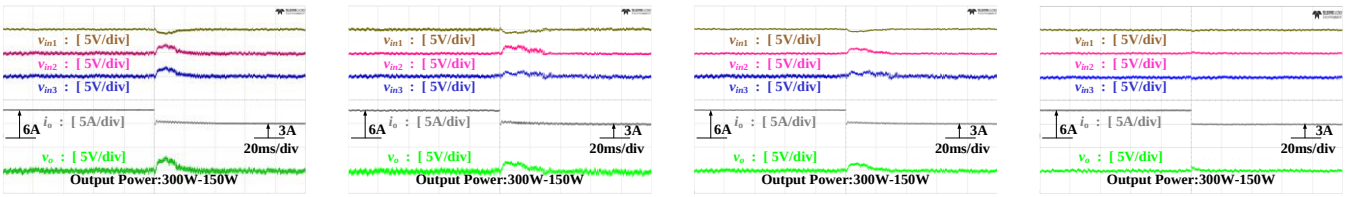
B. Scenario B: Comparison of Transmission Power Variations with Other ADRC Methods

Fig. 9 presents the experimental results corresponding to the decrease in transmission power of the dcSST. Benchmark comparisons are conducted with the ACO-ADRC controller, the PSO-ADRC controller, and the ANN-ADRC controllers. The input voltage is set at v_{in}=150 V, while the output voltage reference is set to v_{oref}=50 V.

As illustrated in Fig. 9, compared with the ACO-ADRC, PSO-ADRC, and ANN-ADRC controllers, the improvements of MSE of v_{in2}, MSE_{vin2} of MASAC-ADRC controllers are 83.3%, 82.2%, and 67.9%. The experimental results demonstrate the excellent dynamic performance and disturbance rejection capabilities of the proposed MASAC-ADRC controller. The waveforms displayed in Fig. 9 validate that the output voltage can be effectively regulated, and the power among different SMs can be well-balanced when the power decreases.

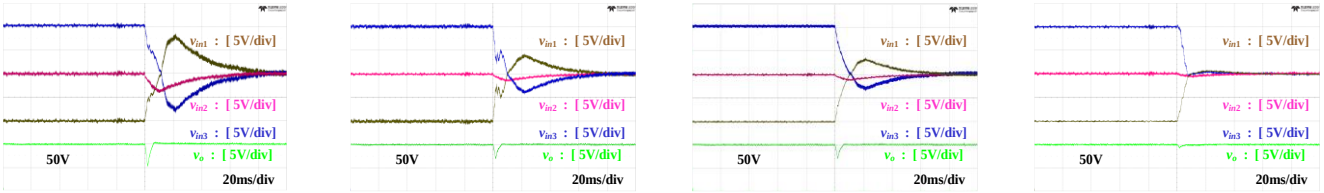
C. Scenario C: Comparison Experiments of Power Balance Variations and Parameter Mismatches with Other Optimal Methods

A new experimental scenario, not previously encountered during training, is introduced as Scenario *C*. The aim is to validate that the proposed controller has superior dynamic performance and disturbance rejection ability.



(a) (b) (c) (d)

Fig. 9. Experimental voltage and current waveforms for a power transition from 300 W to 150 W. (a) ACO-ADRC, (b) PSO-ADRC, (c) ANN-ADRC, and (d) MASAC-ADRC.



(a) (b) (c) (d)

Fig. 10. Experimental results under the 10% LC parameters mismatch and power balance variations. (a) PI, (b) MPC, (c) ANN-ADRC, and (d) Proposed MASAC-ADRC.

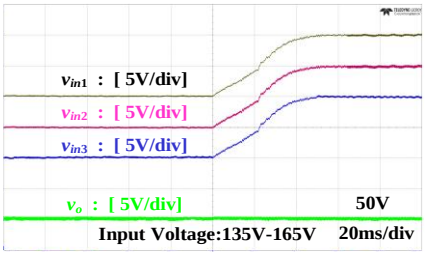


Fig. 11. Experimental waveforms corresponding to Scenario D.1 (300 W constant power load and input voltage variation from 135 V to 165 V).

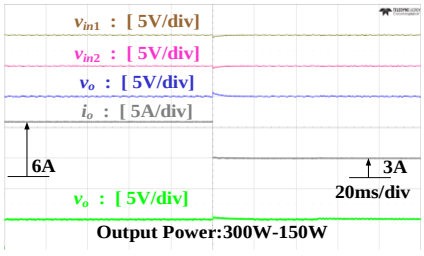


Fig. 12. Experimental waveforms corresponding to Scenario D.2 (150 V input voltage and output power variation from 300 W to 150 W).

A comparative experiment is conducted under inductor parameter mismatch and power balance variations. The benchmarks used for comparison include the PI-based IVS-controlled method [9], the MPC controller, and the ANN-ADRC controller [22]. v_{oref} is set as 50 V and p is set as 300 W. The circuit parameters used in the controller design exhibit a ±10% error, i.e., L_{k1} is 110 µH, and L_{k2} is 90 µH rather than the values presented in Table II.

Initially, the system operates under an unbalanced condition, with the target power distribution performance for the three SMs set to a 4:5:6 ratio. When the power sharing performance reference is altered to a 1:1:1 ratio, the experimental results for transient behavior are depicted in Fig. 10. The corresponding performance indicators in terms of overshoot/undershoot, settling time, and MSEs are consolidated in Table V.

According to Fig. 10 and Table V, the proposed controller exhibits the best dynamic performance in tracking voltage references. Considering the settling time of v_{in1}, t_{vin1}, the proposed MASAC-ADRC controller has an improvement of 91.4%, 87.5%, and 85.9% respectively, compared with the PI, and the ANN-ADRC controller. This demonstrates the enhanced dynamic performance and disturbance rejection capability of the proposed MASAC-ADRC controller.

D. Scenario D: Validation experiments when training conditions are not in the training range

When L_{k1} is 120 µH, and L_{k2} is 80 µH, which are not in the training range, the validation experiments are demonstrated below.

In Scenario D.1, the output power is 300 W. When the input voltage increases from 135 V to 165 V, the corresponding experimental results are shown in Fig. 11. σ_{vo}, t_{vin1}, and MSE_{vin1} are 0.785%, 6.983 ms, and 0.245, respectively. The experimental results demonstrate that the output voltage can be well-regulated, and the power balance among the three SMs can be effectively achieved for the proposed MASAC-ADRC controller during the voltage increase. This result illustrates how the proposed controller can provide satisfactory results under an operating condition that falls outside the considered training range.

In Scenario D.2, when the output power decreases from 300 W to 150 W, the experimental results are shown in Fig. 12. σ_{vo}, t_{vin1}, and MSE_{vin1} are 2.627%, 1.578 ms, and 0.174, respectively.

The experimental results demonstrate the excellent dynamic performance and disturbance rejection capabilities of the proposed MASAC-ADRC controller. The waveforms displayed in Fig. 12 validate that the output voltage can be effectively regulated, and the power among different SMs can be well-balanced when the power decreases.

V. Conclusion

In this paper, a novel MASAC approach has been proposed for real-time parameter optimization of the ADRC controller in dcSST. The proposed methodology effectively addresses output voltage regulation issues and power balance problems under the uncertainties of dc microgrid. By leveraging prior physical knowledge and environmental scenarios, the MASAC

demonstrates faster convergence, inherent adaptability, and efficient rejection of uncertainties. The developed neural network acts as a rapid surrogate model, mapping optimal ADRC parameters based on varying environmental conditions. This approach facilitates the design of environmentally adaptive controller parameters, enabling modular converters to employ different control parameters for specific user application scenarios. According to the comparison experiments, the proposed algorithm shows more than 73.8% improvements in terms of σ_{vo} compared with the ACO-ADRC, PSO-ADRC, and the ANN-ADRC controller when the input voltage rises from 135 V to 150 V. In addition, the MASAC-ADRC controllers show more than 86.8% improvements in tracking voltage references compared with other existing methods under variations of power balance among different SMs.

REFERENCES

[1] S. Kouro, M. Malinowski, K. Gopakumar, J. Pou, L. G. Franquelo, B. Wu, J. Rodriguez, M. Perez, and J. Leon, "Recent advances and industrial applications of multilevel converters," *IEEE Trans. Ind. Electron.*, vol. 57, no. 8, pp. 2553–2580, Aug. 2010.

[2] G. Liang, E. Rodriguez, Y. Zeng, G. Farivar, H. Lam, H. Yuan, and J. Pou, "Reactive power distribution in the cascaded h-bridge converter with unbalanced submodule active power for capacitance reduction," *IEEE Trans. Ind. Electron.*, in press: DOI 10.1109/TIE.2024.3363737.

[3] G. Farivar, W. Manalastas, H. Tafti, S. Ceballos, A. Sanchez-Ruiz, E. Lovell, G. Konstantinou, C. Townsend, M. Srinivasan, and J. Pou, "Grid-connected energy storage systems : state-of-the-art and emerging technologies," *Proc. IEEE*, 2021, in press: DOI 10.1109/JPROC.2022.3183289.

[4] Y. Zhang, L. Ding, N. Hou, and Y. Li, "A direct actual-power control scheme for current-fed dual-active-bridge dc/dc converter based on virtual impedance estimation," *IEEE Trans. Power Electron.*, vol. 37, no. 8, pp. 8963-8975, Aug. 2022.

[5] X. Meng, Y. Jia, Q. Xu, C. Ren, X. Han, and P. Wang, "A novel intelligent nonlinear controller for dual active bridge converter with constant power loads," *IEEE Trans. Ind. Electron.*, 2022, in press: DOI 10.1109/TIE.2022.3170608.

[6] J. Shen, J. Zhang, X. Huang, Q. Lin, and Y. Fang, "Active thermal management method for output-parallel dab dc-dc converters under parameter mismatches and asymmetrical," *IEEE Trans. Power Electron.*, 2022, in press: DOI 10.1109/TPEL.2023.3266287.

[7] D. Ma, W. Chen, and X. Ruan, "A review of voltage/current sharing techniques for series-parallel-connected modular power conversion systems," *IEEE Trans. Power Electron.*, vol. 35, no. 11, pp. 12383–12400, Nov. 2020.

[8] C. Luo and S. Huang, "Novel voltage balancing control strategy for dual-active-bridge input-series-output-parallel dc-dc converters," *IEEE Access*, vol. 8, pp. 103114–103123, 2020.

[9] P. Zumel, L. Ortega, A. Lázaro, C. Fernández, A. Barrado, A. Rodríguez, and M. Hernando, "Modular dual-active bridge converter architecture," *IEEE Trans. Ind. Appl.*, vol. 52, no. 3, pp. 2444–2455, Feb. 2016.

[10] Y. Wang, Y. Guan, O. B. Fosso, M. Molinas, S. Z. Chen, and Y. Zhang, "An input-voltage-sharing control strategy of input-series-output-parallel isolated bidirectional dc/dc converter for dc distribution network," *IEEE Trans. Power Electron.*, vol. 37, no. 2, pp. 1592–1604, Feb. 2022.

[11] N. Hou, P. Gunawardena, X. Wu, L. Ding, Y. Zhang and Y. W. Li, "An input-oriented power sharing control scheme with fast-dynamic response for ISOP DAB dc–dc converter," *IEEE Trans. Power Electron.*, vol. 37, no. 6, pp. 6501–6510, Jun. 2022.

[12] N. Hou, L. Ding, P. Gunawardena, T. Wang, Y. Zhang, and Y. W. Li, "A partial power processing structure embedding renewable energy source and energy storage element for islanded DC microgrid," *IEEE Trans. Power Electron.*, vol. 38, no. 3, pp. 4027-4039, Mar. 2023.

[13] C. Cui, T. Yang, Y. Dai, C. Zhang, and Q. Xu, "Implementation of transferring reinforcement learning for dc-dc buck converter control via duty ratio mapping, " *IEEE Trans. Ind. Electron.*, 2022, in press: DOI 10.1109/TIE.2022.3192676.

[14] J. He, X. Zhang, H. Ma, and C. Cai, "Lyapunov-based large-signal control of three-phase stand-alone inverters with inherent dual control loops and load disturbance adaptivity," *IEEE Trans. Ind. Electron.*, vol. 1, no. 1, pp. 1-10, May 2021.

[15] H. Zhang, Y. Li, Z. Li, C. Zhao, F. Gao, Y. Hu, L. Long, K. Luan, and P. Wang, "Model predictive control of input-series output-parallel dual active bridge converters based dc transformer," *IET Power Electron.*, vol. 13, no. 6, pp. 1144–1152, Feb. 2020.

[16] F. An, W. Song, B. Yu, and K. Yang, "Model predictive control with power self-balancing of the output parallel DAB dc–dc converters in power electronic traction transformer," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 6, no. 4, pp. 1806–1818, Dec. 2018.

[17] H. Lin, H. S. -H. Chung, R. Shen and Y. Xiang, "Enhancing stability of dc cascaded systems with CPLs using MPC combined with NI and accounting for parameter uncertainties," in *IEEE Transactions on Power Electronics*, in press: DOI 10.1109/TPEL.2024.3359672.

[18] Y. Zuo, J. Chen, X. Zhu, C. H. T. Lee, and S. Member, "Different active disturbance rejection controllers based on the same order gpi observer," vol. 69, no. 11, pp. 10969–10983, 2022.

[19] J. Han, "From PID to active disturbance rejection control," *IEEE Trans. Ind. Electron.*, vol. 56, no. 3, pp. 900-906, Mar. 2009.Han, "From PID to active disturbance rejection control," *IEEE Trans. Ind. Electron.*, vol. 56, no. 3, pp. 900–906, 2009.

[20] Z. Yin, C. Du, J. Liu, X. Sun, and Y. Zhong, "Research on auto disturbance-rejection control of induction motors based on an ant colony optimization algorithm," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 3077–3094, Apr. 2018.

[21] Y. Zeng, X. Zhang, S. Mukherjee, A. K. Gupta, C. Sun, and J. Dong, "Adaptive Active Disturbance Rejection Control of DAB Based on PSO," IECON Proc. (Industrial Electron. Conf., vol. 2020-Octob, pp. 2840–2845, 2020.

[22] Y. Zeng, A. Maswood, J. Pou, X. Zhang, Z. Li, C. Sun, S. Mukherjee, A. Gupta, and J.Dong "Active disturbance rejection control using artificial neural network for dual-active-bridge-based energy storage system," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, 2021, in press: DOI 10.1109/JESTPE.2021.3138341.

[23] Y. Tang, etal, "Deep reinforcement learning aided variable-frequency triple-phase-shift control for dual-active-bridge converter," *IEEE Trans. Ind. Electron.*, vol. 70, no. 10, pp. 10506-10515, Oct. 2023.

[24] Y. Zeng, J. Pou, C. Sun, S. Mukherjee, X. Xu, A. Gupta and, J.Dong, "Autonomous input voltage sharing control and triple phase shift modulation method for ISOP-DAB converter in DC microgrid: A multiagent deep reinforcement learning-based method," *IEEE Trans. Power Electron.* vol. 38, no. 3, pp. 2985-3000, Mar. 2023.

[25] S. Jiang, Y. Zeng, Y. Zhu, J. Pou, and K. Georgios, "Stability-oriented multi-objective control design for power converters assisted by deep reinforcement learning", *IEEE Trans. Power Electron.*, pp. 2023, to be published, doi: 10.1109/TPEL.2023.3299979.

[26] S. Zhao, F. Blaabjerg and H. Wang, "An Overview of Artificial Intelligence Applications for Power Electronics," *IEEE Trans. Power Electron.*, vol. 36, no. 4, pp. 4633-4658, Apr. 2021.

[27] Y. Wang, S. Fang, J. Hu, and D. Huang, "Multi-scenarios parameter optimization method for active disturbance rejection control of PMSM based on deep reinforcement learning," *IEEE Trans. Ind. Electron.*, 2022, in press: DOI 10.1109/TIE.2022.3225829.

[28] Y. Zeng, J. Pou, C. Sun, A. Maswood, J. Dong, S. Mukherjee, and A. Gupta, "Multi-agent deep reinforcement learning-aided output current sharing control for input-series output-parallel dual active bridge converter," *IEEE Trans. Power Electron.*, vol. 37, no. 11, pp. 12955-12964, Nov. 2022.

[29] Y. Zeng *et al*., "A physics-informed pattern recognition method for open-circuit fault detection of inverters under unexpected conditions," in *Proc. 46th Annu. Conf. IEEE Ind. Electron. Soc.* (*IECON*), Nov. 2023, pp. 1-5, doi: 10.1109/IECON51785.2023.10311617.

[30] Y. Zeng, A. Maswood, J. Pou, X. Zhang, Z. Li, C. Sun, J.Dong, S. Mukherjee, and A. Gupta, "Deep reinforcement learning based input voltage sharing method for input-series ouput-parallel dual active bridge converter in dc microgrids," in *2021 IEEE Energy Conversion Congress and Exposition* (*ECCE*), Oct. 2021, pp. 3348-3352.

[31] C. Sun, X. Zhang, J. Zhang, M. Zhu, and J. Huang, "Hybrid ISOS modular dc/dc converter constituted by resonant and non-resonant DAB modules," *IEEE Trans. Ind. Electron.*, vol. 69, no. 1, pp. 1062–1069, Feb. 2021.

[32] Y. Zeng, J. Pou, C. Sun, X. Li, G. Liang, Y. Xia, S. Mukherjee, and A. Gupta, "Deep reinforcement learning-enabled distributed uniform control for a dc solid state transformer in dc microgrid," *IEEE Trans. Ind. Electron.*, pp. 2023, to be published, doi: 10.1109/TIE.2023.3294584.

Yu Zeng (Member, IEEE) received the B.S. and M.E. degrees in electrical engineering from Shandong University, China, in 2017 and 2019 respectively, and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore, in 2023. He is currently a Research Fellow at NTU. His research interests include power electronics, artificial intelligence, smart grids, and electric aircraft.

Gaowen Liang (Member, IEEE) received the B.Sc. degree in electrical engineering and automation from the South China University of Technology, Guangzhou, China, in 2018, and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore, in 2022. He is currently a Research Fellow at Energy Research Institute at Nanyang Technological University (ERI@N), Singapore.

His research interest includes the multilevel converters, energy storage systems, renewable energy systems, and smart grid. He is an Editorial Board Member of Chinese Journal of Electrical Engineering. He received the Second Prize Paper Award as the first author at IEEE ECCE in 2022. In 2023, he received the National Scholarship for Outstanding Self-financed Students Abroad from China Scholarship Council.

Qingxiang Liu (Member, IEEE) received the B.Sc. degree in electrical engineering from Wuhan University, Wuhan, China, in 2018, followed by the M.Sc. degree in power engineering and the Ph.D. degree in power electronics from Nanyang Technological University, Singapore, in 2019 and 2024, respectively. Currently, he serves as a Research Fellow at the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

His research interests include modulation and control of power converters, multilevel converters, FACTS devices, and electric vehicles.

He was awarded the Professional Engineers Board Gold Medal in Singapore in 2019.

Ezequiel Rodriguez (Member, IEEE) was born in Tarragona, Spain, in 1994. He received the bachelor's degree in electrical engineering and the master's degree in engineering and technology of electronic systems (topping the 2012 and 2016 graduating cohorts as valedictorian) from Universitat Rovira i Virgili, Tarragona, Spain, in 2016 and 2017, respectively, and the Ph.D. degree in electrical engineering from Nanyang Technological University (NTU), Singapore, in 2022.

He is currently a Postdoctoral Research Fellow with the Energy Research Institute at NTU (ERI@N), Singapore. In addition, he is the co-Director of the Power Electronics and Applications Research Lab at NTU (PEARL@NTU), Singapore. His research interests include control of power electronics converters, with an emphasis on modular multilevel converters for energy storage, solar and FACTS applications. Dr. Ezequiel is the recipient of the 2022 Best Thesis Award by the School of Electrical and Electronic Engineering, NTU.

Josep Pou (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Technical University of Catalonia (UPC)-Barcelona Tech, in 1989, 1996, and 2002, respectively.

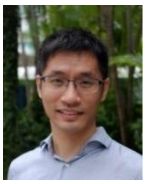
In 1990, he joined the faculty of UPC as an Assistant Professor, where he became an Associate Professor in 1993. From February 2013 to August 2016, he was a Professor with the University of New South Wales (UNSW), Sydney, Australia. He is currently a Professor with the Nanyang Technological University (NTU), Singapore, where he is Cluster Director of Power Electronics at the Energy Research Institute at NTU (ERI@N) and co-Director of the Rolls-Royce @ NTU Corporate Lab. From February 2001 to January 2002, and February 2005 to January 2006, he was a Researcher at the Center for Power Electronics Systems, Virginia Tech, Blacksburg. From January 2012 to January 2013, he was a Visiting Professor at the Australian Energy Research Institute, UNSW, Sydney. He has authored more than 480 published technical papers and has been involved in several industrial projects and educational programs in the fields of power electronics and systems. His research interests include modulation and control of power converters, multilevel converters, renewable energy, energy storage, power quality, HVdc transmission systems, and more-electrical aircraft and vessels.

He is Associate Editor of the *IEEE Journal of Emerging and Selected Topics in Power Electronics*. He was co-Editor-in-Chief and Associate Editor of the *IEEE Transactions on Industrial Electronics*. He received the 2018 IEEE Bimal Bose Award for Industrial Electronics Applications in Energy Systems.

Huamin Jie (Student Member, IEEE) received the B.Eng. degree in electrical engineering from the Wuhan University, Wuhan, China, in 2019, and the M.Sc. degree in power engineering from Nanyang Technological University, Singapore, in 2020. He is currently working toward the Ph.D. degree in electrical engineering with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore.

His research interests include device modeling, electromagnetic interference (EMI), EMI filter design, fault detection, impedance measurement, intentional EMI, and power converter systems. He was the recipient of the Best Paper Awards at the 2022 Asia-Pacific International Symposium on Electromagnetic Compatibility (APEMC) and the 2023 International Conference on Sensing, Measurement, Communication, and Internet of Things Technologies (SMC-IoT).

Xiong Liu (Senior Member, IEEE) received the B.E.and M.Sc. degrees in electrical engineering from Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2008, respectively, and the Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore in 2013. From December 2013 to July 2020, He was working as a Principal Technologist in Rolls-Royce Electrical, Rolls-Royce Singapore Pte. Ltd., Singapore. He is currently an Associate Professor with the Energy Electricity Research Center, International Energy College, Jinan University, Zhuhai, China. His research interests include power electronics, motor drive, and electrical/hybrid propulsion system for marine and aerospace.



Xin Zhang (Senior Member, IEEE) received the Ph.D. degree in Automatic Control and Systems Engineering from the University of Sheffield, U.K., in 2016 and the Ph.D. degree in EEE from Nanjing University of Aeronautics & Astronautics, China, in 2014. From 2017 to 2020, he was an Assistant Professor with EEE, Nanyang Technological University, Singapore. Currently, he is the Professor at Zhejiang University. He services as the AE for IEEE TIE/JESTPE/OJPE, etc. He is interested in stability and AI application in power electronics system.

Janardhana Kotturu (Member, IEEE) received M.Tech and Ph.D. degree in Electrical Engineering from Indian Institute of Technology Roorkee, India, in 2011 and 2019, respectively. He worked as an Assistant Professor at LPU University, Punjab, India from 2011 to 2012. He worked with Rolls-Royce@NTU Corporate Laboratory, Singapore as a Research Scientist from 2018 to 2022. Currently, he is a Senior Power Electronics Engineer at Rolls-Royce Singapore Pte Ltd and jointly working with Rolls-Royce@NTU Corporate Laboratory, Singapore. He dedicates himself to design and development of power electronics converters for medium and high-power applications. His main research interests include high-power density converters, power electronics packaging, isolated and non-isolated DC/DC converters, DC/AC converters, and active power filters.

Amit K. Gupta (Fellow, IEEE) holds a bachelor's degree in electrical engineering from the Indian Institute of Technology (IIT)-Roorkee and a Ph.D. in Electrical Engineering from National University of Singapore (NUS). During 2000-12, he worked for Bechtel Corporation, Samsung Heavy Industries, Delphi Automotive Systems and Vestas Wind Systems. Since August 2012, he is Head of Rolls-Royce Electrical at Rolls-Royce Singapore Pte Ltd. He is Director of the Electrical Program at Rolls-Royce @ NTU Corporate lab and Rolls-Royce Director for the Electrical Power System Integration Lab @ NTU (EPSIL@N).