RESEARCH ARTICLE

Deep Reinforcement Learning Based Control of a Grid Connected Inverter With LCL-Filter for Renewable Solar Applications

ANUGULA RAJAMALLAIAH[1], (Graduate Student Member, IEEE),
SRI PHANI KRISHNA KARRI[1], (Member, IEEE), **MAMDOUH L. ALGHAYTHI**[2], (Member, IEEE),
AND MESHARI S. ALSHAMMARI[2], (Member, IEEE)

[1]National Institute of Technology Andhra Pradesh, Tadepalligudem, Andhra Pradesh 534101, India
[2]Department of Electrical Engineering, College of Engineering, Jouf University, Sakaka, Aljouf 72388, Saudi Arabia

Corresponding author: Anugula Rajamallaiah (arajamallaiah@gmail.com)

ABSTRACT This research paper presents a novel approach to current control in Grid-Connected Inverters (GCI) using Deep Reinforcement Learning (DRL) based Twin Delayed Deep Deterministic Policy Gradient (TD3) method. The study focuses on addressing the limitations of traditional control techniques and state of the art techniques, particularly Proportional-Integral (PI) control and Model Predictive Control (MPC), by leveraging the adaptive and autonomous learning capabilities of DRL. The proposed novel modified TD3-based DRL method learns an optimal control policy directly from raw data, enabling the controller to adapt and improve its performance in real-time. The research includes a comprehensive analysis of the implementation and validation of the modified TD3-based DRL control in a grid-connected three phase three level Neutral Point Clamped (NPC) inverter system with Inductor-Capacitor-Inductor (LCL) filter. Real-time validation experiments are conducted to evaluate the control performance, power transfer capability in grid compliance. Furthermore, a detailed comparison is presented with experimentation, highlighting the advantages of the TD3-based DRL control over PI and MPC control techniques. Robustness checking is performed under various operating conditions, including parameter variations and dynamic conditions in the grid. The results analysis demonstrates that the TD3-based DRL control outperforms traditional PI control techniques in terms of static, dynamic response, and robustness. Additionally, The DRL based grid connected inverter current control method is validated in Renewable Energy Source (RES) solar PV grid integration application.
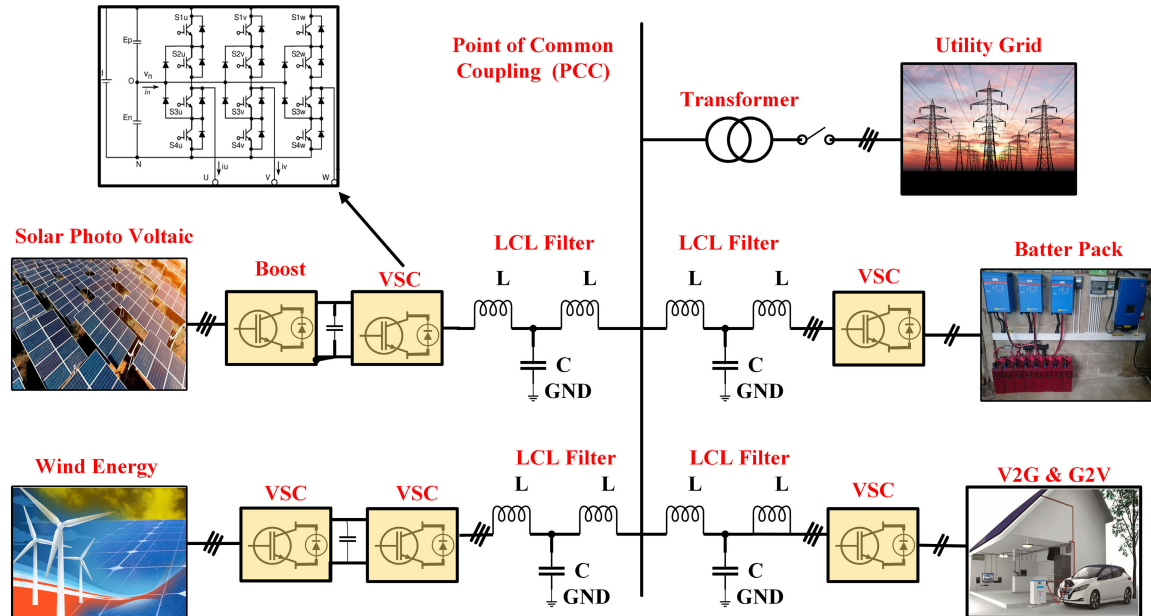
INDEX TERMS Current control technique, deep reinforcement learning, grid-connected inverters, model predictive control, PI control, real-time validation, twin delayed deep deterministic policy gradient method, robustness.

I. INTRODUCTION

The increasing adoption of renewable energy sources, such as solar panels and wind turbines, has led to a growing interest in grid-connected inverters. These inverters play a crucial role in efficiently converting DC power from renewable sources into AC power that is compatible with the grid [1], [2], [3]. Among various inverter configurations, the three-level NPC

The associate editor coordinating the review of this manuscript and approving it for publication was Alfeu J. Sguarezi Filho .

topology has emerged as a preferred choice over traditional two-level inverters for medium-voltage applications. This preference is primarily due to its advantages, including lower switching losses and reduced Total Harmonic Distortion (THD) [4]. In Figure 1, the three-level NPC converter is widely used in grid-connected applications such as renewable energy systems, electric vehicle charging, and industrial motor drives due to its superior performance and efficiency. For instance, they can be connected directly to medium voltage systems without the need for a power transformer,

FIGURE 1. Three phase three level NPC converter with LCL filter is considered as popular configuration in many grid interfacing applications.

and they can operate at a relatively high effective switching frequency, which reduces the THD in the grid/load [5]. Furthermore, compared to a conventional two-level VSI, an NPC inverter typically requires a smaller power filter for a similar application. Additionally, the NPC converter offers lower dv/dt values, higher efficiency, and less strain on the components, which are all considered significant benefits over the conventional two-level topology.

To ensure enhanced performance and grid compliance, the utilization of LCL filter configuration has gained prominence in grid-connected inverters. The LCL filter provides reduced size, superior harmonic filtering capabilities, mitigates resonance issues, enhances stability and robustness, reduces current ripple, and effectively suppresses electromagnetic interference [6], [7], [8].

In the context of grid-connected inverters, effective current control plays a vital role in managing power flow and mitigating harmonics [9], [10]. The inverter adjusts its current output based on the power demand from the grid or the available power from renewable energy sources. Precise control of the injected current facilitates efficient power transfer, ensuring optimal operating conditions while delivering generated energy to the grid [11]. Moreover, grid-connected inverters must limit the presence of harmonics in the injected current to maintain a clean and sinusoidal waveform, thereby enhancing power quality and reducing potential disturbances [12].

DPC enables direct control of active and reactive power flow between the converter and the grid [13], [14]. By minimizing the power error between desired power references and measured actual power, DPC adjusts control signals such as voltage vector selection or switching pattern to achieve the desired power sharing [15], [16]. Traditionally, PI control

is widely employed for current control in GCIs. PI control regulates the converter current by comparing the reference current (obtained from the power control algorithm) with the actual current feedback. The PI controller adjusts the control signals, such as modulation index or switching frequency, to minimize current error and achieve the desired active and reactive power sharing with the grid [18], [19]. However, PI controllers have limitations in terms of dynamic response, adaptability to varying system conditions or disturbances, and handling non-linearities in the system, where as GCI demands faster response for current control when it is used in nested loop. While PI controllers are practical for many control scenarios due to their simplicity, robustness, and cost-effectiveness, more advanced control techniques are required to meet stringent performance requirements or address complex operating conditions [20].

Advanced control techniques, including Sliding Mode Control (SMC), Model Predictive Control (MPC), Adaptive control, and Deep Reinforcement Learning (DRL), have emerged as potential solutions to overcome the limitations of traditional control techniques and address the challenges associated with grid-connected inverters [12].

SMC is a robust control technique that ensures fast response and robustness against parameter variations and disturbances [21]. It uses a sliding surface to guide control actions for regulating active and reactive currents in the converter, enabling desired power sharing with the grid [22], [23], [24]. While SMC exhibits resilience to uncertainties and parameter variations, it can introduce chattering, high-frequency voltage or current oscillations, and control signal saturation, which may impact the achievable control performance [25], [26], [27], [28], [29].

MPC, as an advanced control technique, employs mathematical models to predict the future behavior of the converter and optimizes control signals to achieve desired performance [30], [31], [32], [33]. In the context of GCIs, MPC facilitates control of both active and reactive power sharing while considering constraints, such as grid code requirements, voltage and current limits [34], [35], [36], [37]. However, the real-time solution of complex optimization problems in MPC imposes computational demands and requires accurate modeling, real-time optimization algorithms, and expertise in control engineering [38], [39], [40], [41].

Adaptive control techniques adjust control parameters based on real-time estimation of system parameters or disturbances, ensuring optimal power sharing and stability under varying operating conditions [42], [43], [44], [45]. However, adaptive control relies on accurate parameter estimation, sophisticated algorithms, and additional computational resources, with a trade-off between adaptation and stability [46], [47].

DRL has gained attention in power electronics converter control, offering the ability to learn optimal control policies directly from raw data. DRL techniques find utility in power converter control, including voltage regulation in DC microgrids feeding constant power loads [48], [49], [50], [51], MPPT algorithms [52], [53], smart hybrid power systems coordinated load frequency control [54] and motor drive control [55], [56], [57], [58], [59]. DRL combines deep learning and reinforcement learning to enable autonomous learning and decision-making, accommodating changes in renewable energy generation, load demand, or grid conditions. DRL excels in handling complex and nonlinear power sharing scenarios, optimizing performance based on rewards and penalties obtained from the environment [60].

Within the domain of DRL, a variety of algorithms exists, including Deep Q-Learning (DQN), Deep Deterministic Policy Gradient (DDPG), and the TD3 method. Notably, DQN is aptly designed for discrete action spaces, whereas both DDPG and TD3 are well-suited for continuous action spaces. In the context of our implementation, the TD3 method was selected due to its recognized efficacy in applications involving continuous action spaces. This strategic choice aligns with the specific requirements of power electronic converter controllers, where continuous action spaces are often imperative for achieving optimal performance [61].

This article makes the following contributions to the literature:

- It introduces a DRL-based control architecture for direct current control of a grid-connected three-phase three-level NPC inverter. Conduct a comprehensive assessment of grid current stability under varying conditions, encompassing training, testing, and fluctuating reference scenarios.
- The DRL-based controller has been implemented in real-time using OPAL-RT and a DSP processor in Hardware-in-the-Loop (HIL) mode.

- The robustness analysis is performed on introduced method under parameter variations, grid dynamic conditions, showcasing enhanced performance compared to conventional and state-of-the-art control techniques.
- The DRL-based GCI current controller has been successfully validated in the context of solar PV grid integration applications.

The organization of the article is as follows. Section I discusses the role of GCIs in various applications, highlights the importance of NPC converters, explores LCL filter configurations, addresses the necessity of current control techniques, and reviews existing control techniques in the literature. Section II establishes the mathematical approach to GCI, presents the PI controller, and outlines the MPC control structure. Section III explains the DRL structure, its constituent elements, and the training methodology employed. Section IV delves into the performance evaluation of DRL under both training and testing conditions, along with a comparative performance analysis. Section V focuses on the implementation of a DRL-based grid-connected inverter in a solar PV grid interfacing application, taking into account varying irradiance conditions. Section VI concludes by summarizing the key findings in grid-connected current control methods using DRL and explores potential future research directions in this field.

II. MATHEMATICAL REPRESENTATION OF THE CONTROL STRUCTURE FOR THE GRID CONNECTED INVERTER WITH LCL-FILTER

The schematic representation of the GCI with LCL filter topology can be observed in Fig. 2, which consists of four switches and two clamped diodes per leg. In this configuration, various components are involved. V_{dc} corresponds to the voltage of the direct current (dc) link, while i_{dc} represents the supply current generated by V_{dc}. The dc-link support capacitors are denoted as C_1 and C_2, and they fulfill the role of supporting the dc-link. A fundamental assumption made within this context is the consistent nature of the dc-link voltage. This assumption finds support in the dc-link voltage regulator, which maintains the voltage stability even in scenarios of variations. This stability is achieved through the implementation of mechanisms such as a dc-link chopper in case of renewable solar energy sources, or under voltage protection. As such, the premise of a constant dc-link voltage is deemed realistic and justified. This pivotal assumption consequently allows for the independent operation of the grid- and solar-side components, decoupled from one another. Additionally, the inverter-side filter inductor is represented by L_1, while C_f denotes the filter capacitor. The grid-side filter inductor is referred to as L_2 and R_1, R_2 are parasitic resistances.

To further describe the system, specific currents are defined. The inverter-side three-phase currents are represented as i_{1a}, i_{1b}, and i_{1c}. On the other hand, the grid three-phase currents are denoted as i_{2a}, i_{2b}, and i_{2c}. These
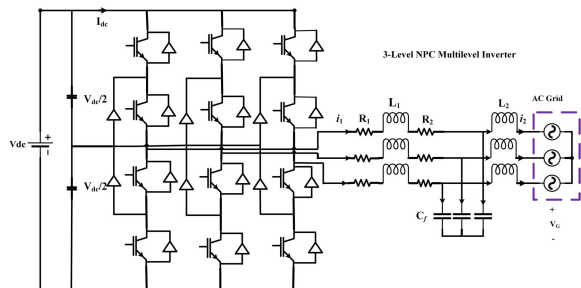
FIGURE 2. Grid connected three phase three level NPC VSC with LCL filter.

currents play crucial roles within the GCI with LCL filters topology. The characterization of grid current and voltage vectors is a fundamental aspect of this study are given by

$$i = \frac{2}{3}(i_a + a \cdot i_b + a^2 \cdot i_c)$$

$$v_{npc} = \frac{2}{3}(v_{a0} + a \cdot v_{b0} + a^2 \cdot v_{c0})$$

$$v_c = \frac{2}{3}(v_{ca} + a \cdot v_{cb} + a^2 \cdot v_{cc})$$

$$v_g = \frac{2}{3}(v_{ga} + a \cdot v_{gb} + a^2 \cdot v_{gc}) \tag{1}$$

where a assumes the value $e^{j2\pi/3}$, and the subscripts a, b, and c correspond respectively to grid phase currents, v_{npc} inverter-switching voltages, v_c voltage across capacitor and v_g grid line voltages.

With a balanced state of capacitor voltages and a consistent dc-link voltage, the voltages generated by the NPC at inverter terminals can be determined by

$$v_x = \begin{cases} \dfrac{s_x \cdot V_{dc}}{2} & \text{if } s_x \in \{-1, 1\} \\ v_n & \text{if } s_x = 0 \end{cases} \tag{2}$$

Here, x assumes values from the set $\{a, b, c\}$, and S_x represents the switching function that depends on the switching signals of the individual leg switches, v_n capacitors difference voltage.

In operation, the three-level GCI can produce three distinct switching states across the *abc* phases, labeled as "N", "O", and "P". These states are represented by the switching vector S, defined as follows:

$$S = \{s_a, s_b, s_c\}$$

where s_a, s_b, and s_c are each within the range $(-1, 0, 1)$, corresponding to the switching states "N", "O", and "P", respectively. Consequently, the three-level GCI encompasses a total of 27 unique combinations of switching states, equivalent to 27 fundamental voltage vectors, each with its specific magnitude distribution as explained in [62].

The state space representation of the above system is given by

$$\begin{bmatrix} \frac{di_1}{dt} \\ \frac{di_2}{dt} \\ \frac{dv_c}{dt} \end{bmatrix} = \begin{bmatrix} -\frac{R_1}{L_1} & 0 & -\frac{1}{L_1} \\ 0 & -\frac{R_2}{L_2} & \frac{1}{L_2} \\ \frac{1}{C} & -\frac{1}{C} & 0 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ v_c \end{bmatrix} + \begin{bmatrix} \frac{1}{L_1} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v_x \\ 0 \\ 0 \end{bmatrix} \tag{3}$$

The primary focus of this analysis is directed towards the NPC grid-side converter and its associated controller. The d-q frame, also known as the synchronous reference frame, is a mathematical coordinate system commonly used in the analysis and control of three-phase power systems. By transforming the three-phase current signals into the d-q frame, the control of the converter becomes more manageable as it allows independent control of the active and reactive power components.

In this configuration, the closed-loop current control aims to regulate the currents flowing through the converter's power switches. The main objective is to achieve accurate and precise control of the output currents while maintaining synchronization with the grid voltage. This control is crucial for various reasons, including maintaining power quality, ensuring efficient energy transfer, and providing robust and stable operation of the converter. The integration of a low-bandwidth Synchronous Reference Frame Phase-Locked Loop (SRF-PLL) plays a crucial role in synchronizing the converter with the Point of Common Coupling (PCC) voltage. The PI controller can be easily implemented in dq-frame about a linear operating region. The reference voltage control vectors from PI controller implementation are given by

$$E_d = V_d + (\omega \cdot L \cdot i_q) + (K_{P_1} \cdot e_1(t) + K_{I_1} \cdot \int e_1(t))$$

$$E_q = V_q - (\omega \cdot L \cdot i_d) + (K_{P_2} \cdot e_2(t) + K_{I_2} \cdot \int e_2(t)) \tag{4}$$

where, V_d, V_q are grid voltage in dq-frame K_{P_1}, K_{P_2} are proportional gains, K_{I_1}, K_{I_2} Integral gains, $e_1(t)$ d-axis current error, $e_2(t)$ q-axis current error. The PI controller operates only after the error signal occurs and also the tuning of the control loops are difficult. This can be overcomed by using the predictive control implementation using mpc technique.

FCS-MPC is crucial in applications where control inputs are constrained, such as power electronics converters control, as it optimizes a finite set of possible inputs, ensuring adherence to physical limitations. The FCS-MPC control method is an advanced control strategy used to optimize the performance of dynamic systems. The first crucial step in designing an MPC system is to develop a mathematical model that accurately captures the dynamics of the controlled process. This model describes how the system responds to changes in inputs over time and serves as the foundation for predicting future behavior. The continuous time mathematical model is represented in (3).

Once the continuous-time model is established, the next step involves discretizing it to align with the control intervals. Discretization allows the formulation of a discrete-time model, which is essential for the subsequent optimization process. The choice of an appropriate sampling time is critical in this phase to strike a balance between accuracy and computational efficiency, the sampling time of one micro second is considered here. The discretization process is

explained using (5)-(7).

$$L\frac{di(t)}{dt} = (v_x(t) - v_c(t)) - Ri(t) \quad (5)$$

$$\frac{di(t)}{dt} = \frac{v_x(t) - v_c(t)}{L} - R \cdot i(t) \quad (6)$$

$$\frac{i(k+1) - i(k)}{\tau_s} = \frac{v_x(k) - v_c(k)}{L} - \frac{R \cdot i}{L} \quad (7)$$

$$i(k+1) = \left(1 - \frac{\tau_s \cdot R}{L}\right) \cdot i(k) + \frac{\tau_s}{L} \cdot (v_x(k) - v_c(k)) \quad (8)$$

The heart of MPC lies in the formulation of a cost function that quantifies the control objectives and constraints. The cost function typically comprises terms related to control effort, deviation from desired setpoints, and any other specific performance goals relevant to the application. The cost function is represented using the (9) for addressing the tracking problem which deals with deviation from desired setpoints. The prediction horizon, representing the duration over which future system behavior is anticipated, and the control horizon, specifying the length of time over which control inputs are optimized, are essential parameters in this formulation. For this power electronic control problem statement, the prediction horizon and control horizon both are set to one for reducing the computational complexity.

$$\text{cost} = (i_{dref} - i_{dactual})^2 + (i_{qref} - i_{qactual})^2 \quad (9)$$

In FCS-MPC, the optimization problem is structured to operate on a finite set of possible control inputs. This is a distinctive feature of this approach, especially beneficial in systems with physical constraints on the control input space, such as power electronics. The Finite Control Set has 19 switching states out of 27, achieved by reducing redundancy for the three-phase three-level inverter. The optimization problem is solved online at each control interval, ensuring real-time adaptability to the dynamic nature of the system

The final step is the iterative online implementation of the MPC algorithm. In flowchart Fig 3, the system follows these steps: it measures the current state, predicts future values through computation, solves the optimization problem, and applies the first optimal control input at each time step. This iterative process continues, allowing MPC to continuously adjust and optimize control inputs based on real-time measurements, making it a powerful and adaptive control strategy in various applications.

The performance of the model predictive control depends on the accurate mathematical modelling, cost function and prediction horizon. But the grid interfaced inverters demands model free operation due to inaccurate prediction with various factors like programmed filter parameters differs with actual values during operation which leads to poor static and dynamic performance. The DRL based controller overcomes these challenges by offering a model-free approach.
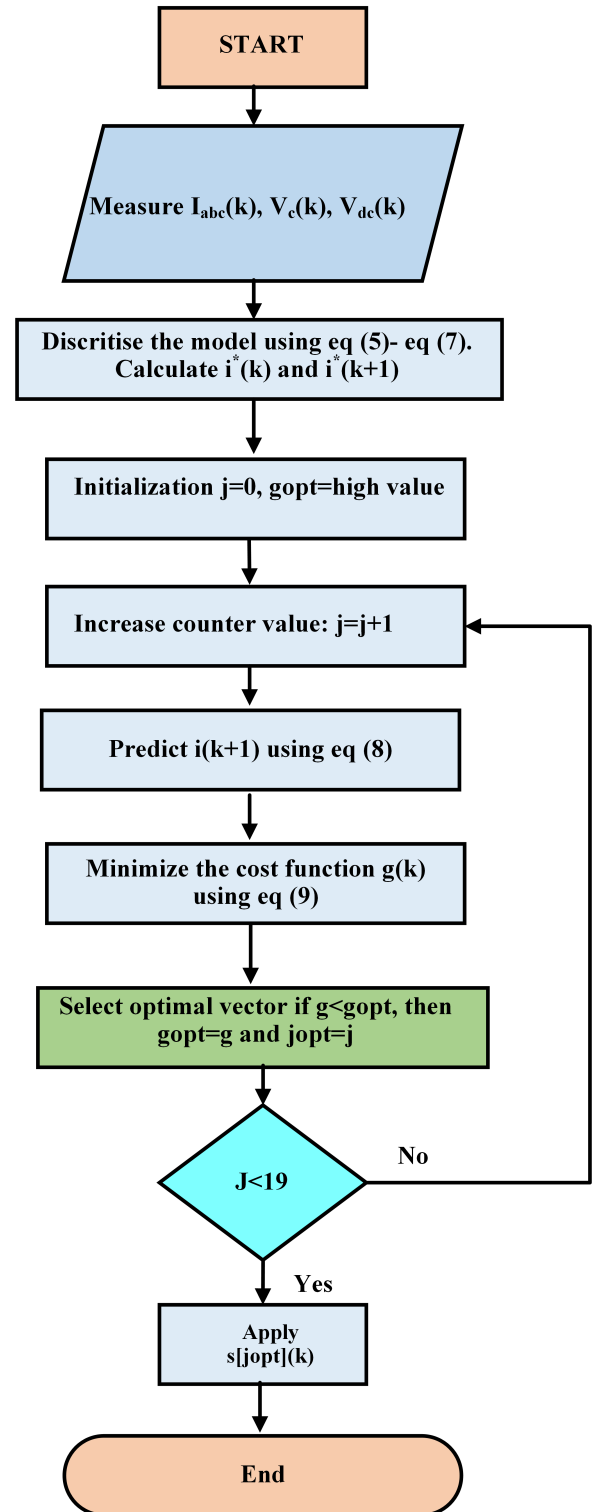

FIGURE 3. Flow chart of model predictive control showing prediction equation, cost function and online optimization.

III. METHODOLOGY

The DRL framework is a sophisticated approach that combines deep learning and reinforcement learning techniques to develop intelligent control policies for complex systems.
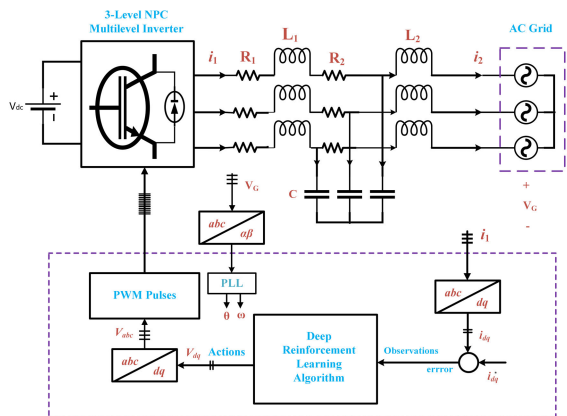
FIGURE 4. Deep reinforcement frame work illustrating agent-environment interaction and the control structure representation.

This framework involves several key components and steps as shown in Fig. 4:

A. FRAME WORK

1) ENVIRONMENT

The environment represents the system or process that the DRL agent interacts with. It provides the agent with observations or states and receives actions in return. In the context of grid connected applications, the environment could be the grid connected three phase NPC inverter with LCL filter, the load it operates on.

2) AGENT

The agent makes the decision by taking the observation as input and gives actions as output using a deep neural network. A deep neural network is used as the function approximator within the DRL framework. It takes the current state as input and outputs the action to be taken by the agent. The deep neural network consists of multiple layers of interconnected nodes that learn to approximate the optimal control policy through training. The neural network takes actor-critic structure from algorithm and it's configuration is mentioned in Table 1.

3) STATE SPACE

The state represents set of observations made in that the current situation or condition of the system. It includes relevant information such as error signal of d-qxis ($e_d(t)$), q-axis ($e_q(t)$), their integral values ($\int e_d(t), \int e_q(t)$), the grid voltages ($V_{dg}(t), V_{qg}(t)$) and the output current of the inverter($I_d(t), I_q(t)$) in d-q frame. The observations set is typically represented using numerical values or sensor measurements and is given by

$$S = \left[e_d(t), e_q(t), \int e_d(t), \int e_q(t), I_d(t), I_q(t), V_{dg}(t), V_{qg}(t) \right]$$

(10)

where, $e_d(t) = I_d^*(t) - I_d(t)$, $e_q(t) = I_q^*(t) - I_q(t)$, $I_d^*(t) \& I_d(t)$ is the reference and actual current of the inverter in along
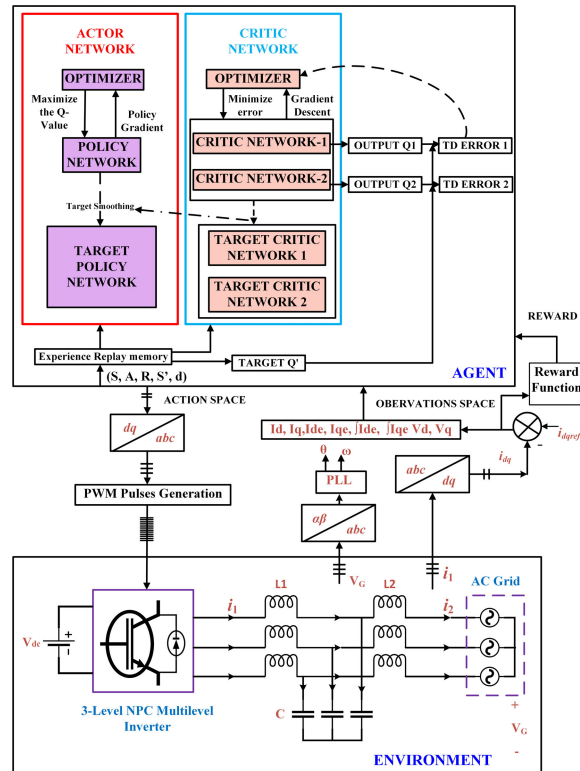

FIGURE 5. DRL frame work illustrating agent-environment interactions, the key components and workflow involved in TD3 algorithm.

d-axis, $I_q^*(t) \& I_q(t)$ is the reference and actual current of the inverter in along q-axis, $V_{dg}(t) \& V_{qg}(t)$ are the grid voltages in dq-frame.

4) ACTION SPACE

The action space defines the set of actions that the DRL agent can choose from. The actions are of two types i.e discrete actions, continuous actions. In inverter current control scenario, the actions can include estimation of continuous time d-q axis voltage signals (E_d, E_q) which will be converted as reference signals in abc frame for PWM pulses generation as shown if Fig. 5. The action space is designed to cover a range of possible control actions that the agent can take to influence the system's behaviour. The limited action space can help the DRL converges very fast and the knowledge on action space limitation can be obtained from dynamics of the environment or conventional control techniques. Here, The action space could be reduced by taking the observation values in per unit basis.

5) REWARD FUNCTION

The reward function defines the quantitative measure of success or performance for the DRL agent. It provides feedback to the agent based on its actions and the resulting system behaviour. The reward function can be designed to encourage desired system behaviours, such as effective tracking performance while penalizing undesired behaviours or control actions that deviate from optimal performance. The

design of the reward function is very crucial step in DRL and is reffered from the paper [63]. The reward function r(t) considered for training is given by

$$r(t) = \begin{cases} 1 - |e(t)|^2, & \text{if } |e(t)| \le 0.001 \\ 0.01 - |e(t)|^2, & \text{if } 0.001 < |e(t)| \le 0.01 \\ 0.001 - |e(t)|^2, & \text{if } 0.01 < |e(t)| \le 0.1 \\ -5|e(t)|^2, & \text{if } 0.1 < |e(t)| \end{cases}$$

(11)

Here, e(t) is the error signal defined as the difference between the desired current and the actual current of the dq-axis. When the reward is calculated along the dq-axis, the error is assigned as $e(t) \leftarrow e_d(t)$ for the d-axis and $e(t) \leftarrow e_q(t)$ for the q-axis.

B. TRAINING PROCESS

The training process involves iteratively updating the deep neural network to improve its performance. The DRL agent interacts with the environment, observes the current state, selects actions based on the current policy, and receives rewards. These experiences are used to update the neural network parameters through back-propagation and optimization algorithms such as stochastic gradient descent. The training process continues until the agent's control policy converges to an optimal or near-optimal solution. By utilizing this DRL framework, the agent gets ability to learn optimal control policies directly from data and interactions with the environment.

C. ALGORITHM: TD3

The TD3 algorithm is an advanced DRL technique used to train agents in complex decision-making tasks. TD3 builds upon the DDPG algorithm and introduces several enhancements to improve stability and sample efficiency. The TD3 algorithm is given in algorithm 1 and is illustrated in Fig. 5. The algorithm consists of the following key components and involves several steps to train an agent in a reinforcement learning task:

1) KEY COMPONENTS

- Actor-Critic Architecture: TD3 utilizes an actor-critic architecture, where the actor network learns the policy (the action selection strategy), and the critic network evaluates the value of the chosen actions. Both the actor and critic networks are deep neural networks that learn to approximate complex functions.
- Delayed Actor Updates: TD3 introduces delayed actor updates to address overestimation issues in DDPG. Instead of updating the actor network at every time step, TD3 delays the update by a fixed number of steps. This helps to decorrelate the actor's updates from the critic's evaluations and stabilizes the learning process.
- Twin Critics: TD3 employs two critic networks instead of a single one used in DDPG. This helps to mitigate the overestimation bias and leads to more accurate

---

Algorithm 1 Twin Delayed Deep Deterministic Policy Gradient (TD3)

1: INPUT: initial policy parameters θ, Q-function parameters ϕ_1, ϕ_2, empty replay buffer $\mathcal{D}$
2: Set target parameters equal to main parameters: $\theta_{\text{target}} \leftarrow \theta, \phi_{\text{target1}} \leftarrow \phi_1, \phi_{\text{target2}} \leftarrow \phi_2$
3: repeat
4: Observe state and select action $a = \text{clip}(\mu_\theta(s) + \epsilon, a_{\text{Low}}, a_{\text{High}})$, where $\epsilon \sim \mathcal{N}$
5: Execute a in the environment
6: Observe next state s', reward r, and done signal d to indicate whether s' is terminal
7: Store (s, a, r, s', d) in replay buffer $\mathcal{D}$
8: **If** s' is terminal, reset environment state
9: **If** it's time to update **then**
10:      **for** j in range(however many updates)
11:      Randomly sample a batch of transitions, $B = \{(s, a, r, s', d)\}$ from $\mathcal{D}$
12:      Compute target actions $a'(s') = \text{clip}\left(\mu_{\theta_{\text{target}}}(s') + \text{clip}(\epsilon, -c, c), a_{\text{Low}}, a_{\text{High}}\right), \epsilon \sim \mathcal{N}(0, \sigma)$
13:      Compute targets $y(r, s', d) = r + \gamma(1 - d) \sum_{i=1}^{2} \min Q_{\phi_{\text{target},i}}(s', a'(s'))$
14:      Update Q-functions by one step of gradient descent using: in 1,2 $\nabla_{\phi_i} \frac{1}{|B|} \sum_{(s,a,r,s',d) \in B} (Q_{\phi_i}(s, a) - y(r, s', d))^2$ for i
15:      If j mod policy delay $= 0$
16:      Update policy by one step of gradient ascent using: $\nabla_\theta \frac{1}{|B|} \sum_{s \in B} ((s, a) - y(r, s', d))^2 Q_{\phi_i}(s, \mu_\theta(s))$
17:      Update target networks with $\phi_{\text{target},i} \leftarrow \rho\phi_{\text{target},i} + (1 - \rho)\phi_i$ $\theta_{\text{target}} \leftarrow \rho\theta_{\text{target}} + (1 - \rho)\theta$
18:      **end if**
19:      **end for**
20: **end if**
21: **until** convergence

---

value estimations. The two critic networks have separate parameters and are trained independently.
- Target Networks and Polyak Averaging: TD3 utilizes target networks to provide stable value estimations. These networks are copies of the actor and critic networks and are updated slowly by polyak averaging. The target networks are used to calculate target values, which are then used for training the actor and critic networks.
- Exploration and Noise: To encourage exploration during training, TD3 adds noise to the selected actions. This noise is typically drawn from a probability distribution, such as Gaussian noise, and helps to discover new actions and avoid getting stuck in local optima.
- Replay Buffer: TD3 employs a replay buffer, which is a memory structure that stores past experiences (state, action, reward, next state). The replay buffer is used to randomly sample and provide training data to the actor and critic networks. By utilizing past experiences, TD3 improves sample efficiency and reduces the correlation between consecutive samples.

TABLE 1. TD3 neural network configuration parameters.

Category	Parameter	Value
Actor	Layer wise neurons	$L_1 = 64, L_2 = 32, L_3 = 1$
	Activation function	ReLU
	Learning rate	0 0.001 s
Critic	Layer wise neurons	$L_1 = 64, L_2 = 32, L_3 = 1$
	Activation function	Tanh
	Learning rate	0.0001 s
General	Discount factor	0.95
	Mini batch size	512
	Experience buffer length	$2 X 10^6$
	Action space	Continuous
	Sampling time	1e-4 sec

- Loss Functions and Optimization: TD3 uses appropriate loss functions and optimization algorithms to update the actor and critic networks. The actor network is trained to maximize the expected cumulative reward, while the critic networks are trained to minimize the temporal difference error between the estimated value and the target value.

2) STEPS INVOLVED

- Initialization: The training process begins by initializing the actor and critic networks with random weights. Additionally, target networks, which are copies of the actor and critic networks, are created and initialized with the same weights as shown in step 1 & 2.
- Interaction with the Environment: The agent interacts with the environment by selecting actions based on the current state and the learned policy. These actions are sent to the environment, which provides the agent with the next state and a reward signal as shown in as shown in step 3-5.
- Replay Buffer Update: The experiences of the agent, consisting of the state, action, reward, and next state, are stored in a replay buffer as shown in step 6-8. This replay buffer serves as a memory that holds past experiences, which will be used for training.
- Training Batch Sampling: From the replay buffer, a batch of experiences is randomly sampled for training as shown in step 9-11. These experiences are used to update the actor and critic networks. The batch size is typically a fixed number of samples.
- Actor Network Update: The actor network is updated by maximizing the expected cumulative reward as shown in step 13. This is done by computing the gradients of the expected cumulative reward with respect to the actor network's parameters as shown in step-14. These gradients are then used to update the actor network's weights through an optimization algorithm, such as stochastic gradient descent.

- Delayed Actor Update: Instead of updating the actor network at every time step, a delay is introduced in the actor update. This delay allows for decorrelation between the actor updates and the critic evaluations. The delay can be a fixed number of steps, ensuring more stable learning.
- Twin Critic Update: The twin critic networks are updated by minimizing the temporal difference error between the estimated value and the target value. The target value is computed using the target networks and is based on the discounted sum of future rewards. The gradients of the temporal difference error are then used to update the critic networks' weights as shown in step-16.
- Target Network Update: The target networks, which provide stable value estimations, are updated slowly using a technique called polyak averaging. This involves updating the target network's weights as a fraction of the current network's weights as shown in step-17. The slow update helps in achieving a smoother learning process and avoids over-fitting.
- Exploration and Noise: To encourage exploration during training, noise is added to the selected actions. This noise can be sampled from a probability distribution, such as Gaussian noise. The noise injection promotes exploration and helps the agent discover new actions and policies beyond local optima.
- Repeat: Steps 2 to 9 are repeated iteratively until the agent's performance reaches a satisfactory level or a specified number of iterations is completed.

Through this process, the agent learns an optimal policy that maximizes cumulative rewards and improves its performance in the given reinforcement learning task.

IV. RESULTS AND DISCUSSION

This section focuses on presenting the outcomes of direct current control of a three-phase NPC inverter connected to the grid. The control methods employed are based on DRL utilizing the TD3 algorithm and conventional PI control and mpc techniques. The PI controller gains are tuned by using Ziegler Nicholas method and are given by

$$K_{P_1} = 5, \quad K_{I_1} = 5000, \quad K_{P_2} = 5, \quad K_{I_2} = 5000 \quad (12)$$

In mpc techniques, the finite control set technique has been implemented. For DRL, the agent training simulations are conducted in the MATLAB 2021a environment using a Dell Precision 5820 workstation equipped with 32GB RAM and a 16GB NVIDIA RTX A4000 GPU. The Real-Time validation process is carried out using the OPAL-RT OP4512 target simulator. A comprehensive hardware setup, as depicted in Fig. 6, was utilized for the complete implementation of Real-Time operations.

In the experimental verification, a precise validation of proposed control system is done utilizing a suite of hardware components, including the OPAL-RT OP4512 Target simulator, TMS320F28379D Launchpad, and a high-performance
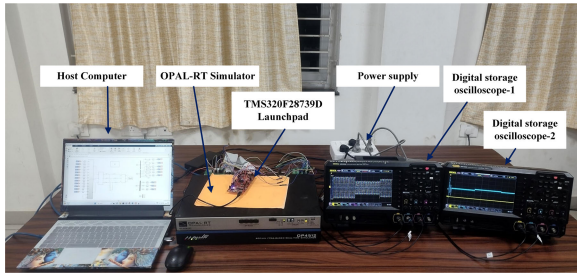
FIGURE 6. Real time hardware implementation setup using OPAL-RT, Texas Instruments C2000 LaunchXL-F28379D.

Host computer as shown in Fig. 8. The OPAL-RT simulator, equipped with an Intel Xeon E3 processor featuring four cores at a clock speed of 3.7GHz, 16GB RAM, and a 256GB SSD memory, serves as the basis for executing real-time simulations. Its extensive set of 2×16 Analog In, 2 x Analog Out, 2×16 Digital In, and 2×16 Digital Out channels provides a versatile interface for seamless data exchange.

The Real-Time simulation verification is coordinated using the RT-LAB 2023.1 software platform, ensuring seamless coordination and synchronization across the OPAL-RT simulator, TMS320F28379D Launchpad, and the Host computer–a Dell Precision 5820 workstation provided with 32GB RAM and a 16GB NVIDIA RTX A4000 GPU.

The TMS320F28379D Launchpad, featuring a 200MHz processor speed, dual C28xCPUs, dual CLAs, 1MB Flash, and advanced functionalities such as 16-bit or 12-bit ADCs, comparators, 12-bit DACs, and PWMs, plays a pivotal role in executing the proposed controller. The Real-Time experimentation unfolds through a structured series of steps.

Initiating with the offline training of the DRL-based TD3 agent in MATLAB, the process yields an optimal agent through the MKL-DNN library. During the training process, a sampling period of one micro second for environment and a sampling time of 0.1 msec for agent is considered. Transitioning to Real-Time execution, the designed environment–a 3-phase 3-level grid-connected inverter with an LCL filter–is executed on the OPAL-RT simulator. Concurrently, the DRL controller is executed on the TMS320F28379D Launchpad.

As shown in the Fig. 7, the organization of hardware connections facilitates a seamless exchange of information, with the OPAL-RT transmitting three inverter currents and three grid voltages via analog output channels. The Launchpad, in turn, receives and processes these signals through ADC channels, generating 12 PWM pulses conveyed through Digital output channels. The OPAL-RT, functioning as the central coordinator, receives these PWM pulses via Digital input channels. The Real-Time execution is explained as shown in flowchart in 7. In the first step, open MATLAB simulated files in accordance with RT-LAB 2023.1 standards on the Host computer in edit mode. Subsequent to successful executions in edit mode then set the configuration for the processor as OPAL-RT Linux (x64-based). The model is then built onto the processor,

I/O configurations are established on the OPAL board, and the target is assigned as OPAL-Linux (x64-based) with rt-server, leveraging hardware-synchronized execution properties. At the same time, the DRL agent has to deploy onto the C2000 Launchpad through MATLAB. The final step involves loading the model onto the OPAL-RT target and initiating the execution of the entire system, thereby validating the effectiveness of the control strategy in a real-time environment.

The results obtained from both training and testing conditions using TD3 method are discussed in Case-I. In Case-II, the efficacy of the proposed DRL-based TD3 method is compared with other control methods and validated the results using the OPAL-RT experimental setup in real-time. In case-III, the proposed control method is tested for fluctuating reference tracking which reflects practical scenarios. In Case-IV, a comprehensive analysis is conducted to compare and assess the dynamic responses of all controllers to determine their respective levels of superiority under normal operating condition and parameter variations. Finally, in case-V, The tracking stability of the dq-currents is assessed under dynamic grid conditions. To facilitate reproducibility, the parameters of grid connected NPC inverter with LCL filter configuration are derived from established real-time research practices, as documented in [64] and [65] and are outlined in Table. 2.

A. CASE-I: ABILITY OF DRL BASED CONTROLLER TO TRACK REFERENCE CURRENT

The agent undergoes training over 890 episodes using the TD3 algorithm on a three-phase NPC inverter with an LCL filter, incorporating dq-reference current variations as defined by (13) - (14).

$$i_d(t) = \begin{cases} -160\,A & 0 \leq t \leq 0.2 \\ -100\,A & 0.2 \leq t \leq 0.4 \\ -60\,A & 0.4 \leq t \leq 0.5 \end{cases} \tag{13}$$

$$i_q(t) = \begin{cases} 0\,A & 0 \leq t \leq 0.5 \end{cases} \tag{14}$$

The training statistics, depicted in Figure 8, illustrate that the training has reached the optimal policy, as evidenced by the constancy of the average reward over more than 50 episodes. Following successful training, the agent is subjected to testing using a different set of reference current variations (deviation from the training values) defined by (15) - (16).

$$i_d(t) = \begin{cases} -150\,A & 0 \leq t \leq 0.2 \\ -70\,A & 0.2 \leq t \leq 0.4 \\ -100\,A & 0.4 \leq t \leq 5 \end{cases} \tag{15}$$

$$i_q(t) = \begin{cases} 0\,A & 0 \leq t \leq 0.5 \end{cases} \tag{16}$$

The waveforms illustrating Grid currents, dq axis current variations, and the total harmonic distortion (THD) of the three phase NPC inverter with LCL configuration are
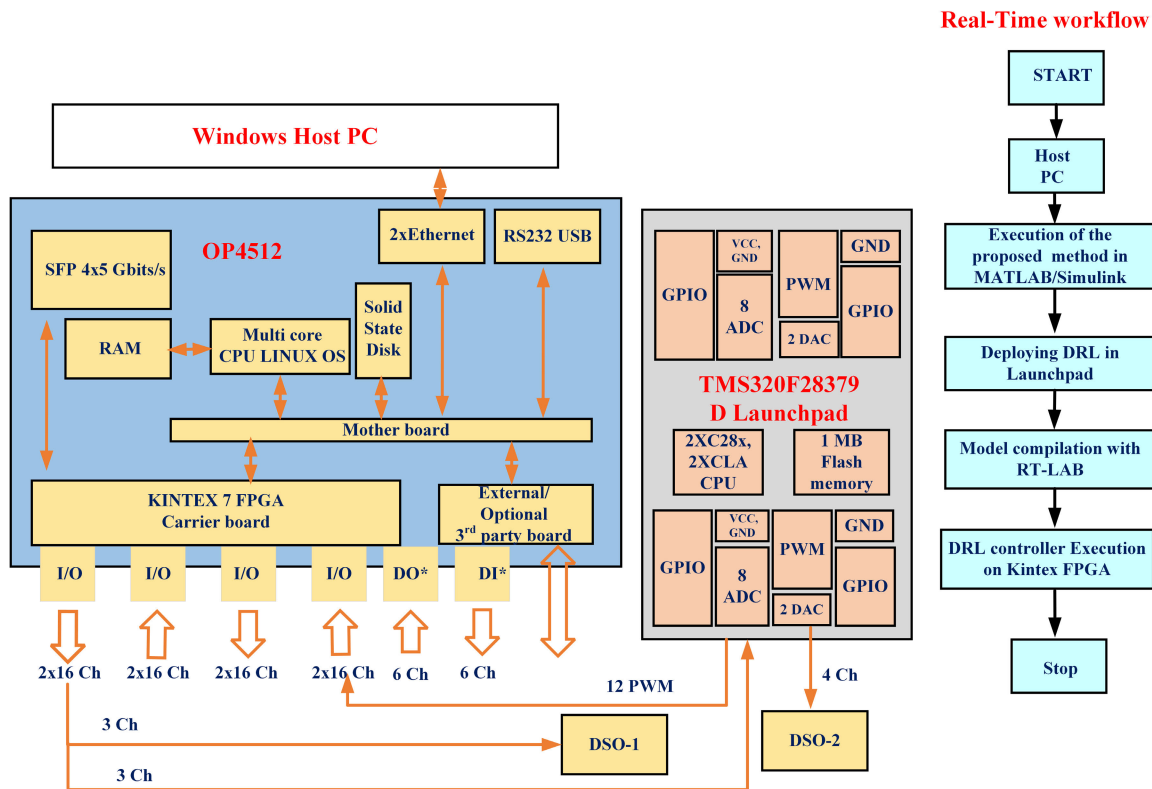
FIGURE 7. Online experimental verification method of the proposed algorithm.

TABLE 2. Three phase NPC inverter with LCL filter configuration parameters.

S. No	Parameter	Value
1	Inverter power rating	100 KVA
2	DC Link voltage	800 volts
3	Output voltage	415 V_{L-L}
4	Frequency	50 Hz
5	Switching frequency	10 KHz
6	Resonating frequency	1 KHz
7	Inductance	1.5 mH
8	Capacitance	100 micro Farad
9	Sampling Time	1 μ sec

FIGURE 8. The training statistics illustrate that the training has reached the optimal policy.

B. CASE-II: REAL-TIME HIL VALIDATION OF DRL BASED CONTROLLER, PI AND MPC CONTROL TECHNIQUES

In order to validate the DRL algorithm as a Real-Time controller, simulations are conducted on the OPAL-RT platform and Texas C2000 launchpad. The results pertaining to output grid current regulation and dq-axis current regulation using the TD3 method are presented in Fig. 15- Fig. 16. Additionally, the corresponding outcomes obtained with the PI method and MPC control technique are depicted in Fig. 17 - Fig. 20. The THD graphs of all three methods are shown in Fig. 21 - Fig. 23. All controllers exhibit satisfactory steady-state and dynamic performance under normal operating conditions, however, their robustness, static
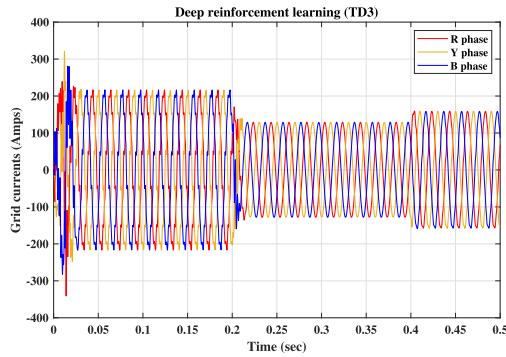
presented in Fig. 9-11 under the training condition, as well as in Fig. 12-14 under the testing condition. From these graphs, It is can be observed the grid currents could track the reference currents effectively and the THD of the grid currents is within IEEE 519 grid code condition. It is also evident that the agent has successfully acquired knowledge of the underlying patterns and relationships within the training data and can effectively generalize this knowledge to unseen data.

FIGURE 9. Three phase grid side currents using TD3 method with training case.

FIGURE 10. Tracking performance of dq axis current using TD3 method with training case.

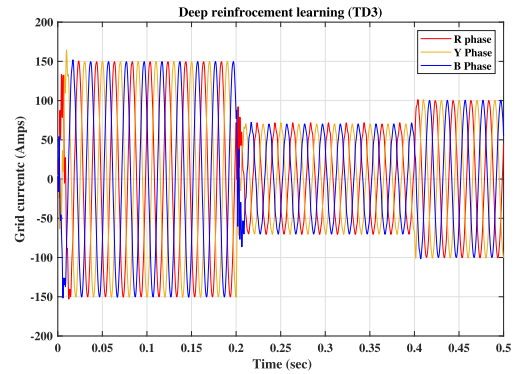**FIGURE 11.** The THD of Three phase grid currents using TD3 method with training case, 3.34%.

FIGURE 12. Three phase grid currents using TD3 method under testing set.
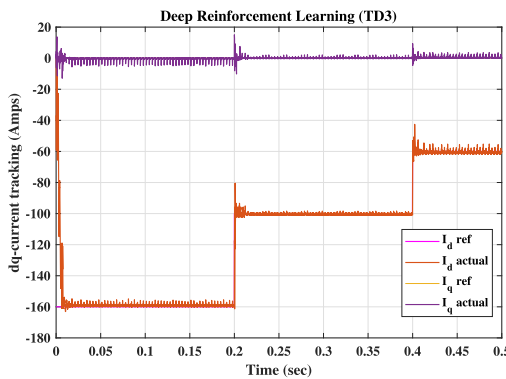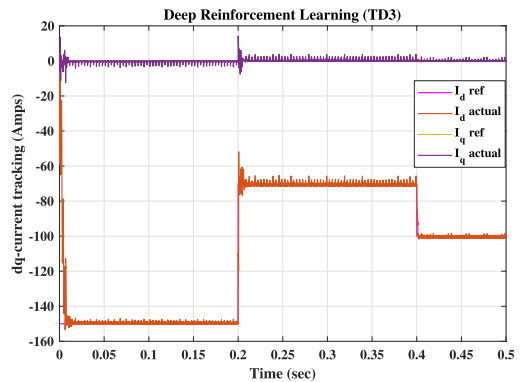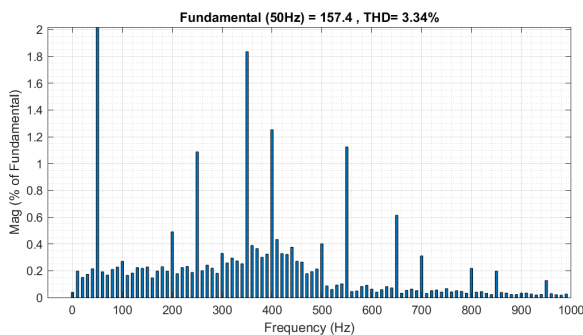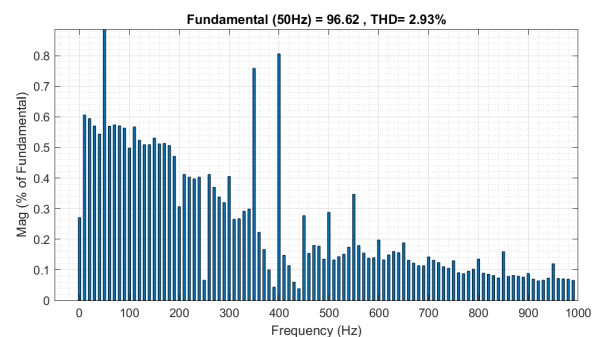

FIGURE 13. Tracking performance of dq axis current using TD3 method under testing set.

FIGURE 14. The THD of three phase grid currents using TD3 method under testing set, 2.93%.

and dynamic performance characteristics are compared in the subsequent case.

C. CASE-III: ABILITY TO TRACK FLUCTUATING REFERENCE CURRENT

GCI systems are commonly employed to establish connections between wind turbines and solar photovoltaic arrays with the electric power grid. The challenge arises from the fact that the power output from these sources fluctuates considerably due to varying weather conditions. Consequently, the reference current for the GCI system experiences constant variations over time. For instance, within periods shorter than an hour, wind speed can be approximated by considering a slowly varying mean speed, denoted as V_w, along with N

sinusoidal components featuring frequencies ω_i, amplitudes A_i, and random phases ϕ_i, as detailed in (17).

$$v_{dc}(t) = V_{dc} + \sum_{n=1}^{N} A_n \cos(\omega_n t + \phi_n) \quad (17)$$

In the context of this fluctuating and gusty wind scenario, the power from a wind turbine is generated with a dynamic negative d-axis current values. Simultaneously, the q-axis reference current is set to zero, indicating a lack of reactive power. This configuration aligns with the typical behavior of wind power production under such turbulent wind conditions.

Furthermore, the depicted Fig. 24 highlights the impressive performance of DRL based controller in effectively tracking
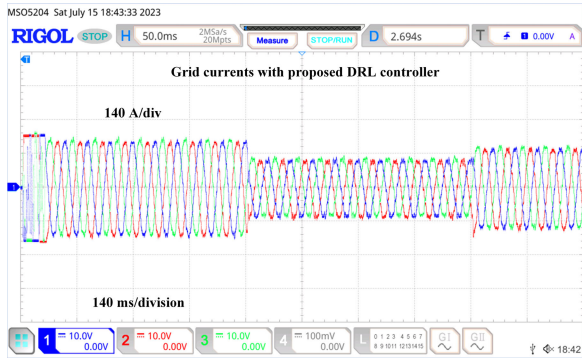
FIGURE 15. Three phase grid side currents waveforms using TD3 method under Real-Time validation.
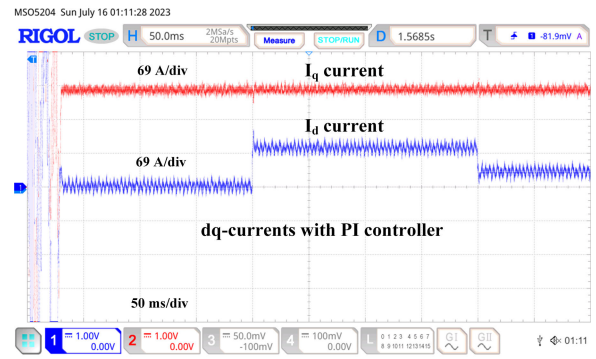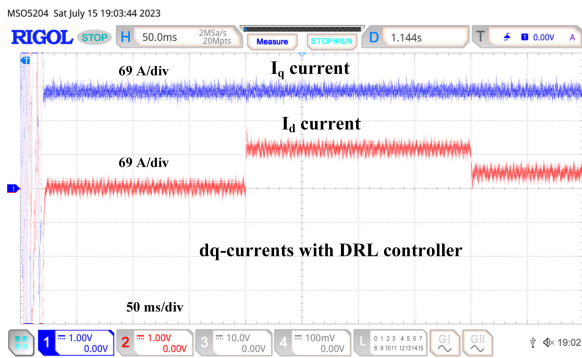

FIGURE 16. Tracking performance of dq axis currents using TD3 method under real time validation.

FIGURE 17. Three phase grid side currents waveforms using PI method under real-time validation.

FIGURE 18. Tracking performance of dq axis currents using PI method under real time validation.
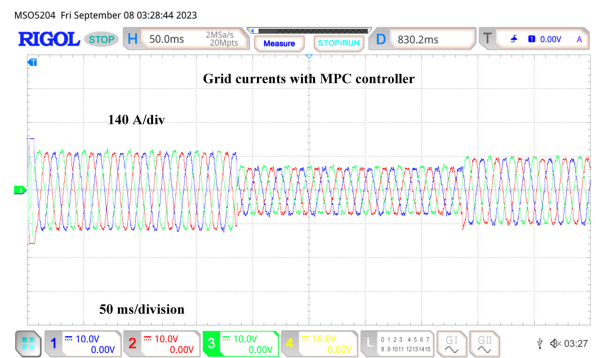

FIGURE 19. Three phase grid side currents waveforms using MPC method under real-time validation.
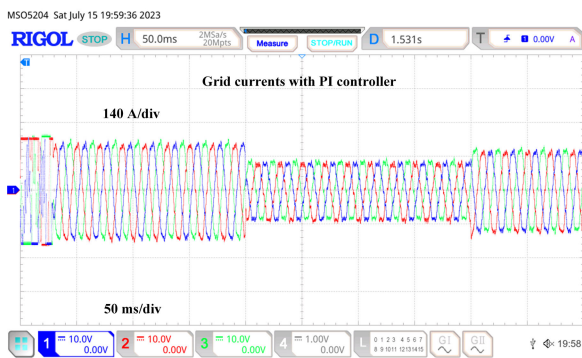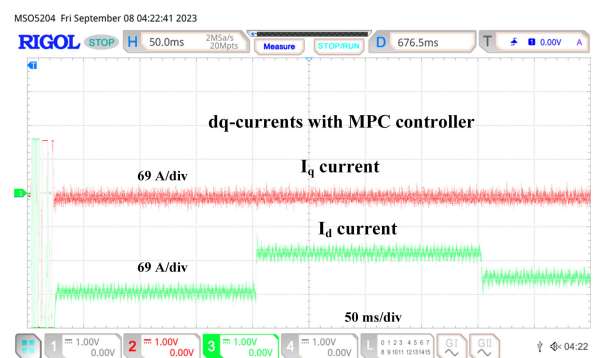

FIGURE 20. Tracking performance of dq axis currents using MPC method under real time validation.

the variable reference current within the power converter switching environment.

D. CASE-IV: ROBUST PERFORMANCE EVALUATION OF PROPOSED METHOD UNDER PARAMETER VARIATION

Stability analysis for tracking performance has long been a central concern within traditional control systems, particularly in the context of GCIs. Research in this domain typically delves into the performance of GCIs when subjected to variations in system parameters or when dealing with unbalanced or distorted AC system conditions. To illustrate, [66] conducts a sensitivity analysis employing a small-signal model to assess GCI behavior amidst changing system parameters. Similarly, [67] introduces a control strategy

aimed at enhancing GCI performance when confronted with varying system conditions.

In this case, a comprehensive comparison is conducted among the proposed TD3 method, the conventional PI control and MPC methods to check the performance facing dynamic changes in parameters. The study focused on two specific variations: changes in the resistance and inductance of the grid filter. The results showed a fine impact on the controller, with noticeable effects from shifts in grid-filter inductance rather than alterations in resistance. Notably, the controller remained robust in the face of changes in grid-filter resistance. An interesting observation emerged, indicating that the controller was less affected when the grid-filter inductance exceeded the nominal value. Conversely, adverse
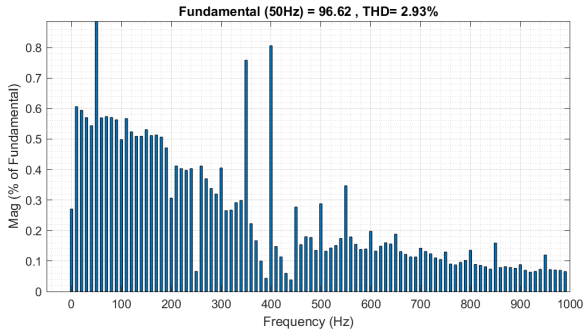
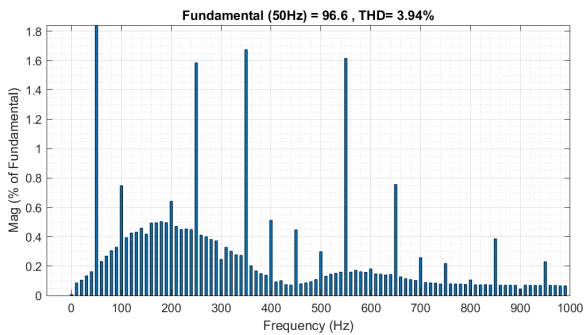**FIGURE 21.** The THD in grid side currents using TD3 method, 2.93%.

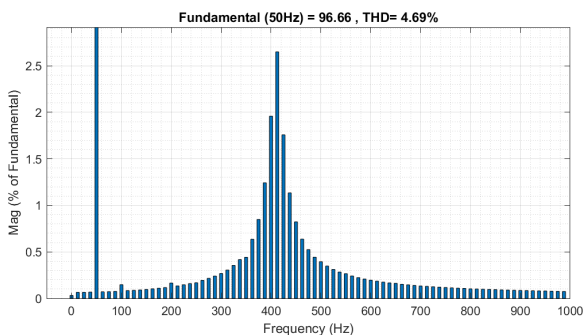**FIGURE 22.** The THD in grid side currents using PI method, 3.94%.

FIGURE 23. The THD in grid currents using MPC method, 4.69%.

FIGURE 24. Tracking performance of dq axis currents using TD3 method under fluctuating reference current.

FIGURE 25. Static and dynamic response of the dq- current tracking comparison for the PI, MPC, DRL control techniques a) under nominal parameter values, b) under parameter mismatch, 25% above nominal values, c) under parameter mismatch, 25% above nominal values.

effects occurred when the inductance fell below the nominal value. The evaluation focuses on aspects such as effectiveness in static and dynamic performance, as well as grid code conditions. Figure 25 a) illustrates the dq-axis current regulation achieved with the TD3 method, the PI controller, MPC controller and their comparison under nominal operating conditions. Figure 25 b) & c) shows the current tracking performance under filter inductance parameter below 50% to nominal value and above 25% to nominal value respectively. The steady state and the dynamic performance analysis is tabulated in Table. 3 for all the three cases,

The key findings from the Table. 3 indicate that when the actual inductance is reduced to 50% below the nominal value, the DRL-based controller meets the IEEE 519 Grid code conditions by maintaining Total Harmonic Distortion (THD) below 5% while the PI control and MPC techniques yields exceeding 5% under such conditions. However, when the actual inductance exceeds the nominal value by 25%,
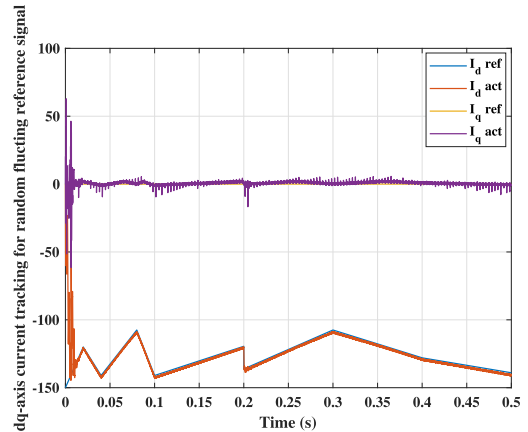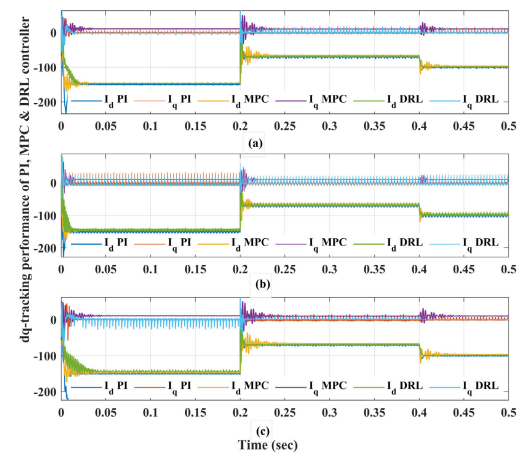
the THD of the DRL controller remains within grid code specifications whereas the other two exceeds the limits. Importantly, if the actual inductance falls more than 50% below the nominal value, significant adverse effects emerge, leading to pronounced distortion and imbalance in the grid current. The inferences from the Table. 3 are given as follows.

- In terms of Overshoot, the order of robustness from most to least robust is DRL > MPC > PI. The DRL controller demonstrates the lowest percentage overshoot, consistently at 0% during increased step change under all parameter variations. It is highly robust in terms of overshoot.
- For Settling Time, the order is DRL < MPC < PI, with the DRL controller having the fastest response. The settling time of the DRL controller is the shortest among the three controllers, indicating rapid response to set point changes.
- Regarding Steady State Error, the order is DRL < MPC < PI, with the DRL controller showing the best steady-state performance. The DRL controller maintains a

TABLE 3. Static and transient performance metrices for PI, MPC, DRL control techniques under various conditions.

Controller	Parameter variation	% Overshoot	Settling time (msec)	steady state error	THD
PI	50 % below	52.6	10	8.5%	6.04 %
	Nominal	56 %	15	4.2%	3.19 %
	25 % above	49.3 %	10	10%	15.25 %
MPC	50 % below	18 %	15	11.4%	8.49 %
	Nominal	15.3 %	25	7.14%	4.69 %
	25 % above	16 %	35	14.2%	11.01 %
DRL	50 % below	10.4 %	2.5	4.2%	4.29 %
	Nominal	9 %	5	1.4%	2.93%
	25 % above	10.1 %	2.5	5.7%	4.09%
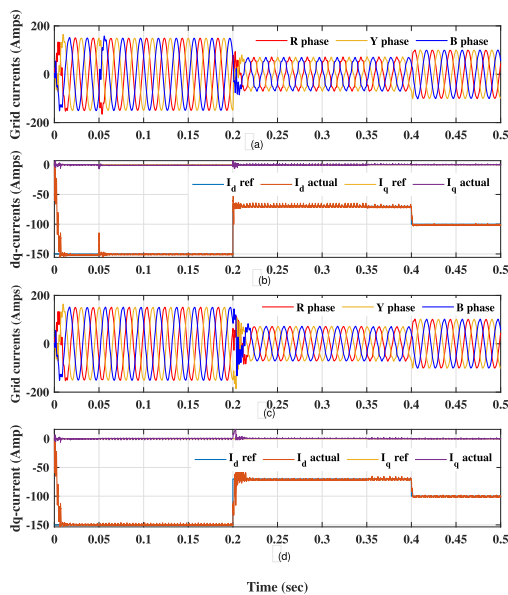
FIGURE 26. Stability of the dq-current demonstrated under dynamic grid voltage conditions.

FIGURE 27. THD of grid currents under dynamic voltage rise condition.

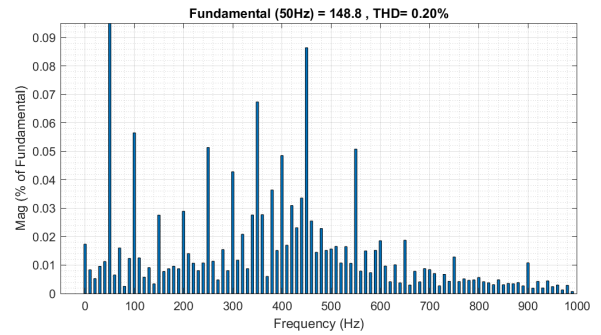**FIGURE 28.** THD of grid currents under dynamic voltage dip condition.

consistently low steady-state error, indicating excellent steady-state performance.

- In terms of THD, the order is MPC < DRL < PI, with the MPC controller generally maintaining lower THD values. The THD of the DRL controller is moderate and relatively consistent, making it a robust choice in terms of THD.
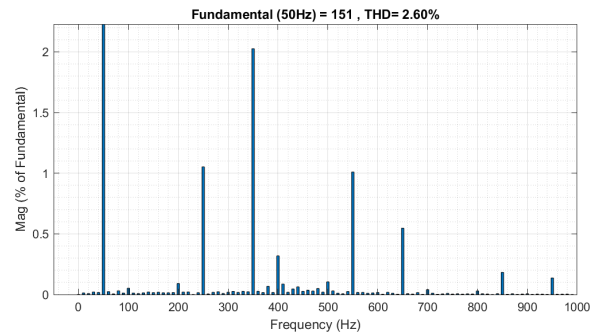
This comparative study underscores the superior stability and performance of the DRL-based controller, outperforming both conventional PI and MPC control methods when confronted with variable grid-filter inductance conditions.

E. CASE-V: ROBUST PERFORMANCE EVALUATION OF PROPOSED METHOD UNDER GRID DYNAMIC CONDITIONS

In this case, the controller's stability is rigorously assessed under dynamic grid voltage conditions, where voltage fluctuations range between 0.9 and 1.1 volts per unit. This dynamic scenario is deliberately induced from time t = 0.05 sec to t = 0.35 sec, simulating both voltage rise and voltage sag conditions. The Figure 26, illustrates that the controller

exhibits stability, effectively tracking the reference current even under voltage swell and sag dynamics. Furthermore, the THD measured under these conditions adheres to the IEEE 519 grid code specifications as shown in Fig. 27 - Fig. 28. This comprehensive evaluation underscores the controller's adeptness in maintaining stability and performance across diverse and dynamic operational scenarios.

V. EVALUATION OF DRL BASED CONTROLLER IN SOLAR GRID INTEGRATION APPLICATION

In various applications related to renewable energy and microgrids, the control strategy for Grid-Connected Converters (GCC) often adopts a nested-loop structure comprising an inner faster current control loop and a slower outer voltage control loop. This architecture generates d-axis and q-axis current references, denoted as i_d^*, i_q^* respectively, which are fed into the current loop controller. The d-axis loop is employed for regulating the dc-link voltage, while the q-axis loop serves for reactive power or grid voltage support control. The control operation involves a PI controller that generates the d-axis current reference through error signals between measured and reference dc-link voltages, along with a q-axis current reference derived from the difference between actual and desired reactive power.

To illustrate the concept further, Figure 29 provides a schematic representation of the DRL based controller within nested-loop control configuration. This configuration typically corresponds to the integration of distributed energy resources into the grid, as depicted in Fig 1. The grid is depicted on the right, while a RES, such as a solar power with boost converter, is depicted on the left. The power

TABLE 4. Boost converter parameters for solar PV system.

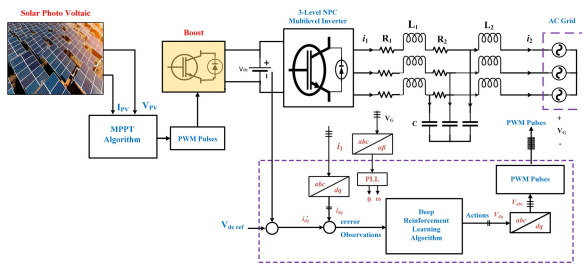| S. No	Parameter	Value
1	Maximum power (Pmp)	100 KW
2	Maximum voltage (Vmp)	290 V
3	Maximum current (Imp)	345.5 A
4	Boost Inductor	14.5 mH
5	Capacitor	3227 μF
6	DC link voltage	800 V
7	DC switching frequency	10 kHz

FIGURE 29. DRL based grid connected inverter current controller is used for solar PV grid integration application.
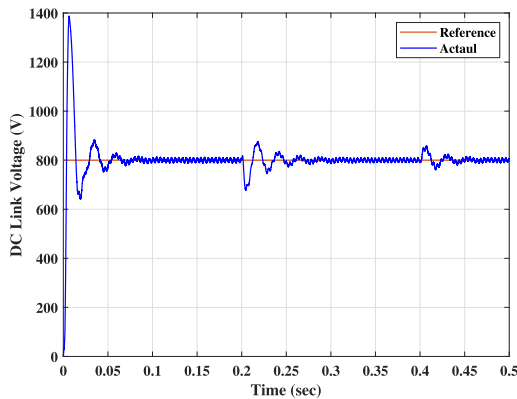

FIGURE 30. DC-Link voltage maintained at 800v using DRL based control technique.

FIGURE 31. dq-currents tracking performance in solar PV grid integration application under varying irradiance condition.
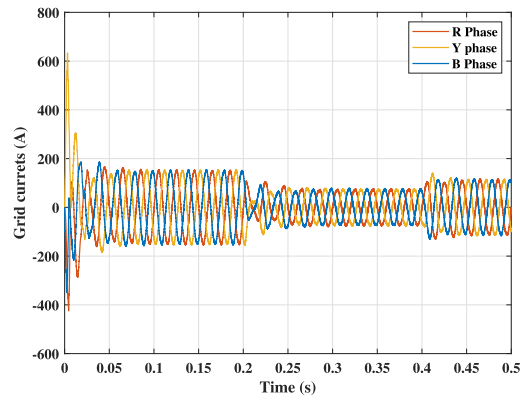
FIGURE 32. Three phase grid currents in solar PV grid integration application under varying irradiance condition.
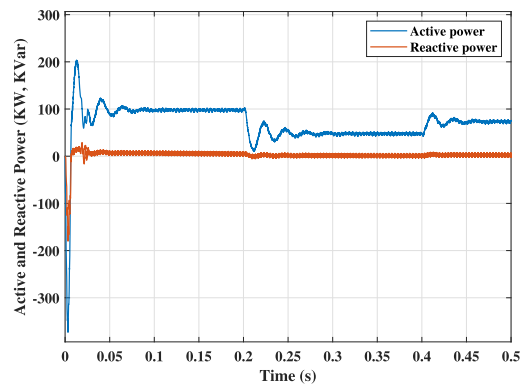
FIGURE 33. Active and reactive power waveforms in solar PV grid integration application using DRL based controller.

flow involves transfer from the RES to the grid through the dc-link capacitor and the GCI. For ease of reproducibility, the parameters of the boost converter and grid-side inverter are derived from common real-time research practices [68], [69], [70] and are detailed in Table. 4. Evaluating the performance of the DRL based control approach in the nested-loop structure, Fig 30 showcases its effectiveness.

Prior to t=0.2 seconds, the RES generates an active power of 100 kW at solar irradiance of $1000W/m^2$, and the GCC's reactive power reference is 0 Var, indicating that the GCI should absorb a small reactive power from the grid. Initially, the dc-link voltage is set at 800V. Despite the absence of synchronization control at the system's outset, the DRL

based control swiftly adjusts both the dc-link voltage and the GCI reactive power around their reference values. Notably, oscillations are minimal. t=0.2 seconds, the RES's irradiance shifts to $500W/m^2$, leading to reduced active power delivery to the grid through the dc-link and GCI. While the reactive power reference remains unchanged, the dc-link voltage falls. The DC-link vloltage profile is given in Fig.30 Yet, the DRL based controller rapidly regulates the dc-link voltage around the reference value. Further changes are demonstrated: at t=0.4 seconds with RES's irradiance shifts to $750W/m^2$, the dq- current tracking performance in solar grid connected

application is given in Fig.31. The grid currents and the active reactive power wave forms are given in Fig.32-Fig.33. All the figures illustrating that the DRL controller consistently exhibits excellent performance in meeting the requirements of nested-loop control, successfully adapting to various reference changes and efficiently regulating both dc-link voltage and reactive power. This reinforces the DRL based control approach's efficacy in ensuring stable and precise grid-connected converter operations.

VI. CONCLUSION

The research paper contributes significantly to the literature on GCIs by introducing and validating the TD3-based DRL control architecture for direct current control of a three-phase three-level NPC inverter. The assessment of grid current stability has been conducted under conditions involving training, testing, and fluctuating reference scenarios. The proposed technique has been experimentally validated. Real-time implementation in Hardware-in-the-Loop (HIL) mode using OPAL-RT and a DSP processor further confirms the feasibility of the DRL-based controller. The proposed control technique has undergone robustness analysis under parameter variations, grid dynamic conditions, showcasing enhanced performance compared to conventional and state-of-the-art control techniques such as the PI control and the MPC control method. It should be noted especially the successful validation of the DRL-based GCI current controller in solar PV grid integration applications. The study provides valuable information on the application of DRL in power electronics and demonstrates its effectiveness in achieving superior control performance. The findings pave the way for future research and development of advanced control strategies for grid integration of renewable energy sources. The outer loop PI controller can also be replaced by using multi-agent deep reinforcement learning control techniques as a future scope.

REFERENCES

[1] D. M. R. Korada and M. K. Mishra, "Fixed switching frequency model predictive current control for grid-connected inverter with improved dynamic and steady state performance," *IEEE Access*, vol. 11, pp. 104094–104105, 2023.

[2] P. Montero-Robina, C. Albea, F. Gómez-Estern, and F. Gordillo, "Hybrid modeling and control of three-level NPC rectifiers," *Control Eng. Pract.*, vol. 130, Jan. 2023, Art. no. 105374.

[3] R. Rosso, X. Wang, M. Liserre, X. Lu, and S. Engelken, "Grid-forming converters: Control approaches, grid-synchronization, and future trends—A review," *IEEE Open J. Ind. Appl.*, vol. 2, pp. 93–109, 2021.

[4] M. B. Saïd-Romdhane, M. W. Naouar, I. Slama-Belkhodja, and E. Monmasson, "Robust active damping methods for LCL filter-based grid-connected converters," *IEEE Trans. Power Electron.*, vol. 32, no. 9, pp. 6739–6750, Sep. 2017.

[5] U.-M. Choi, H.-H. Lee, and K.-B. Lee, "Simple neutral-point voltage control for three-level inverters using a discontinuous pulse width modulation," *IEEE Trans. Energy Convers.*, vol. 28, no. 2, pp. 434–443, Jun. 2013.

[6] J. Lai, X. Yin, X. Yin, and L. Jiang, "Fractional order harmonic disturbance observer control for three-phase LCL-type inverter," *Control Eng. Pract.*, vol. 107, Feb. 2021, Art. no. 104697.

[7] R. N. Beres, X. Wang, M. Liserre, F. Blaabjerg, and C. L. Bak, "A review of passive power filters for three-phase grid-connected voltage-source converters," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 4, no. 1, pp. 54–69, Mar. 2016.

[8] J. Kukkola and M. Hinkkanen, "Observer-based state-space current control for a three-phase grid-connected converter equipped with an LCL filter," *IEEE Trans. Ind. Appl.*, vol. 50, no. 4, pp. 2700–2709, Jul. 2014.

[9] D. Pérez-Estévez, J. Doval-Gandoy, A. G. Yepes, and Ó. López, "Positive- and negative-sequence current controller with direct discrete-time pole placement for grid-tied converters with LCL filter," *IEEE Trans. Power Electron.*, vol. 32, no. 9, pp. 7207–7221, Sep. 2017.

[10] S. Gulur, V. M. Iyer, and S. Bhattacharya, "A dual-loop current control structure with improved disturbance rejection for grid-connected converters," *IEEE Trans. Power Electron.*, vol. 34, no. 10, pp. 10233–10244, Oct. 2019.

[11] H. Yuan, H. S. Lam, N. Beniwal, S.-C. Tan, J. Pou, and S. R. Hui, "Direct-switch duty-cycle control of grid-connected N-level neutral-point-clamped converter," *IEEE Trans. Ind. Electron.*, vol. 70, no. 9, pp. 8624–8633, Sep. 2023.

[12] H. Wu and X. Wang, "Control of grid-forming VSCs: A perspective of adaptive fast/slow internal voltage source," *IEEE Trans. Power Electron.*, vol. 38, no. 8, pp. 10151–10169, Aug. 2023.

[13] P. Roncero-Sánchez, V. Feliu-Batlle, and A. García-Cerrada, "Design and comparison of state-feedback and predictive-integral current controllers for active- and reactive-power control in renewable energy systems," *Control Eng. Pract.*, vol. 17, no. 2, pp. 255–266, Feb. 2009.

[14] E. Heydari, A. Y. Varjani, and D. Diallo, "Fast terminal sliding mode control-based direct power control for single-stage single-phase PV system," *Control Eng. Pract.*, vol. 104, Nov. 2020, Art. no. 104635.

[15] J. Hu, "Improved dead-beat predictive DPC strategy of grid-connected DC–AC converters with switching loss minimization and delay compensations," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 728–738, May 2013.

[16] D. Zhi, L. Xu, and B. W. Williams, "Improved direct power control of grid-connected DC/AC converters," *IEEE Trans. Power Electron.*, vol. 24, no. 5, pp. 1280–1292, May 2009.

[17] J. P. Roselyn, C. P. Chandran, C. Nithya, D. Devaraj, R. Venkatesan, V. Gopal, and S. Madhura, "Design and implementation of fuzzy logic based modified real-reactive power control of inverter for low voltage ride through enhancement in grid connected solar PV system," *Control Eng. Pract.*, vol. 101, Aug. 2020, Art. no. 104494.

[18] J. Dannehl, F. W. Fuchs, S. Hansen, and P. B. Thøgersen, "Investigation of active damping approaches for PI-based current control of grid-connected pulse width modulation converters with LCL filters," *IEEE Trans. Ind. Appl.*, vol. 46, no. 4, pp. 1509–1517, Jul. 2010.

[19] J. Dannehl, F. W. Fuchs, and P. B. Thøgersen, "PI state space current control of grid-connected PWM converters with LCL filters," *IEEE Trans. Power Electron.*, vol. 25, no. 9, pp. 2320–2330, Sep. 2010.

[20] D. Pérez-Estévez, J. Doval-Gandoy, A. G. Yepes, Ó. López, and F. Baneira, "Enhanced resonant current controller for grid-connected converters with LCL filter," *IEEE Trans. Power Electron.*, vol. 33, no. 5, pp. 3765–3778, May 2018.

[21] J. Liu, Y. Yin, W. Luo, S. Vazquez, L. G. Franquelo, and L. Wu, "Sliding mode control of a three-phase AC/DC voltage source converter under unknown load conditions: Industry applications," *IEEE Trans. Syst. Man, Cybern. Syst.*, vol. 48, no. 10, pp. 1771–1780, Oct. 2018.

[22] G. V. Hollweg, P. J. D. de Oliveira Evald, R. V. Tambara, and H. A. Gründling, "A robust adaptive super-twisting sliding mode controller applied on grid-tied power converter with an LCL filter," *Control Eng. Pract.*, vol. 122, May 2022, Art. no. 105104.

[23] M. Bahloul, J. P. Gaubert, D. Mehdi, and S. K. Khadem, "Design of backstepping-sliding mode controller for a bidirectional grid connected DC/AC converter," in *Proc. 7th Int. Conf. Syst. Control (ICSC)*, 2018, pp. 288–293.

[24] F. Sebaaly, H. Vahedi, H. Y. Kanaan, N. Moubayed, and K. Al-Haddad, "Design and implementation of space vector modulation-based sliding mode control for grid-connected 3L-NPC inverter," *IEEE Trans. Ind. Electron.*, vol. 63, no. 12, pp. 7854–7863, Dec. 2016.

[25] R. Guzman, L. G. de Vicu na, M. Castilla, J. Miret, and H. Martin, "Variable structure control in natural frame for three-phase grid-connected inverters with LCL filter," *IEEE Trans. Power Electron.*, vol. 33, no. 5, pp. 4512–4522, May 2018.

[26] L. T. Martins, M. Stefanello, H. Pinheiro, and R. P. Vieira, "Current control of grid-tied LCL-VSI with a sliding mode controller in a multiloop approach," *IEEE Trans. Power Electron.*, vol. 34, no. 12, pp. 12356–12367, Dec. 2019.

[27] B. Long, W. Mao, P. Lu, J. Rodriguez, J. M. Guerrero, K. T. Chong, and Y. Teng, "Passivity fractional-order sliding-mode control of grid-connected converter with LCL filter," *IEEE Trans. Power Electron.*, vol. 38, no. 6, pp. 6969–6982, Jun. 2023.

[28] R. P. Vieira, L. T. Martins, J. R. Massing, and M. Stefanello, "Sliding mode controller in a multiloop framework for a grid-connected VSI with LCL filter," *IEEE Trans. Ind. Electron.*, vol. 65, no. 6, pp. 4714–4723, Jun. 2018.

[29] B. Long, P. J. Lu, K. T. Chong, J. Rodriguez, and J. Guerrero, "Robust fuzzy-fractional-order nonsingular terminal sliding-mode control of LCL-type grid-connected converters," *IEEE Trans. Ind. Electron.*, vol. 69, no. 6, pp. 5854–5866, Jun. 2022.

[30] G. V. Hollweg, R. V. Tambara, J. R. Massing, L. C. Borin, E. Mattos, G. G. Koch, C. R. D. Osório, and V. F. Montagner, "Model reference adaptive controllers with improved performance for applications in LCL-filtered grid-connected converters," *Control Eng. Pract.*, vol. 138, Sep. 2023, Art. no. 105591.

[31] T. Dragicevic, C. Zheng, J. Rodriguez, and F. Blaabjerg, "Robust quasi-predictive control of LCL-filtered grid converters," *IEEE Trans. Power Electron.*, vol. 35, no. 2, pp. 1934–1946, Feb. 2020.

[32] A. Mora, R. Cardenas, R. P. Aguilera, A. Angulo, P. Lezana, and D. D. Lu, "Predictive optimal switching sequence direct power control for grid-tied 3L-NPC converters," *IEEE Trans. Ind. Electron.*, vol. 68, no. 9, pp. 8561–8571, Sep. 2021.

[33] S. Vazquez, A. Marquez, R. Aguilera, D. Quevedo, J. I. Leon, and L. G. Franquelo, "Predictive optimal switching sequence direct power control for grid-connected power converters," *IEEE Trans. Ind. Electron.*, vol. 62, no. 4, pp. 2010–2020, Apr. 2015.

[34] N. N. Nam, M. Choi, and Y. I. Lee, "Model predictive control of a grid-connected inverter with LCL filter using robust disturbance observer," *IFAC-PapersOnLine*, vol. 52, no. 4, pp. 135–140, 2019.

[35] W. R. Sultana, S. K. Sahoo, S. Sukchai, S. Yamuna, and D. Venkatesh, "A review on state of art development of model predictive control for renewable energy applications," *Renew. Sustain. Energy Rev.*, vol. 76, pp. 391–406, Sep. 2017.

[36] J. Scoltock, T. Geyer, and U. K. Madawala, "Model predictive direct power control for grid-connected NPC converters," *IEEE Trans. Ind. Electron.*, vol. 62, no. 9, pp. 5319–5328, Sep. 2015.

[37] N. Panten, N. Hoffmann, and F. W. Fuchs, "Finite control set model predictive current control for grid-connected voltage-source converters with LCL filters: A study based on different state feedbacks," *IEEE Trans. Power Electron.*, vol. 31, no. 7, pp. 5189–5200, Jul. 2016.

[38] E. Liu, Y. Han, A. S. Zalhaf, P. Yang, and C. Wang, "Performance evaluation of isolated three-phase voltage source inverter with LC filter adopting different MPC methods under various types of load," *Control Eng. Pract.*, vol. 135, Jun. 2023, Art. no. 105520.

[39] J. S. Costa, A. Lunardi, P. C. Ribeiro, I. B. Da Silva, D. A. Fernandes, and A. J. S. Filho, "Performance-based tuning for a model predictive direct power control in a grid-tied converter with L-filter," *IEEE Access*, vol. 11, pp. 8017–8028, 2023.

[40] W. Yang, S. Miao, Z. Liu, J. Han, Y. Xiong, and Q. Tu, "Model predictive direct power control of grid-connected converters considering unbalanced filter inductance and grid conditions," *J. Modern Power Syst. Clean Energy*, vol. 9, no. 6, pp. 1279–1288, Nov. 2021.

[41] J. R. Fischer, S. A. González, I. Carugati, M. A. Herrán, M. G. Judewicz, and D. O. Carrica, "Robust predictive control of grid-tied converters based on direct power control," *IEEE Trans. Power Electron.*, vol. 29, no. 10, pp. 5634–5643, Oct. 2014.

[42] P. Li, R. Li, Y. He, and J. Zhang, "Adaptive finite control set model predictive control for three-phase inverters connected to distorted grid with fewer voltage sensors," *Control Eng. Pract.*, vol. 116, Nov. 2021, Art. no. 104936.

[43] G. V. Hollweg, P. J. D. D. O. Evald, E. Mattos, L. C. Borin, R. V. Tambara, and V. F. Montagner, "Self-tuning methodology for adaptive controllers based on genetic algorithms applied for grid-tied power converters," *Control Eng. Pract.*, vol. 135, Jun. 2023, Art. no. 105500.

[44] J. Liu, X. Sun, B. Ren, W. Song, and P. Wheeler, "Strong adaptability control based on dual-division-summation current control for an LCL-type grid-connected inverter," *IEEE Trans. Power Electron.*, vol. 37, no. 12, pp. 14157–14172, Dec. 2022.

[45] F. Gonzalez-Espin, I. Patrao, E. Figueres, and G. Garcera, "An adaptive digital control technique for improved performance of grid connected inverters," *IEEE Trans. Ind. Informat.*, vol. 9, no. 2, pp. 708–718, May 2013.

[46] P. Li, X. Tong, Y. He, and T. Chen, "Adaptive model predictive control for PUC grid-connected inverter system with inaccurate parameters," *Control Eng. Pract.*, vol. 139, Oct. 2023, Art. no. 105652.

[47] M. Stefanello, J. R. Massing, and R. P. Vieira, "Robust control of a grid-connected converter with an LCL-filter using a combined sliding mode and adaptive controller in a multi-loop framework," in *Proc. 41st Annu. Conf. IEEE Ind. Electron. Soc.*, Nov. 2015, pp. 3726–3731.

[48] O. Zandi and J. Poshtan, "Voltage control of DC–DC converters through direct control of power switches using reinforcement learning," *Eng. Appl. Artif. Intell.*, vol. 120, Apr. 2023, Art. no. 105833.

[49] 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.*, vol. 70, no. 3, pp. 2887–2896, Mar. 2023.

[50] M. Hajihosseini, M. Andalibi, M. Gheisarnejad, H. Farsizadeh, and M. Khooban, "DC/DC power converter control-based deep machine learning techniques: Real-time implementation," *IEEE Trans. Power Electron.*, vol. 35, no. 10, pp. 9971–9977, Oct. 2020.

[51] M. Gheisarnejad, A. Akhbari, M. Rahimi, B. Andresen, and M.-H. Khooban, "Reducing impact of constant power loads on DC energy systems by artificial intelligence," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 12, pp. 4974–4978, Dec. 2022.

[52] L. Avila, M. De Paula, I. Carlucho, and C. Sanchez Reinoso, "MPPT for PV systems using deep reinforcement learning algorithms," *IEEE Latin Amer. Trans.*, vol. 17, no. 12, pp. 2020–2027, Dec. 2019.

[53] V.-H. Bui, T.-T. Nguyen, and H.-M. Kim, "Distributed operation of wind farm for maximizing output power: A multi-agent deep reinforcement learning approach," *IEEE Access*, vol. 8, pp. 173136–173146, 2020.

[54] R. Loka, R. Dubey, and A. M. Parimi, "Coordinated load frequency control of a smart hybrid power system using the DEMA-TD3 algorithm," *Control Eng. Pract.*, vol. 134, May 2023, Art. no. 105480.

[55] Y. Wang, S. Fang, J. Hu, and D. Huang, "A novel active disturbance rejection control of PMSM based on deep reinforcement learning for more electric aircraft," *IEEE Trans. Energy Convers.*, vol. 38, no. 2, pp. 1461–1470, 2023.

[56] Y. Wang, S. Fang, J. Hu, and D. Huang, "Multiscenarios parameter optimization method for active disturbance rejection control of PMSM based on deep reinforcement learning," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 10957–10968, Nov. 2023.

[57] S. Bhattacharjee, S. Halder, Y. Yan, A. Balamurali, L. V. Iyer, and N. C. Kar, "Real-time SIL validation of a novel PMSM control based on deep deterministic policy gradient scheme for electrified vehicles," *IEEE Trans. Power Electron.*, vol. 37, no. 8, pp. 9000–9011, Aug. 2022.

[58] Y. Wang, S. Fang, and J. Hu, "Active disturbance rejection control based on deep reinforcement learning of PMSM for more electric aircraft," *IEEE Trans. Power Electron.*, vol. 38, no. 1, pp. 406–416, Jan. 2023.

[59] Y. Wang, S. Fang, and D. Huang, "An improved model-free active disturbance rejection deadbeat predictive current control method of PMSM based on data-driven," *IEEE Trans. Power Electron.*, vol. 38, no. 8, pp. 9606–9616, Aug. 2023.

[60] Y. Jiang, W. Gao, J. Na, D. Zhang, T. T. Hämäläinen, V. Stojanovic, and F. L. Lewis, "Value iteration and adaptive optimal output regulation with assured convergence rate," *Control Eng. Pract.*, vol. 121, Apr. 2022, Art. no. 105042.

[61] J. Wu, Q. M. J. Wu, S. Chen, F. Pourpanah, and D. Huang, "A-TD3: An adaptive asynchronous twin delayed deep deterministic for continuous action spaces," *IEEE Access*, vol. 10, pp. 128077–128089, 2022.

[62] T. Zhao, M. Zhang, C. Wang, and Q. Sun, "Model-free predictive current control of three-level grid-connected inverters with LCL filters based on Kalman filter," *IEEE Access*, vol. 11, pp. 21631–21640, 2023.

[63] C. Cui, N. Yan, B. Huangfu, T. Yang, and C. Zhang, "Voltage regulation of DC–DC buck converters feeding CPLs via deep reinforcement learning," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 3, pp. 1777–1781, Mar. 2022.

[64] P. Falkowski and A. Sikorski, "Finite control set model predictive control for grid-connected AC–DC converters with LCL filter," *IEEE Trans. Ind. Electron.*, vol. 65, no. 4, pp. 2844–2852, Apr. 2018.

[65] J. Scoltock, T. Geyer, and U. K. Madawala, "A model predictive direct current control strategy with predictive references for MV grid-connected converters with LCL-filters," *IEEE Trans. Power Electron.*, vol. 30, no. 10, pp. 5926–5937, Oct. 2015.

[66] E. Figueres, G. Garcera, J. Sandia, F. Gonzalez-Espin, and J. C. Rubio, "Sensitivity study of the dynamics of three-phase photovoltaic inverters with an LCL grid filter," *IEEE Trans. Ind. Electron.*, vol. 56, no. 3, pp. 706–717, Mar. 2009.

[67] S. Alepuz, S. Busquets-Monge, J. Bordonau, J. A. Martinez-Velasco, C. A. Silva, J. Pontt, and J. Rodriguez, "Control strategies based on symmetrical components for grid-connected converters under voltage dips," *IEEE Trans. Ind. Electron.*, vol. 56, no. 6, pp. 2162–2173, Jun. 2009.

[68] I. Poonahela, S. Bayhan, H. Abu-Rub, M. M. Begovic, and M. B. Shadmand, "An effective finite control set-model predictive control method for grid integrated solar PV," *IEEE Access*, vol. 9, pp. 144481–144492, 2021.

[69] B. Zhang, W. Wu, Y. Yang, N. Gao, J. Chen, E. G. Koutroulis, H. S. Chung, M. Liserre, and F. Blaabjerg, "A novel simplified finite control set repeat model predictive control for grid-connected inverters," *IEEE Trans. Ind. Electron.*, vol. 70, no. 11, pp. 11324–11333, Nov. 2023.

[70] S. He, D. Zhou, X. Wang, and F. Blaabjerg, "Passivity-based multisampled converter-side current control of LCL-filtered VSCs," *IEEE Trans. Power Electron.*, vol. 37, no. 11, pp. 13848–13860, Nov. 2022.

SRI PHANI KRISHNA KARRI (Member, IEEE) received the M.Tech. degree in signal processing and the Ph.D. degree in multidimensional signal processing and machine learning from IIT Kharagpur, in 2011 and 2017, respectively. He is currently an Assistant Professor with the Electrical Engineering Department, NIT Andhra Pradesh. Following the Ph.D. degree, he expanded his expertise during a Postdoctoral Fellowship with IIT Kharagpur, from 2017 to 2018. During this phase, he delved into developing deep reinforcement learning (RL)-based image analysis frameworks and pushing the boundaries of image analysis by integrating advanced deep RL techniques. His research interests include signal processing, machine learning, and image analysis.

MAMDOUH L. ALGHAYTHI (Member, IEEE) received the B.S. degree in electrical engineering from Jouf University, Sakakah, Saudi Arabia, in 2012, the M.S. degree in electrical engineering from Southern Illinois University, Carbondale, IL, USA, in 2015, and the Ph.D. degree in electrical engineering from the University of Missouri, Columbia, MO, USA, in 2020. He is currently an Assistant Professor with the Department of Electrical Engineering, Jouf University, where he is also the Chair of the Electrical Engineering Department and the Coordinator of the Master Program in Renewable Energy. He has authored or coauthored many journal articles and conference papers. His current research interests include design and modeling of power electronics converters, renewable energy systems, reliability aspects of DC–DC converters, smart grids, energy management, and microgrids. He is a member of the IEEE Power Electronics Society.

ANUGULA RAJAMALLAIAH (Graduate Student Member, IEEE) was born in June 1991. He received the bachelor's degree in electrical and electronics engineering from JNTU Hyderabad, Telangana, India, in 2012, and the master's degree in instrumentation and control from NIT Calicut, Kerala, India, in 2015. He is currently pursuing the Ph.D. degree with NIT Andhra Pradesh, specializing in deep reinforcement learning implementation for power converters control in electrical engineering applications. He was an Assistant Professor with the SR Engineering College, Warangal (currently SR University), from July 2015 to August 2020. His research interest includes optimal control of power converters in electrical engineering application using deep reinforcement learning.

MESHARI S. ALSHAMMARI (Member, IEEE) received the B.Sc. degree in electrical engineering from the University of Hail, Saudi Arabia, in 2014, and the M.E. degree in energy systems engineering and the Ph.D. degree in electrical and electronics engineering from the University of Galway, in 2017 and 2022, respectively. He is currently an Assistant Professor with the Electrical Engineering Department, Jouf University, Sakaka, Saudi Arabia. His research interests include the analysis of AC and DC distribution systems, DC/AC power converters for utility grid and renewable energy applications, and renewable energy integration systems. In addition, he was awarded the Best Poster at the Seventh NUI Galway Research Day, in 2017, and Best Paper at ICRERA, in 2021.

● ● ●