IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Implementation of Deep Reinforcement Learning for Model-free Switching And Control of a 23-level Single DC Source Hybrid Packed U-Cell (HPUC)

POURIA QASHQAI[1], (Member, IEEE), MOHAMMAD BABAIE[1], (Member, IEEE), RAWAD ZGHEIB[2], (Member, IEEE), KAMAL AL-HADDAD[1], (Life Fellow, IEEE)

[1]Department of Electrical Engineering, École de Technologie Supérieure (ÉTS), Montreal, QC H3C 1K3, Canada
[2]Institut de recherche d'Hydro-Québec (IREQ), Varennes, QC J3X 1S1, Canada

Corresponding author: Pouria Qashqai (e-mail: pouria.qashqai.1@ens.etsmtl.ca).

ABSTRACT This paper proposes a novel Deep Reinforcement Learning (DRL) method for controlling a 23-level Single DC Source Hybrid Packed U-Cell (HPUC) converter. The HPUC topology generates a high number of voltage levels with minimal components but presents control challenges due to its numerous switching states and dynamic charging behavior. Unlike conventional control methods, which require accurate models and are sensitive to noise and parameter mismatches, DRL offers a model-free and resilient approach to the non-linear control of such complex systems. A Deep Q-Network (DQN) agent which is inherently model-free and suited for high-dimensional state spaces and discrete action spaces, is employed to address these issues. To validate the proposed method, simulations were conducted in the MATLAB/Simulink environment. The obtained results demonstrated the satisfactory performance of the proposed DRL method, achieving a Total Harmonic Distortion (THD) of 2.71% in the output current under steady-state, maintaining stable capacitor voltage balancing, and exhibiting rapid dynamic response (e.g., settling within approximately 40 ms for current step changes). Furthermore, its resilience was highlighted by its ability to maintain control despite a 25dB SNR noise condition and up to 15% variations in capacitor values.

INDEX TERMS Deep Reinforcement Learning (DRL), Hybrid PUC (H-PUC), Intelligent Control, Machine Learning (ML), Multilevel Inverter (MLI), Packed U-Cell (PUC)

I. INTRODUCTION

POWER electronics converters play a crucial role in modern life due to their numerous applications in renewable energies, smart grids, and electric vehicles. This necessitates the implementation of advanced converter topologies requiring more efficient and reliable control methods [1]–[4].

One of the prominent examples of such topologies is the 23-level Single DC Source Hybrid Packed U-cell (H-PUC) Converter [5]. This converter generates a high number of voltage levels with a reduced component count by combining two Packed U-cell (PUC) [6], [7] converters. However, its numerous switching states and the time-varying behavior of the DC-link capacitors lead to high control complexity.

Linear control methods (e.g., PID) are straightforward but struggle with the non-linearities of modern converters [8]. Non-linear schemes such as Feedback Linearization and Sliding Mode Control improve robustness, but the former needs an accurate model and the latter can require high switching frequencies [9], [10]. Fuzzy Logic Control, while powerful in handling uncertainties, depends heavily on the quality of its rule base [11]. Genetic Algorithm-Based Control can optimize effectively but is computationally expensive [12]. Other advanced approaches, including H-Infinity Control [13], Predictive Functional Control [14], and Kalman Filter-Based Control [15], also exhibit high computational costs and sensitivity to parameter variations. Finally, Model Predictive Control (MPC) [16]–[18] predicts future behavior to choose optimal actions, but also requires significant computing power and an accurate model, making it prone to degradation under parameter variations, noise, and uncertainties.

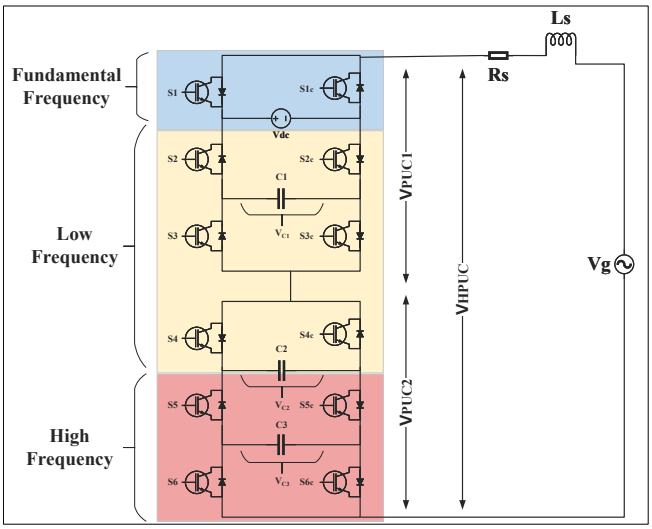

FIGURE 1. **Topology of a 23-level Hybrid Packed U-cell (HPUC) comprised of two cascaded PUC5 converters with capacitor voltages V_{C1}, V_{C2}, and V_{C3} indicated.**

TABLE 1. **Switching states and capacitor charging directions of the upper leg (PUC1) and the lower leg (PUC2). The upward arrow (↑) indicates capacitor charging, while the downward arrow (↓) indicates capacitor discharging and "–" indicates no change in capacitor charge.**

Upper Leg (PUC1)					
State	S_1	S_2	S_3	C_1	V_{PUC1}
1	0	0	0	–	0
2	0	0	1	↓	$-V_{C1}$
3	0	1	0	↑	$V_{C1} - V_{dc}$
4	0	1	1	–	$-V_{dc}$
5	1	0	0	–	V_{dc}
6	1	0	1	↑	$V_{dc} - V_{C1}$
7	1	1	0	↓	V_{C1}
8	1	1	1	–	0
Lower Leg (PUC2)					
State	S_4	S_5	S_6	C_2	C_3	V_{PUC2}
1	0	0	0	–	–	0
2	0	0	1	–	↓	$-V_{C3}$
3	0	1	0	↓	↑	$V_{C3} - V_{C2}$
4	0	1	1	↑	–	$-V_{C2}$
5	1	0	0	–	–	V_{C2}
6	1	0	1	↑	–	$V_{C2} - V_{C3}$
7	1	1	0	↓	–	V_{C3}
8	1	1	1	–	–	0

Recent advancements in artificial intelligence, particularly machine learning (ML), along with increased commercially available computational power, have increased the popularity of AI in power electronics research [19]. Some studies have investigated machine learning for modeling [20]–[23] as well as control [24]–[26] of power electronics.

Deep reinforcement Learning (DRL) [27], a subset of machine learning, is a modern control method that combines the generalization power of deep neural networks [28] with reinforcement learning. It shows excellent potential in complex non-linear systems such as power systems with a notable degree of uncertainty in the environment as well as parameter variations [19]. DRL by learning directly from interaction with the system, can overcome t,he need for precise mathematical models, a significant challenge for complex topologies like the HPUC. This model-free approach inhrenetly offers better resilience to unmodeled dynamics, parameter mismatch, and noise compared to model-dependent techniques like MPC or feedback linearization. DRL works by learning the optimal control policy through trial and error via interactions with the system. Despite requiring large training data, high computational power, and challenging reward function procedure, DRL is used in many areas that require advanced non-linear control such as robotics [29], [30], video games [31], and electric vehicles [32]. Its applications in power systems and power electronics include but are not limited to the optimization of power distribution [33], energy management of distributed energy sources [34], dynamic load management [35], Integrated Energy Systems (IES) management [36], and smart-grid operations [37].

Only recently, DRL has been used as a method for control of power electronics converters [38]. In [39], DRL is applied as a solution to the shortcomings of conventional methods for control of DC/DC buck converter when feeding constant power load (CPL). For large dynamic changes in CPLs, conventional control algorithms often do not demonstrate satisfactory performance. Even though the DRL method proposed in this study has mitigated this issue and can deliver satisfactory performance, the DRL agent is utilized for gain tuning in the feedback loop. Therefore, not only is auxiliary control required, but also it does not take advantage of a significant benefit of using DRL which is its model-free nature.

In [40], a model-free DRL control is proposed for a DC/DC buck converter with CPL showing strong dynamic performance despite significant CPL changes. However, its reliance on precise measurements leads to sensitivity to noise and error accumulation between the model and the actual converter.

In [41], DRL is used to control a DC/DC buck converter to be resilient against uncertainties and parameter changes. However, it relies on an offline pre-trained converter model for an extended state observer (ESO) to adjust to variations by observing differences between the model and the converter.

Research into applying DRL in power electronics is growing and gaining popularity. However, its advantages in control of power electronics converters are still in the early stages [19]. Some studies [42], focus on DRL control for simpler converters. Others, such as [43], perform real-time simulations as a safe substitute for experimental results under harsh operational points. Some studies explore more complex converters like the Neutral Point Clamped (NPC) converter, as seen in [44], [45]). While [45] shows promising steady-state performance, it lacks testing under dynamic changes, noise, and uncertainties.

Based on the control and switching method proposed in [45], a DRL-based control method for NPC is developed in [46]. The method proposed in this paper is shown to demonstrate satisfactory performance in steady-state as well

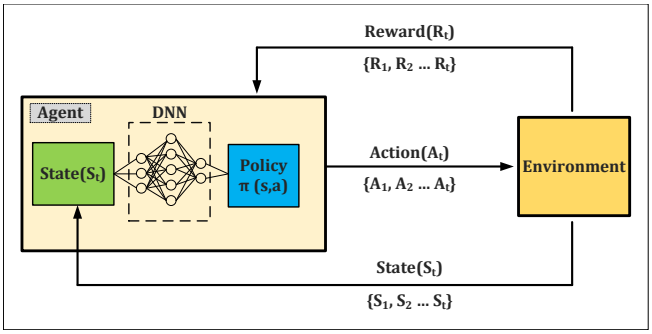

FIGURE 2. Diagram of DRL agent aiming to obtain the optimal policy of $\pi(s, a)$ at the time step of t

as dynamic changes. Moreover, the proposed DRL method is compared against MPC through simulation and experimental results and it is proven that the DRL method is more resilient against uncertainties, noise, and parameter variations. However, the proposed method does not explore implementation on a fairly complex topology with demanding control considerations such as HPUC. To explore this gap in the literature, this paper proposes a similar implementation of DRL for the 23-level H-PUC converter, aiming to take advantage of DRL's adaptability and robustness to control an efficient converter that demands a challenging control. The main objective of this research is to demonstrate the feasibility and effectiveness of DRL in switching and controlling the complex 23-level H-PUC converter.

This paper is structured as follows. Section II introduces the modern 23-level HPUC topology, highlighting its strengths and weaknesses, especially emphasizing its control challenges. In section III, the fundamentals of DRL are discussed and the proposed control method is presented in section IV. Moreover, Section V is dedicated to results. In this section, simulation results obtained from the MATLAB/Simulink environment are presented and analyzed. Lastly, the findings of this paper are discussed in section VI.

II. SINGLE DC SOURCE HYBRID PACKED U-CELL (HPUC)

The Hybrid Packed U-Cell also known as HPUC [5] as shown in Fig. 1, is a modern topology based on two cascaded Packed U-Cell (PUC) [6] converters in order to generate multiple voltage levels in the output. Instead of a second DC source for the lower leg PUC converter, a capacitor is used. The PUC in the upper leg is responsible for generating low-frequency voltage generation whereas the lower leg is mostly responsible for generating high-frequency voltages. This modification enables HPUC to generate multiple voltage levels using a single DC source despite making the control of the HPUC significantly more complex compared to PUC.

Despite providing these advantages over other combinations of PUC topology, voltage balancing in HPUC is rather challenging. As seen in Table 1, each sub-module consists of six switches and each switch has two states of ''on'' and ''off''. Thus the total number of combinations among the switches is $2^3 = 8$ per sub-module. Therefore, the whole topology has a combination of $8 \times 8 = 64$ switching states as shown in (1).

$$S_i = \overline{S_i} = \begin{cases} 1 & \text{if } S_i \text{ is on} \\ 0 & \text{if } S_i \text{ is off} \end{cases} \qquad i = 1, 2, \ldots, 6 \qquad (1)$$

Where S_i and $\overline{S_i}$ are complementary switches that are never on and off simultaneously to prevent short circuits and other complications.

As shown in Table 1, depending on the instantaneous voltages across each capacitor and corresponding to different switching combinations, various currents will be drawn from them. The arrows in Table 1 indicate the direction of capacitor current: $\downarrow$ indicates discharging (current flowing out of the positive terminal), and $\uparrow$ indicates charging (current flowing into the positive terminal). Thus, the voltage of each capacitor is dependent on the capacitor in its upper position which makes voltage balancing relatively challenging. Therefore, conventional linear control methods will struggle to deliver satisfactory performance when handling such a complex system let alone being resistant to uncertainties. To mitigate this issue, a model predictive control (MPC) capable of effectively regulating the capacitor voltages and efficiently generating the output voltages is proposed in [5]. Using DRL the mathematical model of the HPUC converter is not required. Similarly, voltage balancing equations are not required to be obtained because the DRL agent spontaneously discovers the optimal policy for voltage balancing based on observations and reward functions. Therefore, the inherent model-free characteristics of the DRL method can provide advantages in scenarios where a model of the HPUC converter is either not readily accessible or obtaining one proves to be challenging. In the next section, the DRL method is described in detail.

III. FUNDAMENTALS OF DEEP REINFORCEMENT LEARNING (DRL)

In this section initially the fundamentals of deep enforcement learning will be introduced. Then the proposed control method based on DRL for intelligent control of HPUC will be discussed. Lastly, the advantages and limitations of the proposed DRL method will be discussed.

A. REINFORCEMENT LEARNING

Deep Reinforcement Learning (DRL) combines deep neural networks (DNN) with reinforcement learning (RL) to control complex and non-linear systems. In DRL, as shown in Fig. 2, an agent interacts with the environment and selects actions based on the observed states and received rewards. The objective is to update the policy of $\pi(s, a)$ in each iteration to achieve the optimal policy based on actions A_t applied to the environment and inspecting the observations S_t and calculating the instantaneous and accumulative rewards of R_t where t is the time step of the experience episode.

Reinforcement learning can be expressed as a four-tuple Markov Decision Process (MDP) of $\{s, a, pa, ra\}$ to find an

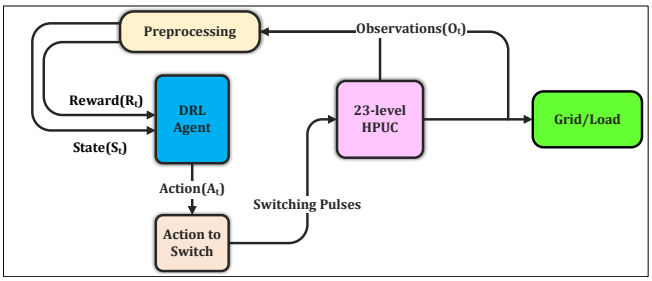

FIGURE 3. **Block diagram of the proposed DRL method for model-free control of a grid-connected 23-level single DC source HPUC**

optimal policy of π^*, to maximize long-term rewards:

$$\pi^* = \arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t \mid s_0, \pi\right] \tag{2}$$

Where π^* is the optimal policy, t is the time step, γ is the discount factor which is a number between 0 and 1 indicating how actions affect the accumulative reward in the long term. Ultimately, s_0 is the initial state and r_t is the instantaneous reward at the time step of t.

There are various algorithms to achieve the aforementioned optimal policy for a given RL agent. The most prominent ones in the literature are as follows:

- **Value-based** which solves the Bellman optimality equation to find the best value function for selecting actions in each state.
- **Policy-based** which alternates between estimating and improving a policy based on its value function.
- **Q-Learning** which is a model-free and non-policy method that estimates the value of state action pairs and updates them based on its calculated rewards.
- **Actor-Critic Methods** which combines policy-based (actor) and value-based (critic) methods, where the actor chooses actions and the critic evaluates them.

B. Q-LEARNING

Both value iteration and policy iteration algorithms are model-based. Unlike such algorithms, Q-learning [47] is an off-policy model-free algorithm that learns the action-value function directly from experience episodes without requiring an accurate model of the environment. It approximates the values i.e. ''Q-values'' corresponding to taking each action in a given state. Q-learning uses (3) to update Q-values in each iteration of interactions with the environment.

$$Q(s,a) \leftarrow Q(s,a) + \alpha\Big[R(s,a) + \gamma \max_{a'} Q(s',a') - Q(s,a)\Big] \tag{3}$$

Where $Q(s,a)$ is the Q-value for taking the action of a in the state of s, while α is the learning rate, $R(s,a)$ is the instantaneous reward after taking action of a in the state of s, γ is the discount factor and ultimately, a' and s' are the next action and state after taking action a in the state of s. Using this equation, the new value of $Q(s,a)$ is updated after each iteration.

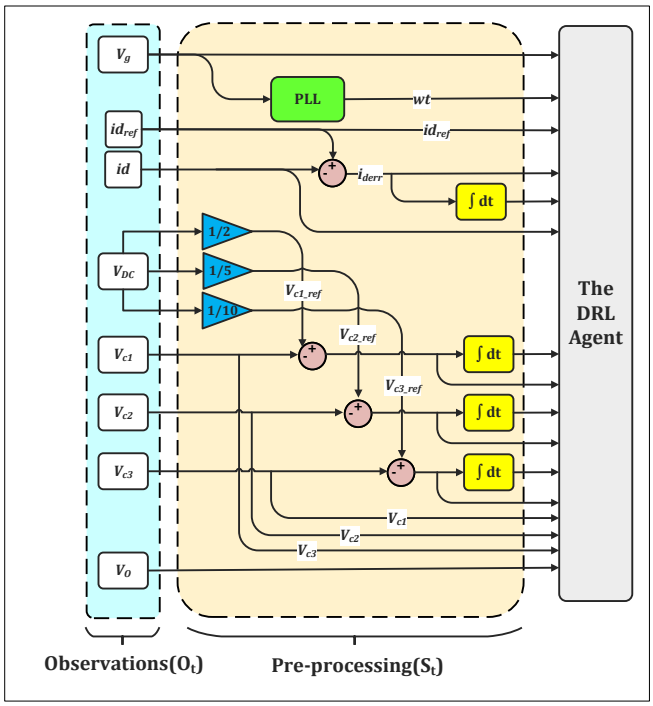

FIGURE 4. **Block Diagram of the state space generation unit comprised of observations (measurements) and pre-processing**

C. DEEP REINFORCEMENT LEARNING

As stated earlier, conventional reinforcement learning algorithms use numerical methods to find the optimal policies or values. Despite being easier to implement and less demanding in terms of computational burden, such algorithms are not suited for fairly complex and high-dimensional applications. To mitigate these limitations, deep neural networks (DNNs) are integrated into reinforcement learning, giving rise to a modern approach known as "Deep Reinforcement Learning". Taking advantage of neural networks not only enables DRL to handle high-dimensional and complex problems but also leverages the data-driven nature of neural networks. Being data-driven empowers DRL algorithms to learn optimal control of various applications without acquiring an accurate model of their corresponding systems. Additionally, neural networks are excellent for generalization. Thus, these types of control methods are well suited for non-linear control schemes where unseen data is likely to degrade the performance of conventional non-linear control methods.

In this paper, a DNN-based Q-learning method called Deep Q-Network (DQN), is used due to its suitability for model-free environments and discrete action spaces.

IV. THE PROPOSED CONTROL METHOD BASED ON REINFORCEMENT LEARNING

Using an approach similar to [46], a control method based on DRL is developed and proposed in this section.

TABLE 2. Popular DRL Agent Types Supported by MATLAB.

Agent	Type	Action Space	On/Off Policy
Q-Learning (Q) Agents	Value-Based	Discrete	Off
SARSA Agents	Value-Based	Discrete	On
Deep Q-Network (DQN) Agents	Value-Based	Discrete	Off
Policy Gradient (PG) Agents	Policy-Based	Discrete or Continuous	On
Actor-Critic (AC) Agents	Actor-Critic	Discrete or Continuous	On
Deep Deterministic Policy Gradient (DDPG) Agents	Actor-Critic	Continuous	Off
Twin-Delayed Deep Deterministic Policy Gradient (TD3) Agents	Actor-Critic	Continuous	Off
Soft Actor-Critic (SAC) Agents	Actor-Critic	Continuous	Off
Proximal Policy Optimization (PPO) Agents	Actor-Critic	Discrete or Continuous	On
Trust Region Policy Optimization (TRPO) Agents	Actor-Critic	Discrete or Continuous	On
Model-Based Policy Optimization (MBPO) Agents	Actor-Critic	Discrete or Continuous	Off

A. INTRODUCTION

The overall diagram of the proposed method is shown in Fig. 3. As seen, the proposed method consists of a pre-processing block that feeds the observation signals obtained via measurements to the DRL agent. Using these signals that act as the state space, the agent can estimate the model of the converter and learn its dynamic behavior. This block is equal to the input layer of the DNN block as shown in Fig. 2. The DRL the agent finds the optimal control policy based on previous experience, current observations, and the accumulated reward.

The part of the DRL agent block responsible for the generation of actions is equal to the output layer of the DNN. To improve efficiency and to reduce training times of the agent. Two actions are generated. One for the upper leg and one for the lower leg of the HPUC. In the literature, this approach is known as multi-dimensional action space. This enables the agent to observe and learn the effect of switching states on each corresponding leg, resulting in a faster convergence to the optimal switching. Finally, the "actions to switch" block takes the actions from the DRL agent which are discrete values, usually integers. This block then maps these values to their corresponding switching states for the upper leg and lower leg of the HPUC using a look-up table.

In the following subsections, initially, the agent type is selected and justified. Moreover, observations (signals measured) are selected to generate the state space, and the idea behind choosing each one is discussed. Furthermore, to effectively achieve the optimal control policy the reward functions specific to this application are designed. Lastly, the action space and the method of mapping the numerical actions generated by the proposed DRL agent to the switching gates of the HPUC converter are introduced.

B. CHOOSING THE AGENT TYPE

The most popular agent types supported by MATLAB [48] with their characteristics are described in Table 2 An off-policy agent is well-suited for the model-free approach proposed in this paper. Q-learning-based agents are preferred in comparison to actor-critic-based agents due to the sensitivity of the actor-critic approach towards hyper-parameters. In this paper, Deep Q-Network (DQN), a form of Q-learning algorithm is used for training the DRL agent. The DQN agent type is not only model-free but also is excellent for handling high-dimensional state spaces in which, action space is discrete. These characteristics make it meet the requirements for model-free control of a complex converter with discrete action space (i.e. switching states) such as HPUC.

C. CREATING THE STATE SPACE (OBSERVATIONS)

Different measurements are required to be applied to the DRL agent to effectively capture the dynamic behavior of the converter. These measurements are called observations in the literature. The following equation demonstrates the observation array of the DRL agent:

$$O_t = \left[i_d,\ v_g, v_o, V_{DC}, i_{dref}, v_{c1}, v_{c2}, v_{c3}\right]^T \tag{4}$$

Where O_t is the observations array, i_d is the output current, v_g is the grid voltage, v_o is the output voltage of the converter, V_{DC} is the input DC voltage, i_{dref} is the reference current, and finally v_{c1}, v_{c2}, and v_{c3} are the voltages across capacitors c_1, c_2 and c_3, respectively.

As illustrated in Fig. 4, the pre-processing unit manipulates the measurements provided by the observations unit to form the ultimate state space of the controller. The generated state space of the controller, marked as S_t is shown in (5):

$$\begin{aligned} S_t = \Big[& i_d,\ v_g,\ v_o,\ v_{c1},\ v_{c2},\ v_{c3}, \\ & i_{derr},\ \int i_{derr},\ i_{dref},\ \omega t,\ v_{c1err},\ v_{c2err},\ v_{c3err}, \\ & \int v_{c1err},\ \int v_{c2err},\ \int v_{c3err} \Big]^T \end{aligned} \tag{5}$$

Where i_{derr} is the instantaneous error between the reference current i_d and actual output current i_g while $\int i_{derr}$ is its accumulative error. Similarly v_{c1err}, v_{c2err}, and v_{c3err} are the errors between actual and reference voltages of capacitors, c_1, c_2, and c_3 respectively while $\int v_{c1err}$, $\int v_{c2err}$, and $\int v_{c3err}$ are their accumulative counterparts. Ultimately ωt is the angular speed of the grid voltage equal to $2\pi f$ where f is the line frequency of v_g.

The diagram of the observations and pre-processing block is depicted in Fig. 4. As illustrated, in order to capture the

accumulative effect of the regulated parameters (i.e. i_d, v_{c1}, v_{c2}, and v_{c3}) their accumulative errors are also calculated and added as state parameters. Principally, only six signal measurements and two variables are needed for this control scheme. However, the state array connected to the DRL agent has 16 elements as shown in Fig. 4. Although raw measurement signals can be likewise utilized as state space, the preprocessing unit provides more insight into the dynamics of the converter for the DRL agent and improves its efficiency and reduces the training times significantly.

D. REWARD FUNCTION

The most challenging part of a DRL application is to design a reward function. The challenge arises from the fact that the optimal operation may seem intuitive to humans with greater insight but the DRL agent may struggle to find an optimal policy through mathematical equations of its reward function. On the other hand, a poorly designed reward function can result in a policy that leads to the optimal reward by exploiting the system in ways that are not desirable or sometimes even hazardous. For instance, let's consider a DRL agent is used to control a Buck converter, and its reward function is designed to aim for minimizing loss and puts significantly high emphasis on achieving this goal. Therefore, there is a chance that the DRL agent may find out that there is a hefty penalty for power loss thus it may opt to avoid turning on its switch(es) to achieve the highest reward possible. This example demonstrates the importance of designing carefully curated reward functions as the most important step in implementing DRL for the control of various applications including power electronics.

In this paper, a reward function is designed that empowers the DRL agent to achieve four objectives efficiently while prioritizing each objective based on its importance.

There are four objectives. For $\forall i \in \{1, 2, 3, 4\}$ let R_i be an objective and $\Delta(i)$ be its corresponding error, aiming to minimize the errors between the reference values and the actual values of i_d, v_{c1}, v_{c2}, and v_{c3}, respectively. A composite reward function is designed to address these multi-objective requirements. Considering Δ_j as the error for objective j. The reward component R_j for each objective $j \in \{i_d, v_{c1}, v_{c2}, v_{c3}\}$ is defined as:

$$R_j(\Delta_j) = \begin{cases} 1 - \log_{10}\big(1 + K_1|\Delta_j|\big), & \text{if } \epsilon_{th} < |\Delta_j| < \Delta_{large} \\ 1 - K_2\sqrt{|\Delta_j|}, & \text{if } |\Delta_j| \geq \Delta_{large} \\ R_{max}, & \text{if } |\Delta_j| \leq \epsilon_{th} \end{cases} \tag{6}$$

where:

- $\Delta_j = (\text{actual}_j - \text{reference}_j)$.
- K_1, K_2 are positive scaling factors (e.g., $K_1 = 10, K_2 = 1$ used in this study).
- ϵ_{th} is a small threshold defining near-perfect tracking (e.g., 10^{-4}).
- Δ_{large} is a threshold distinguishing small/medium errors from large errors (e.g., 1).

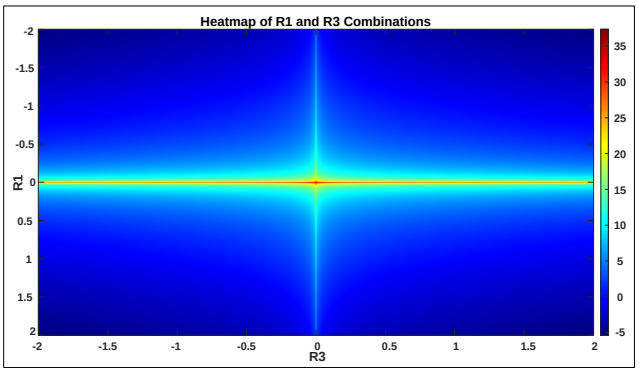

FIGURE 5. Heat map of the reward functions R_1 and R_3 representing $\Delta(i_d)$ and $\Delta(v_{c3})$

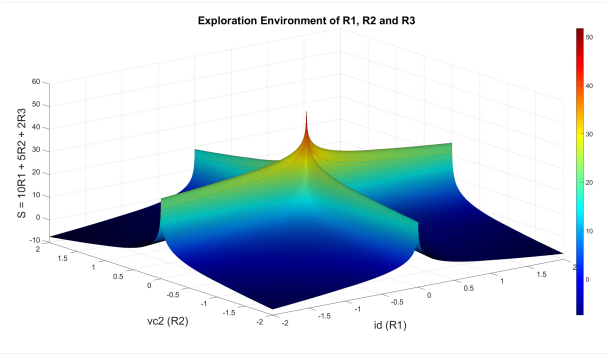

FIGURE 6. Exploration environment of a combination of the reward functions R_1, R_2, and R_3 representing $\Delta(i_d)$, $\Delta(v_{c2})$, and $\Delta(v_{c3})$

- R_{max} is the maximum reward component for achieving very low error (e.g., $1 - \log_{10}(1 + K_1\epsilon_{th})$ or a similiar fixed high value).

The logarithmic term for small to medium errors provides a steep gradient, encouraging fine-tuning. The square root term for large errors offers a less aggressive penalty than linear or quadratic, preventing conservative approach when the agent is far from the target. The constant high reward for negligible errors stabilizes training. The total instantaneous reward R_t is a weighted sum of these components:

$$R_t = w_{i_d} R_{i_d}(\Delta_{i_d}) + \sum_{k=1}^{3} w_{v_{ck}} R_{v_{ck}}(\Delta_{v_{ck}}) \tag{7}$$

where w_{i_d} and $w_{v_{ck}}$ are weighting factors reflecting the relative importance of each objective. In this paper, these coefficients were selected empirically through an iterative tuning process, prioritizing current control ($w_{i_d} = 10$) as the primary objective, followed by the voltage balancing of the capacitors with decreasing priority ($w_{v_{c1}} = 5$, $w_{v_{c2}} = 3$, and $w_{v_{c3}} = 2$). This prioritization scheme was found to yield a good balance between fast current response and stable capacitor voltage regulation during training explorations.

To demonstrate the exploration environment of the agent, a heat map for a combination of R_1 and R_3 is illustrated in Fig. 5. As seen, the agent gives higher priority to achieving the minimum error for i_d which is represented by R_1 and v_{c3}

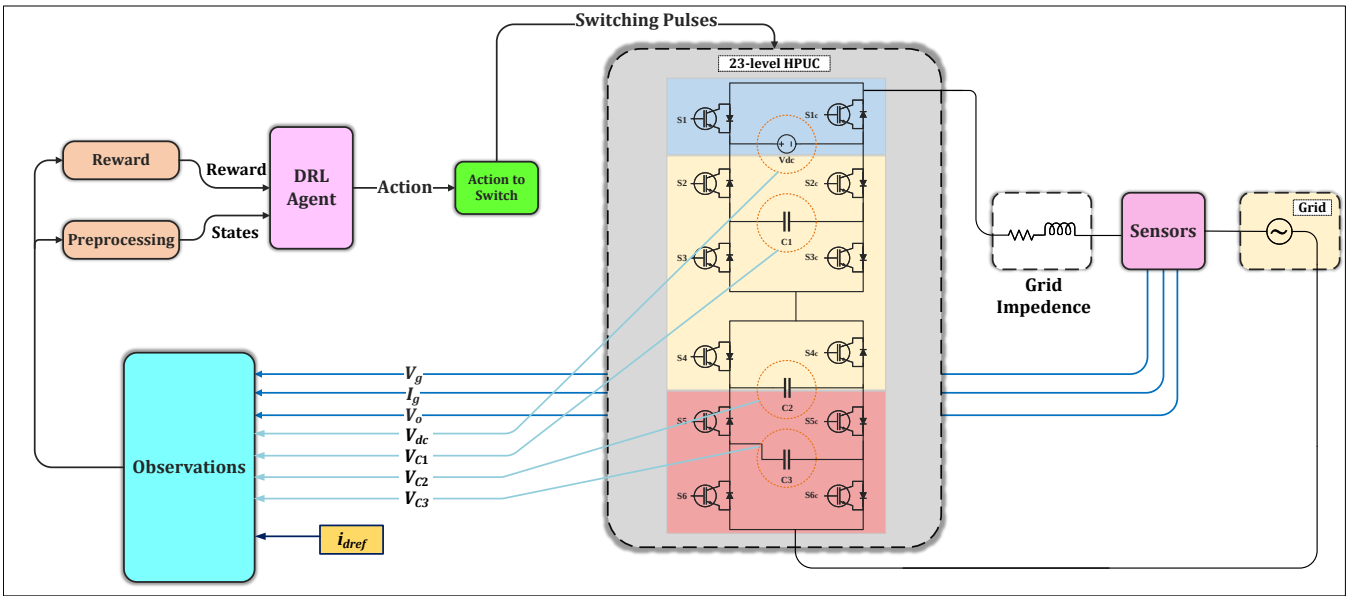

FIGURE 7. Diagram of a 23-level single DC source HPUC, controlled by the proposed DRL method

which is represented by R_3. It is evident from the heatmap that the DRL agent puts a higher priority on regulating the output current i_d in comparison to balancing the voltage of v_{c3}.

Ultimately, since depicting a four-dimensional diagram is not feasible, to visualize the reward exploration environment that the DRL agent can explore, it is illustrated as a three-dimensional map of exploration between the reward functions R_1, R_2, and R_3 as shown in Fig. 6. The figure shows that the agent explores the reward function with a low reward variation for medium and large errors resulting in a plateau of rewards when all objectives deviate too far from their reference values. Once the agent reaches areas that correspond to minimal errors, it receives higher rewards. The overall reward of "R" peaks when all objectives have near-zero errors. This configuration not only encourages the agent to stay in near-zero error areas but also allows it to fine-tune the policy to achieve the smallest error possible while respecting the priority of accomplishing each objective.

E. ACTION SPACE

In order to generate the action space, consider the topology of the HPUC converter as shown in Fig. 1. As stated earlier, the HPUC converter is comprised of 12 switches, 6 on each submodule. Since each switch and its complimentary pair have two states of On and Off. Therefore, there are $2^3 = 8$ combinations at each submodule (PUC1 and PUC2) of the converter. However, only 8 distinct voltage levels are generated by different combinations of the switching states as is listed in Table 1. Thus, the total number of permitted switching states (i.e. $8 \times 8 = 64$) is considered for creating the action space A_t. Since these switching states are incomprehensible for the DRL agent, an integer number is used to represent each combination. Thus, the discrete two-dimensional action space

TABLE 3. Simulation Parameters

Parameter	Value	Unit
Peak Grid Voltage	170	V
Grid Frequency	60	Hz
Grid Resistance (per phase)	0.1	Ω
Grid Inductance (per phase)	5	mH
DC-link Capacitors	2000	µF
Simulation Time Step (Ts)	50	µs
DC-link Voltage	160	V

in this method would be a matrix as shown in (8).

$$A_t = \begin{bmatrix} 0 & 1 & \dots & 6 & 7 \\ 0 & 1 & \dots & 6 & 7 \end{bmatrix} \tag{8}$$

A lookup table linked to the DRL agent's output is utilized to map each action, represented as an integer as shown in (8), into its corresponding switching state. This process is illustrated in Fig. 3. As seen, the "Action to Switch" block enables the mapping of actions to switching signals.

F. FINALIZING THE DRL CONTROLLER

Fig. 7. depicts the schematic of the proposed DRL controller connected to the 23-level single DC source HPUC. The actions generated by the DRL agent remain unchanged until the following sampling period of the agent, which may be different from the simulation's sampling time. Therefore, the DRL agent's sampling period can be considered as the switching frequency (f_{sw}) of the converter.

V. SIMULATION RESULTS

In order to assess the performance of the proposed approach in controlling a 23-level HPUC converter, it is implemented in the Matlab/Simulink simulation environment. Different

TABLE 4. Training Parameters

Parameter	Value
Parameter	Value
Layer Size of the State Path	[224,224]
Layer Size of the Action Path	124
Learning Rate	0.01
Gradient Threshold	1
Normalization	None
Bias Learn Rate Factor	0
Double DQN	No
Target Smooth Factor	0.001
Experience Buffer Length	10^5
Mini Batch Size	640
Discount Factor	0.9
Score Averaging Window Length	5
Epsilon Greedy Exploration Epsilon	1
Epsilon Decay	0.05
Epsilon Min	0.01
Agent Sampling Time	100µs

operation modes, dynamic changes, noise in measurements, parameter changes, and uncertainties are studied.

Table 3 and Table 4 illustrate the simulation parameters and the training parameters, respectively. By applying these values along with the previously obtained reward function, the following results are obtained in various operation conditions.

1) Steady-state operation

To evaluate the steady-state operation, the performance of the proposed DRL method is examined. Fig. 8. illustrates the steady-state operation of the HPUC converter using DRL. The system demonstrates satisfactory steady-state performance. The steady-state error for the output current i_d tracking is maintained within $\pm 0.2A$ of the reference. The capacitor voltage balancing is also effective, with voltage errors for V_{C1}, V_{C2}, and V_{C3} kept within $\pm 5V$ of their respective setpoints (e.g., V_{C1} around $V_{DC}/2 = 80V$). The system exhibits a relatively small rise time, reaching 90% of the target current in approximately 30 ms from startup. The Total Harmonic Distortion (THD) of the output current is measured as 2.71% ensuring that THD remains within the acceptable range of standards.

2) Dynamic response

Additionally, in order to assess the dynamic response of the DRL method, step changes in the active and reactive power delivered to the grid using the HPUC converter are studied. As shown in Fig. 9, a step change from $i_d = 10$A, $i_q = 0$A to $i_d = 8$A, $i_q = 0$A at $t = 200$ms and another one to $i_d = 10$A, $i_q = 0.4$A at $t = 300$ms are applied to the HPUC converter controlled by the DRL agent. As observed in Fig. 9, the DRL agent effectively tracks the step changes in i_d and i_q. For the step change in i_d from 10A to 8A at t=200ms, the output current settles to the new reference within approximately 40 ms with minimal overshoot. A similar responsive behavior is observed for the subsequent change at t=300ms, demonstrating the agent's ability to quickly adapt to varying power demands while maintaining capacitor voltage stability. The results demonstrated good performance when facing these dynamic changes.

3) Noise, parameter variations, and uncertainty

The main objectives of utilizing DRL for the control of power electronic converters, such as HPUC, are its model-free approach, and the generalization characteristics of this control method. In the previous sub-sections, the satisfactory performance of the DRL agent without prior knowledge of the topology under steady-state as well as the dynamic changes of the aforementioned converter are demonstrated. In this section, the converter is utilized under various non-regular operational points to assess the resilience of the DRL method under uncertain conditions.

The HPUC converter is set to deliver a current of $i_d = 6$A to the grid using the DRL agent as shown in Fig. 10. At $t = 100$ms, an active power increase of $i_d = 10$A is administrated. As seen, the DRL agent followed this dynamic change closely. Moreover, a reactive power increase from $i_q = 0$A to $i_q = 0.2$A is added at $t = 200$ms, and the agent displayed a favorable response to this step change.

Moreover, a white Gaussian noise (WGN) is added to the output current (i_g) signal. The magnitude of the added WGN is calculated using (V-3) and (V-3) to achieve a signal-to-noise (SNR) ratio of 25dB. Where P is the power and A is the amplitude of the signal or the added noise.

$$SNR = \frac{P_{\text{signal}}}{P_{\text{noise}}} = \left(\frac{A_{\text{signal}}}{A_{\text{noise}}}\right)^2 \tag{9}$$

$$SNR_{\text{dB}} = 10\ \log_{10}\left(\frac{P_{\text{signal}}}{P_{\text{noise}}}\right) \tag{10}$$

The aforementioned noise is added at $t = 300$ms to the HPUC converter when it is controlled by the DRL agent as shown in Fig. 10. As depicted, the introduction of WGN (SNR = 25dB) at $t = 300$ms causes a slight increase in the output current ripple, but the THD remains below 5%, and the DRL method maintains stable control. This verifies the robust performance and superior capability of the DRL method in resisting performance degradation in noisy environments.

Additionally, to evaluate the performance of the DRL method when facing parameter variations, a degradation manifested as a reduction of 15% in the capacity of all three DC link capacitors of the HPUC converter is administrated at $t = 400$ms. As shown in Fig. 10, following the 15% reduction in capacitance at $t = 400$ms, the peak-to-peak voltage ripple on the DC-link capacitors increased by approximately $\pm 20V$ initially, but the DRL agent adapted, and the system remained stable, with the output voltage distortion being minor. These ripples become less pronounced in a short period. The output voltage becomes slightly distorted as well.

At $t = 500$ms, a 5% increase in the grid voltage is applied to the HPUC. As seen in Fig. 10, this change results in slightly higher voltage ripples in the capacitors.

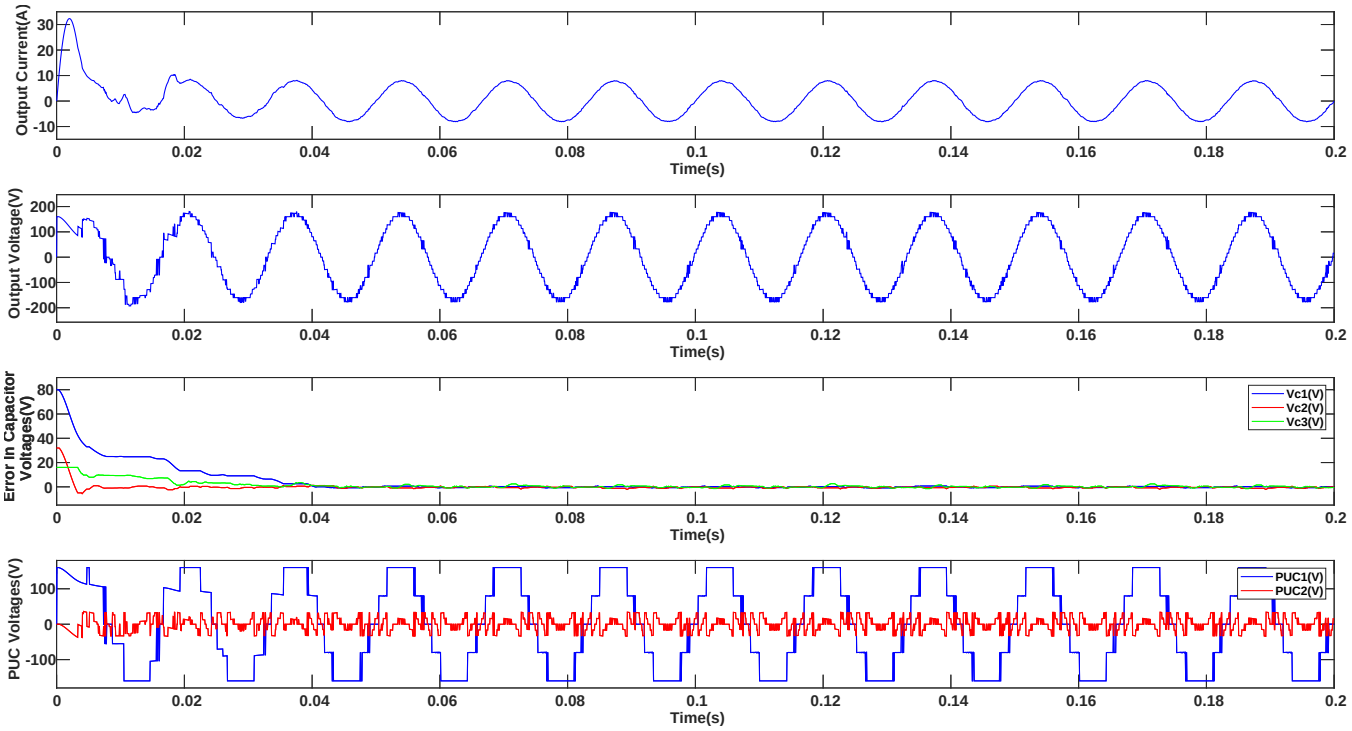

FIGURE 8. **Waveforms of the output current(A), output voltage(V), errors of the DC-link capacitors(V) and voltages across submodules PUC1 and PUC2 in steady state operation controlled by the DRL agent**

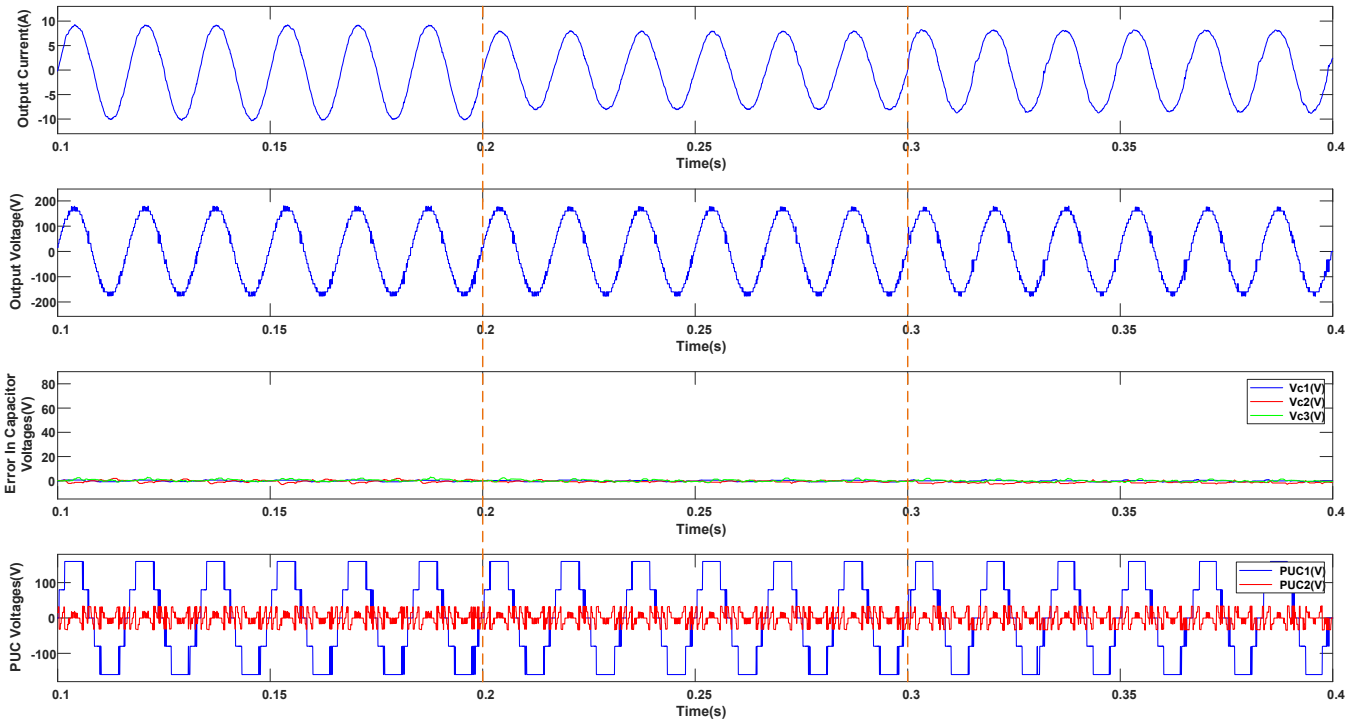

FIGURE 9. **Waveforms of the output current(A), output voltage(V), errors of the DC-link capacitors(V) and voltages across submodules PUC1 and PUC2 controlled by the DRL agent facing step changes from $i_d = 10$A; $i_q = 0$A to $i_d = 8$A; $i_q = 0$A at $t = 200$ms (indicated by dashed line) and $i_d = 10$A; $i_q = 0.4$A at $t = 300$ms (indicated by dashed line)**

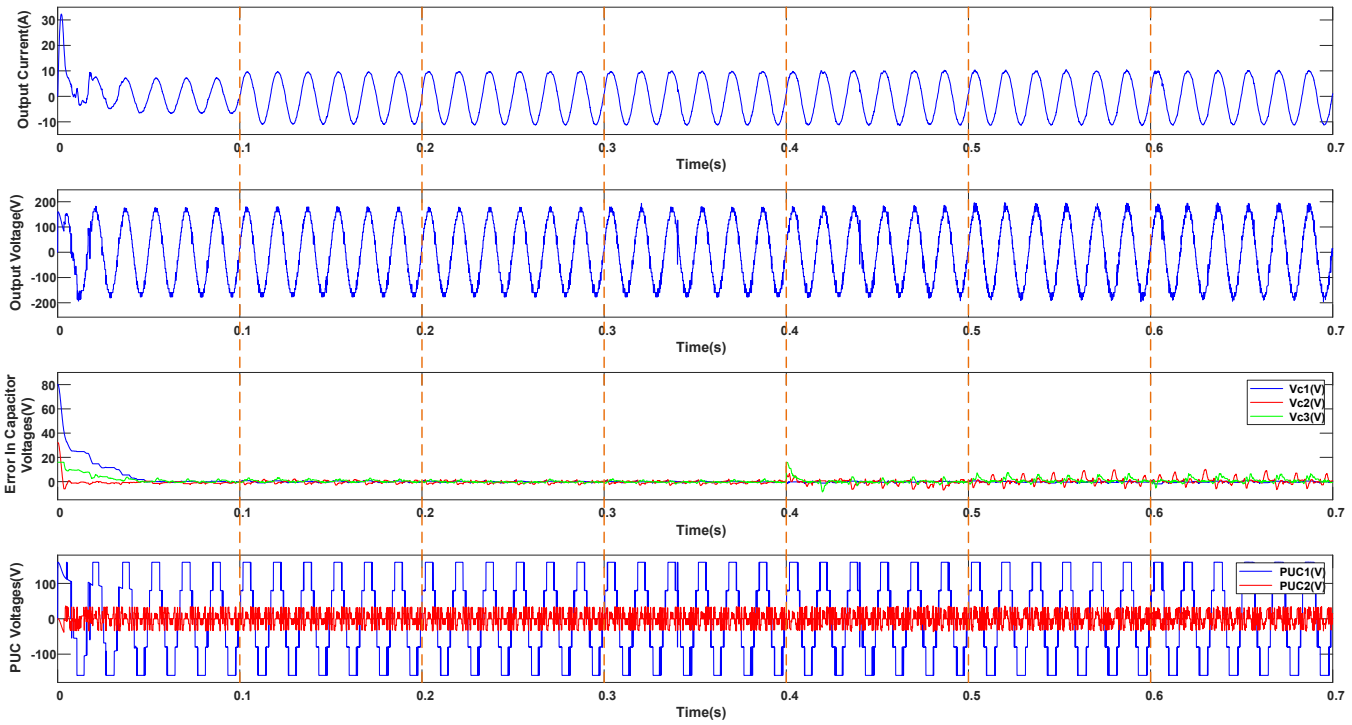

FIGURE 10. **Waveforms of the output current(A), output voltage(V), errors of the DC-link capacitors(V) and voltages across submodules PUC1 and PUC2 controlled by the DRL agent under accumulate effect of active power changes at $t = 100$ms, reactive power changes at $t = 200$ms, noise in measurements at $t = 300$ms, capacitor degradation at $t = 400$ms, grid voltage changes at $t = 500$ms and line impedence changes at $t = 600$ms**

On top of all the previously mentioned changes, an increase of 28% is applied to the line impedance at $t = 600$ms. The cumulative effect of all these changes imposes severe challenges to traditional control methods. However, the DRL method only exhibits a moderate distortion when all the previous conditions impact its performance at $t = 600$ms. The cumulative effect at $t = 600$ms shows a moderate increase in output voltage distortion and capacitor voltage ripple (e.g., peak errors temporarily reaching $\pm 15V$), but the system does not lose stability, underscoring the DRL agent's robustness.

Considering the results obtained in this section, it can be concluded that the DRL agent, despite requiring prior training with relatively long training times, not only can obtain the optimal switching and control policy of the HPUC converter in a model-free approach but also demonstrates relatively high resilience towards uncertainties, parameter variations, noise, and dynamic responses.

VI. CONCLUSION

Based on the results obtained in this paper, it is shown that the implementation of Deep Reinforcement Learning (DRL) for controlling the 23-level Single DC source Hybrid Packed U-Cell (HPUC) converter offers significant advantages over traditional methods that face challenges such as requiring an accurate model of the converter as well as sensitivity to parameter variations, noise, and other uncertainties, leading to performance degradation or even instability. In contrast, DRL provides a modern approach to nonlinear control of power electronics. Without a prior model of the converter and only by training its neural network layers and acquiring optimal control policy through interactions with the converter, DRL demonstrates higher resilience towards uncertainties, parameter variations, noise, and dynamic changes. The model-free nature of DRL, powered by its interactive training procedure and the implementation of deep neural networks improve their adaptability and effectiveness in addressing the control challenges of complex power electronics converters such as HPUC. Simulations in MATLAB/Simulink confirmed the satisfactory performance of the proposed method, achieving a low THD of 2.71% for the output current, demonstrating effective capacitor voltage balancing with errors kept within $\pm 5V$ under nominal conditions, and showcasing robust dynamic responses (e.g., settling times around 40 ms for significant load changes). The DRL agent's resilience was particularly evident when subjected to a 25dB SNR noise environment and 15% capacitor degradation, where it maintained operational stability and acceptable performance. Therefore, a promising approach is introduced that empowers further studies on advanced and intelligent control of complex power electronics converters. Future work will focus on experimental validation of the proposed DRL controller to further assess its real-world performance and practical implementation challenges.

REFERENCES

[1] G. Zhang, Z. Li, B. Zhang, and W. A. Halang, "Power electronics converters: Past, present and future," vol. 81, pp. 2028–2044.
[2] P. Qashqai, A. Sheikholeslami, H. Vahedi, and K. Al-Haddad, "A Review

on Multilevel Converter Topologies for Electric Transportation Applications," in *2015 IEEE Vehicle Power and Propulsion Conference (VPPC)*, pp. 1–6.

[3] S. Impram, S. V. Nese, and B. Oral, "Challenges of renewable energy penetration on power system flexibility: A survey," vol. 31, p. 100539.

[4] R. A. Krishna and L. P. Suresh, "A brief review on multi level inverter topologies," in *2016 International Conference on Circuit, Power and Computing Technologies (ICCPCT)*. IEEE, pp. 1–6.

[5] K.-R. Sorto-Ventura, M. Abarzadeh, K. Al-Haddad, and L. A. Dessaint, "23-level single DC source hybrid PUC (H-PUC) converter topology with reduced number of components: Real-time implementation with model predictive control," vol. 1, pp. 127–137.

[6] H. Vahedi, H. Y. Kanaan, and K. Al-Haddad, "PUC converter review: Topology, control and applications," in *IECON 2015-41st Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 004 334–004 339.

[7] H. Vahedi, P.-A. Labbé, and K. Al-Haddad, "Sensor-less five-level packed U-cell (PUC5) inverter operating in stand-alone and grid-connected modes," vol. 12, no. 1, pp. 361–370.

[8] S. Bacha, I. Munteanu, and A. I. Bratcu, "Power electronic converters modeling and control," vol. 454, p. 454.

[9] M. A. Mahmud, T. K. Roy, S. Saha, M. E. Haque, and H. R. Pota, "Robust Nonlinear Adaptive Feedback Linearizing Decentralized Controller Design for Islanded DC Microgrids," vol. 55, no. 5, pp. 5343–5352.

[10] H. Komurcugil, S. Biricik, S. Bayhan, and Z. Zhang, "Sliding mode control: Overview of its applications in power converters," vol. 15, no. 1, pp. 40–49.

[11] M. Wang and Y. Wang, "Fuzzy Neural-Network-based Output Tracking Control for Nonlinear Systems with Unknown Dynamics," in *2020 Chinese Automation Congress (CAC)*, pp. 5124–5129.

[12] J. V., "GENETIC ALGORITHM BASED SOLUTION IN PWM CONVERTER SWITCHING FOR VOLTAGE SOURCE INVERTER FEEDING AN INDUCTION MOTOR DRIVE," vol. 27, no. 2, pp. 45–60.

[13] D. Mishra and S. Mandal, "Voltage regulation of dc-dc boost converter using h-infinity controller," in *2020 International Symposium on Devices, Circuits and Systems (ISDCS)*, 2020, pp. 1–5.

[14] H. Liu and S. Li, "Speed control for PMSM servo system using predictive functional control and extended state observer," vol. 59, no. 2, pp. 1171–1183.

[15] D. Çelik, H. Ahmed, and M. E. Meral, "Kalman filter-based super-twisting sliding mode control of shunt active power filter for electric vehicle charging station applications," vol. 38, no. 2, pp. 1097–1107.

[16] M. Babaie, M. Sharifzadeh, M. Mehrasa, G. Chouinard, and K. Al-Haddad, "Supervised learning model predictive control trained by ABC algorithm for common-mode voltage suppression in NPC inverter," vol. 9, no. 3, pp. 3446–3456.

[17] J. Rodriguez, C. Garcia, A. Mora, S. A. Davari, J. Rodas, D. F. Valencia, M. Elmorshedy, F. Wang, K. Zuo, and L. Tarisciotti, "Latest advances of model predictive control in electrical drives—Part II: Applications and benchmarking with classical control methods," vol. 37, no. 5, pp. 5047–5061.

[18] J. Rodriguez, C. Garcia, A. Mora, F. Flores-Bahamonde, P. Acuna, M. Novak, Y. Zhang, L. Tarisciotti, S. A. Davari, and Z. Zhang, "Latest advances of model predictive control in electrical drives—Part I: Basic concepts and advanced strategies," vol. 37, no. 4, pp. 3927–3942.

[19] P. Qashqai, H. Vahedi, and K. Al-Haddad, "Applications of artifical intelligence in power electronics," in *2019 IEEE 28th International Symposium on Industrial Electronics (ISIE)*. IEEE, pp. 764–769.

[20] P. Qashqai, K. Al-Haddad, and R. Zgheib, "Deep neural network-based black-box modeling of power electronic converters using transfer learning," in *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*. IEEE, pp. 1–6.

[21] ——, "Modeling Power Electronic Converters Using A Method Based on Long-Short Term Memory (LSTM) Networks," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pp. 4697–4702.

[22] P. Qashqai, R. Zgheib, and K. Al-Haddad, "GRU and LSTM comparison for black-box modeling of power electronic converters," in *IECON 2021–47th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 1–5.

[23] ——, "A Programmatical Method for Real-time Simulation of Black-box LSTM-based Models of Power Electronic Converters in Hypersim," in *2022 IEEE 1st Industrial Electronics Society Annual On-Line Conference (ONCON)*. IEEE, pp. 1–5.

[24] M. Babaie, M. Sharifzadeh, and K. Al-Haddad, "Adaptive ANN based Single PI Controller for Nine-Level PUC Inverter."

[25] M. Babaie, M. Saeidi, M. Sharifzadeh, A. Hamadi, K. Al-Haddad, and A. Chandra, "Hybrid ANN-Linear controller for maximum PV energy harvesting in grid-tied packed e-cell inverter," in *2020 International Symposium on Power Electronics, Electrical Drives, Automation and Motion (SPEEDAM)*. IEEE, pp. 871–875.

[26] V. Kůrková, Y. Manolopoulos, B. Hammer, L. Iliadis, and I. Maglogiannis, Eds., *Artificial Neural Networks and Machine Learning – ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III*, ser. Lecture Notes in Computer Science. Springer International Publishing, vol. 11141. [Online]. Available: http://link.springer.com/10.1007/978-3-030-01424-7

[27] Y. Li. Deep Reinforcement Learning: An Overview. [Online]. Available: http://arxiv.org/abs/1701.07274

[28] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," vol. 521, no. 7553, pp. 436–444.

[29] H. Nguyen and H. La, "Review of deep reinforcement learning for robot manipulation," in *2019 Third IEEE International Conference on Robotic Computing (IRC)*. IEEE, pp. 590–595.

[30] W. Zhao, J. P. Queralta, and T. Westerlund, "Sim-to-real transfer in deep reinforcement learning for robotics: A survey," in *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*. IEEE, pp. 737–744.

[31] K. Shao, Z. Tang, Y. Zhu, N. Li, and D. Zhao. A Survey of Deep Reinforcement Learning in Video Games. [Online]. Available: http://arxiv.org/abs/1912.10944

[32] Y. Wan, J. Qin, Q. Ma, W. Fu, and S. Wang, "Multi-agent DRL-based data-driven approach for PEVs charging/discharging scheduling in smart grid," vol. 359, no. 4, pp. 1747–1767.

[33] Y. Gao and N. Yu, "Deep reinforcement learning in power distribution systems: Overview, challenges, and opportunities," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*. IEEE, pp. 1–5.

[34] S. Lee and D.-H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," vol. 18, no. 1, pp. 488–497.

[35] A. Sheikhi, M. Rayati, and A. M. Ranjbar, "Dynamic load management for a residential customer; reinforcement learning approach," vol. 24, pp. 42–51.

[36] T. Yang, L. Zhao, W. Li, and A. Y. Zomaya, "Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning," vol. 235, p. 121377.

[37] Y. Li, C. Yu, M. Shahidehpour, T. Yang, Z. Zeng, and T. Chai, "Deep Reinforcement Learning for Smart Grid Operations: Algorithms, Applications, and Prospects."

[38] D. Cao, W. Hu, J. Zhao, G. Zhang, B. Zhang, Z. Liu, Z. Chen, and F. Blaabjerg, "Reinforcement learning and its applications in modern power and energy systems: A review," vol. 8, no. 6, pp. 1029–1042.

[39] M. Hajihosseini, M. Andalibi, M. Gheisarnejad, H. Farsizadeh, and M.-H. Khooban, "DC/DC Power Converter Control-Based Deep Machine Learning Techniques: Real-Time Implementation," vol. 35, no. 10, pp. 9971–9977.

[40] T. Yang, C. Cui, and C. Zhang, "On the Robustness Enhancement of DRL Controller for DC-DC Converters in Practical Applications," in *2022 IEEE 17th International Conference on Control & Automation (ICCA)*, pp. 225–230.

[41] C. Cui, N. Yan, B. Huangfu, T. Yang, and C. Zhang, "Voltage Regulation of DC-DC Buck Converters Feeding CPLs via Deep Reinforcement Learning," vol. 69, no. 3, pp. 1777–1781.

[42] D. Alfred, D. Czarkowski, and J. Teng, "Model-Free Reinforcement-Learning-Based Control Methodology for Power Electronic Converters," in *2021 IEEE Green Technologies Conference (GreenTech)*, pp. 81–88.

[43] M. Gheisarnejad, H. Farsizadeh, and M. H. Khooban, "A Novel Nonlinear Deep Reinforcement Learning Controller for DC–DC Power Buck Converters," vol. 68, no. 8, pp. 6849–6858.

[44] J. Wang, R. Yang, and Z. Yao, "Efficiency Optimization Design of Three-Level Active Neutral Point Clamped Inverter Based on Deep Reinforcement Learning," in *2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2)*, pp. 605–610.

[45] P. Qashqai, K. Al-Haddad, and R. Zgheib, "A new model-free space vector modulation technique for multilevel inverters based on deep reinforcement learning," in *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, pp. 2407–2411.

[46] P. Qashqai, M. Babaie, R. Zgheib, and K. Al-Haddad, "A Model-Free Switching and Control Method for Three-Level Neutral Point Clamped Converter Using Deep Reinforcement Learning," vol. 11, pp. 105 394–105 409.
[47] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," vol. 7, pp. 133 653–133 667.
[48] Reinforcement Learning Agents - MATLAB & Simulink. [Online]. Available: https://www.mathworks.com/help/reinforcement-learning/ug/create-agents-for-reinforcement-learning.html

POURIA QASHQAI (MEMBER, IEEE) received the B.Sc. degree in electrical engineering from the University of Isfahan, Isfahan, Iran, in 2013, the M.Sc. degree in power electronics from Babol Noshirvani University of Technology, Babol, Iran, in 2015, and the Ph.D. degree in electrical engineering from the École de Technologie Supérieure (ÉTS), Montreal, QC, Canada, in 2025. During his Ph.D., his research was part of a collaborative project with the Hydro-Québec Research Institute (IREQ) and OPAL-RT Technologies. He has also worked as a Power System Simulation Research Intern at IREQ and as an Electrical Modeling & Simulation Specialist at OPAL-RT Technologies. He is currently a Power System Specialist with Eaton - CYME International T&D, Brossard, QC, Canada. As a doctoral student, he was a member of the Groupe de Recherche en Électronique de Puissance et Commande Industrielle (GRÉPCI). His research interests include AI applications in power electronics and power systems, such as deep reinforcement learning and recurrent neural networks, real-time simulation and co-simulation for power grids, and the modeling and control of power converters.

MOHAMMAD BABAIE (MEMBER, IEEE) received the Ph.D. degree in electrical engineering from École de technologie supérieure (ÉTS), Montreal, Canada, where he was recognized on the university honor list for outstanding academic and research achievements. He is currently a Power Electronics Design Engineer at Socomec and previously worked at OPAL-RT Technologies and ÉTS as a Research Associate. His expertise includes the design and real-time control of power electronic converters for electric vehicle (EV) chargers, grid-tied systems, and microgrids. His research interests span model predictive control, intelligent and adaptive control algorithms, multilevel inverter topologies, and power hardware-in-the-loop (PHIL) experimentation. He has developed FPGA- and MCU-based embedded control platforms, designed experimental testbeds for microgrid and converter validation, and contributed to patented technologies in the field. Dr. Babaie has authored numerous IEEE journal and conference papers and serves as an active reviewer for IEEE Transactions and flagship conferences in power electronics and industrial applications.

RAWAD ZGHEIB (MEMBER, IEEE) received the dual master's degree from Paris, France, and Beirut, Lebanon, and the Ph.D. degree in electrical engineering from Montreal, Canada. He is currently a Research Project Manager with G&W Electric Company involved in developing and managing proposals for government-sponsored projects and leading the execution of approved initiatives. His projects provide opportunities to develop cutting-edge technologies and will play a pivotal role in advancing innovation strategies. He brings more than six years of expertise in research and development with a strong focus on advancing the energy transition for electrical utilities. He has successfully led innovative research projects aimed at the digitalization of electrical substations and the simulation of extreme scenarios to strengthen grid resilience-critical areas for the future of energy systems.

KAMAL AL-HADDAD (LIFE FELLOW, IEEE) received the B.Sc.A. and M.Sc.A. degrees from the University of Quebec à Trois-Rivières, Trois-Rivières, QC, Canada, in 1982 and 1984, respectively, and the Ph.D. degree from the Institute National Polytechnique, Toulouse, France, in 1988. Since June 1990, he has been a Professor with the Electrical Engineering Department, École de Technologie Supérieure, Montreal, QC, Canada, where he has been the Senior Canada Research Chair of Electric Energy Conversion and Power Electronics since 2002. He is a consultant and has established a very solid link with many Canadian and international industries working in the field of power electronics, electric transportation, aeronautics, and telecommunications. He successfully transferred and implemented twenty-four technologies to Canadian and international companies. His research interests include highly efficient static power converters, harmonics and reactive power control using hybrid filters, voltage-level multiplier, resonant and multilevel converters including the modeling, control, and development of prototypes for various industrial applications in electric traction, renewable energy, sinewave power supplies for drives, BESS and renewable energy devices. Prof. Al- Haddad is a Member of the Academy of Sciences, Fellow of the Royal Society of Canada, and a Fellow Member of the Canadian Academy of Engineering. He was the IEEE IES President during 2016–2017. He is an Associate Editor for the IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS and IES Distinguished Lecturer. He was the recipient of the 2014 IEEE IES Dr.-Ing. Eugene Mittelmann Achievement Award, and the IEEE 2023-2024 division VI director. Professor Al-Haddad is the recipient of the 2023 Medal in Power Engineering and the 2024 Québec Industrial Association Research and development ADRIQ Innovation price.