Reinforcement Learning based Control for Non-Isolated DC-DC Converters

Kaushik K Gajula
Member, IEEE

***Abstract*—In this technical report, we will review the application of Reinforcement Learning (RL) in electrical systems and apply the algorithm to control the non-isolated dc-dc converters i.e., Buck, Boost and Buck-Boost. The RL algorithm takes action by updating the action value function, also called the Q-function. The reward function is calibrated to choose the best possible action on the basis of its effort in minimizing the error state i.e., the difference between measurement and reference. Mathematical derivations will be supported by simulation results to prove the application of this theory.**

***Index Terms*—Power converters and systems, Non-isolated dc-dc converters, Reinforcement learning, Q-Learning.**

I. Introduction

MODERN electrical grids are undergoing a fundamental transformation: increasing penetration of renewable energy sources, distributed energy resources (DER), and electric vehicles are creating unprecedented levels of complexity, uncertainty, and variability in both generation and demand [1], [2]. Conventional control paradigms—model predictive control (MPC), PID controllers, and optimization-based dispatch solvers—while mature and well-understood, are increasingly challenged by this dynamism [3]. Reinforcement learning (RL), a branch of machine learning wherein an agent learns optimal control policies through interaction with an environment, has emerged as a highly promising methodology for addressing these challenges [4], [5].

Voltage regulation is one of the most studied RL applications in power systems. Applied Deep Q-Network (DQN) to transmission voltage control, proposing open-source environments and demonstrating scalability up to 500-bus systems is presented in [6]. Evaluation of Deep Reinforcement Learning (DRL)-based voltage control on real operational contexts (IEEE 14-bus, Illinois 200-bus, ISO New England) and identified key performance bottlenecks for practical deployment is presented in [7].

Graph-based methods have improved RL's ability to capture network topology. In [8] a Graph Convolutional Network–Soft Actor-Critic (GCN-SAC) architecture for autonomous voltage control, addressing challenges in leveraging topological features and computational efficiency is demonstrated. A comprehensive survey by [9] covers value-based, policy-based, model-assisted, meta-learning, multi-agent, and hierarchical RL approaches for voltage control, spanning studies from 2017–2025.

For distribution networks with high renewable penetration, deep RL for volt-var control (VVC) has attracted particular attention. DRL methods can handle the unpredictability of solar and wind, reducing power losses and voltage violations in active distribution networks (ADNs) [10].

DC-DC converters—buck, boost, buck-boost, and more exotic topologies—are the workhorses of power electronics, appearing in electric vehicle EV power-trains, battery chargers, photo voltaic PV systems, and DC microgrids. Their control is traditionally based on small-signal linearization and PID or sliding-mode controllers, which can fail to self-adapt under large disturbances, mode transitions (continuous vs. discontinuous conduction mode), or parameter uncertainty.

DRL has emerged as a compelling model-free alternative. A Proximal Policy Optimization (PPO) based DRL controller was demonstrated for DC-DC buck converters operating in both continuous conduction mode (CCM) and discontinuous conduction mode (DCM) with resistive and inductive loads; key design innovations included a chattering-reduction reward function and neural network architecture optimization, improving voltage regulation and output smoothness over traditional PWM-based schemes [11]. A comparative evaluation of PPO against traditional PID and sliding-mode controllers showed that the RL-based approach achieves shorter settling times and better stability during step-response transients [12].

A real-time DRL approach for buck converter control explicitly addresses the controller time-delay problem inherent in digital signal processors, constructing an augmented virtual decision process to eliminate the delay effect and achieving stable output voltage regulation with fast transient response via direct gate control—without a traditional PWM modulator [13]. Direct switch-level control using RL, bypassing the small-signal model entirely, was demonstrated for voltage regulation of DC-DC converters in [14].

A comprehensive review of DRL for power converter control (2025) tabulates applications spanning DC-DC, AC-DC, DC-AC, and bidirectional topologies, covering adaptive predictive horizon control with Twin Delayed Deep Deterministic Policy Gradient (TD3), coordinated voltage control of fuel cell systems using multi-agent Deep Deterministic Policy Gradient (DDPG)/ Multi-Agent Deep Deterministic Policy Gradient (MADDPG), and topology-adaptive control of buck converters in DC microgrids [15].

DC microgrids aggregate DERs (PV, batteries, fuel cells), DC loads, and DC-DC converters on a common DC bus. Key control objectives include DC bus voltage regulation, proportional current sharing among distributed generators (DGs), and economic energy dispatch—often simultaneously. Constant power loads (CPLs), which exhibit negative incremental impedance, pose a particular stability threat that conventional

linear controllers handle poorly.

A DRL framework based on the TD3 algorithm was proposed specifically for voltage regulation of DC-DC buck converters feeding CPLs in DC microgrids, demonstrating superior robustness and transient performance over conventional controllers under varying load conditions [16].

For hybrid energy storage systems (HESS) in DC microgrids—combining batteries, super-capacitors, and hydrogen storage—a two-layer DRL control strategy was proposed for optimal power-sharing considering the output response characteristics of each storage element, addressing both power balance and storage longevity [17].

In this report, the following notation will be followed:

- If function $\mathcal{F}$ and its parameters a, b are all at iteration k then,

$$\mathcal{F}_{a,b}[k]$$

- If function $\mathcal{F}$ is at iteration $k+1$ and its parameters a, b are in iteration k then,

$$\mathcal{F}_{a[k],b[k]}[k-1]$$

II. Control by Reinforcement Learning

This section provides a review on RL algorithm. Furthermore, application of the same on dc-dc converters will be presented.

A. Reinforcement Learning

In RL, the objective of an agent (learner/decision-maker) is to maximize/optimize the policy π, given by

$$\pi(a|s) = \mathbb{P}(A_t = a|S_t = s). \tag{1}$$

Equation (1) states that a policy is defined as the probability for an agent to take an action a when it is in state s. An optimal policy is derived from the function $Q_{s[k],a}$, which is the maximum possible action value that can be yielded over a long-term from the state $s[k]$ while introducing an action a. The total expected long-term reward starting from state s at iteration k while making perfect decisions, is denoted by $V^*_{s[k]}$.

$$V^*_s[k] = \max_{a \in A_t} Q^*_{s[k],a}. \tag{2}$$

Based on the action value function, the optimal policy generated for the future iterations is given by,

$$\pi^*_s[k+1] = \arg\max_a Q_{s,a}[k]. \tag{3}$$

A Q-Learning based RL strategy where the function $Q_{s,a}$ is updated over each iteration k is given by:

$$\begin{aligned} Q_{s,a}[k+1] = Q_{s,a}[k] + \alpha\Big(r[k] + \cdots \\ \gamma \max_a (Q_{s[k+1],a}) - Q_{s,a}[k]\Big). \end{aligned} \tag{4}$$

Using the state $s[k+1]$, $Q_{s,a}$ picks the best action a and consequentially optimizes the policy over the reward function at each iteration k. Furthermore, in (4) the reward at iteration k denoted by $r[k]$. The learning rate, discount factor are represented by α and γ respectively. Such that,

$$\alpha, \gamma \in (0, 1]. \tag{5}$$

By picking α's value closer to 0, it would mean that the agent learns slowly and relies more on past experiences. Conversely, a value closer to 1 would mean, the agent learns quickly, adapting rapidly to new rewards. Similarly, for γ if the value is closer to 0 then the agent only cares about immediate/short-term rewards r. Otherwise, a value closer to 1 would make the agent place higher value on long-term goals.

In contrast to the off-policy approach of Q-Learning, the State-Action-Reward-State-Action (SARSA) operates on-policy making it an easier choice to implement on the dc-dc converters. The previous equation (Q-Learning) in (4) is then written as,

$$\begin{aligned} Q_{s,a}[k+1] = Q_{s,a}[k] + \alpha\Big(r[k] + \cdots \\ \gamma Q_{s[k+1],a[k+1]}[k] - Q_{s,a}[k]\Big). \end{aligned} \tag{6}$$

In (6), once again the Q-function/action value function is updated over each iteration. However, this time it is by using the information from the next step/iteration and not maximizing.

B. Reinforcement Learning for dc-dc converters

To apply RL algorithm for the dc-dc converters we consider the states as:

$$\begin{pmatrix} s_v \\ s_i \end{pmatrix} = \begin{pmatrix} v - v^{ref} \\ i - i^{ref} \end{pmatrix}, \tag{7}$$

which is based on the error function. The capacitor voltage and the inductor current measurements are denoted by v and i respectively. The reward function is given by,

$$\begin{pmatrix} r_v[k] \\ r_i[k] \end{pmatrix} = \begin{pmatrix} -\Big(v[k] - v^{ref}[k]\Big)^2 - \Big(i[k] - i^{ref}[k]\Big)^2 \\ -\Big(i[k] - i^{ref}[k]\Big)^2 - \Big(d[k]\Big)^2 \end{pmatrix}. \tag{8}$$

Based on the action a the Q function is then taken as follows:

$$\begin{pmatrix} Q_{s_v,i^{ref}}[k+1] \\ Q_{s_i,d}[k+1] \end{pmatrix} = \begin{pmatrix} r_v[k+2] + \gamma r_v[k+3] \\ r_i[k+2] + \gamma r_i[k+3] \end{pmatrix}. \tag{9}$$

The action value function, Q's relation to the reward function r, is given by

$$\begin{aligned} Q_{s,a}[k] = \mathbb{E}_\pi\Big[r[k+1] + \gamma r[k+2] \cdots \\ + \gamma^2 r[k+3] + \cdots \Big| s[k], a[k]\Big]. \end{aligned} \tag{10}$$

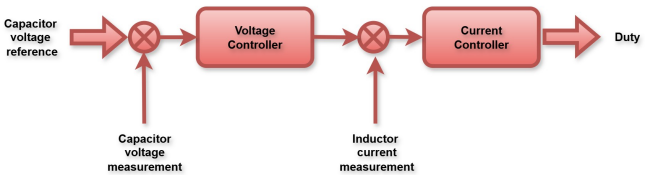
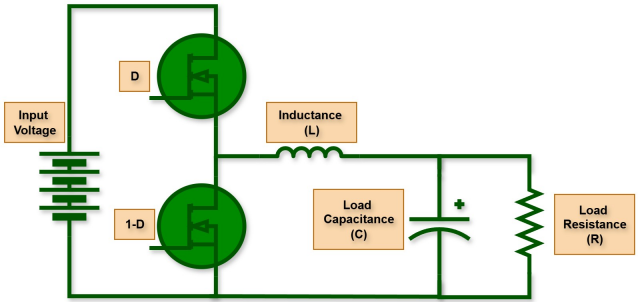


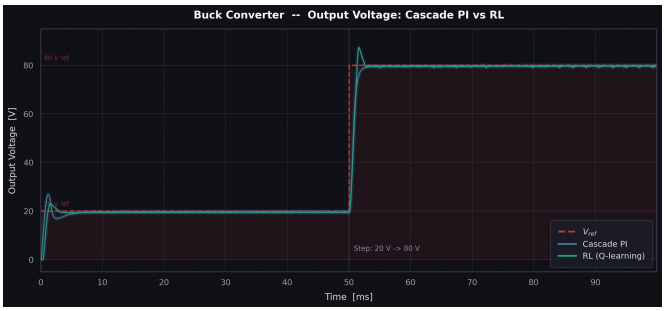
Fig. 1. Cascade PI Controller.

III. Simulation Results

In this section, the RL strategy is implemented and compared to a typical cascaded PI control strategy for non-isolated dc-dc converters. The architecture of the cascaded PI controller including the voltage control outer loop (slower) and the current control inner loop (faster) is presented in Fig.1. In Figs. 2, 3 and 4 the topologies of the dc-dc non-isolated power converters and their respective voltage control results are presented.



(a) Topology of Buck Converter.



(b) Change in output voltage reference from 20 V to 80 V.

Fig. 2. Case 1: Buck Converter.

A. Case 1: DC-DC Buck Converter

The buck converter topology together with the controller output to the switch is presented in Fig. 2a. The voltage control comparison between the cascaded PI control strategy and the RL strategy is presented in Fig. 2b. As seen in the figure, the voltage is risen from an initial value of 20 V to 80 V at the load of the buck converter while the input voltage was set at 100 V. RL shows good steady state and transient response to the change introduced at t = 0.05 s and at initialization which is fairly similar to the control performance by the cascaded PI controller. For a non-isolated dc-dc converter, state vector is given by:

$$x(t) = \begin{bmatrix} i(t) \\ v(t) \end{bmatrix} \tag{11}$$

where:

- i = Inductor current
- v = Capacitor (output) voltage

The input is v_{in} (source voltage), and the output is the capacitor voltage, v.

For ON-state, use the switch denoted by D:

$$\frac{di}{dt} = \frac{v_{in} - v}{L}, \quad \frac{dv}{dt} = \frac{i - \frac{v}{R}}{C} \tag{12}$$

For OFF-state use the switch denoted by $1 - D$:

$$\frac{di}{dt} = \frac{-v}{L}, \quad \frac{dv}{dt} = \frac{i - \frac{v}{R}}{C} \tag{13}$$

In matrix form:

ON-state (switch denoted by D):

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix}}_{A_{\text{on}}} x + \underbrace{\begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix}}_{B_{\text{on}}} v_{in} \tag{14}$$

OFF-state (switch denoted by $1 - D$):

$$\dot{x} = \underbrace{\begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix}}_{A_{\text{off}}} x + \underbrace{\begin{bmatrix} 0 \\ 0 \end{bmatrix}}_{B_{\text{off}}} v_{in} \tag{15}$$

Let D be the duty cycle ($0 \leq D \leq 1$). The averaged state-space model is:

$$A = DA_{\text{on}} + (1 - D)A_{\text{off}}$$
$$B = DB_{\text{on}} + (1 - D)B_{\text{off}}$$

Thus:

$$\dot{x} = Ax + Bv_{in} \tag{16}$$

$$y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} x \tag{17}$$

Since $A_{\text{on}} = A_{\text{off}}$, we have:

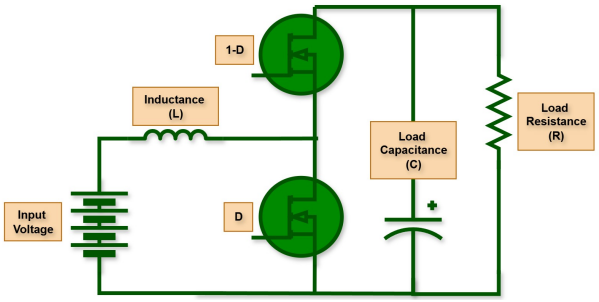
$$A = \begin{bmatrix} 0 & -\frac{1}{L} \\ \frac{1}{C} & -\frac{1}{RC} \end{bmatrix} B = \begin{bmatrix} \frac{D}{L} \\ 0 \end{bmatrix} \tag{18}$$

Therefore, the averaged model is:

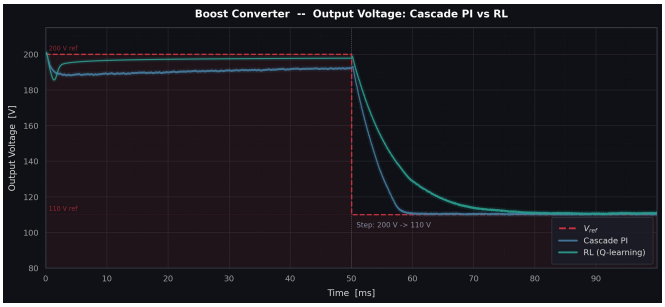
$$\begin{aligned} \frac{di}{dt} &= -\frac{v}{L} + \frac{Dv}{L} \\ \frac{dv}{dt} &= \frac{i}{C} - \frac{v}{RC} \end{aligned} \tag{19}$$

TABLE I. DC power electronics simulation parameters.

No	Parameters	Value
1	Switching frequency	50 $\times 10^3$Hz
2	Controller frequency	25 $\times 10^3$Hz
3	Load resistance (R)	30 Ω
4	Load capacitance (C)	510 μF
5	Inductance (L)	1 mH
6	Input Voltage (v_{in})	100 V



(a) Topology of Boost Converter.



(b) Change in output voltage reference from 200 V to 110 V.

Fig. 3. Case 2: Boost Converter.

B. Case 2: DC-DC Boost Converter

For the boost converter, the topology is presented in Fig. 3a. The voltage control comparison between the newly introduced RL strategy and the traditional cascaded PI controller is presented in Fig. 3b. For an input voltage of 100 V, the load voltage is dropped from an initial value of 200 V to 110 V at t = 0.05 s. RL strategy shows competitive results in comparison to the cascaded PI controller.

Keeping the naming consistent with the state space modeling for a buck converter in the previous subsection, the average state space model for a boost converter is given by:

$$\begin{bmatrix} \frac{di(t)}{dt} \\ \frac{dv(t)}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1-D(t)}{L} \\ \frac{1-D(t)}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} \frac{1}{L} \\ 0 \end{bmatrix} v_{in}(t) \quad (20)$$

The averaged state equations are:

$$\begin{aligned} \frac{di}{dt} &= -\frac{1-D(t)}{L} v + \frac{v_{in}}{L}, \\ \frac{dv}{dt} &= \frac{1-D(t)}{C} i - \frac{v}{RC}. \end{aligned} \quad (21)$$

To address the control variable $D(t)$, which is directly tied to the states in a boost converter, small-signal linearization is applied around a steady state operation point as follows:

$$\begin{aligned} i(t) &= i + \hat{i}(t) \\ v(t) &= v + \hat{v}(t) \\ D(t) &= D + \hat{D}(t) \end{aligned} \quad (22)$$

Linearizing using the equations (21), (22),

$$\frac{d\hat{i}}{dt} = -\frac{1-D}{L}\hat{v} + \frac{v}{L}\hat{D}, \quad (23)$$

$$\frac{d\hat{v}}{dt} = \frac{1-D}{C}\hat{i} - \frac{i}{C}\hat{D} - \frac{\hat{v}}{RC}. \quad (24)$$

The matrix form representation for the above equations is then given by:

$$\dot{\hat{\mathbf{x}}}(t) = \underbrace{\begin{bmatrix} 0 & -\frac{1-D}{L} \\ \frac{1-D}{C} & -\frac{1}{RC} \end{bmatrix}}_{A_{\text{lin}}} \hat{\mathbf{x}}(t) + \underbrace{\begin{bmatrix} \frac{v}{L} & \frac{1}{L} \\ -\frac{i}{C} & 0 \end{bmatrix}}_{B_{\text{lin}}} \begin{bmatrix} \hat{D}(t) \\ \hat{v}_{in} \end{bmatrix} \quad (25)$$

where $\hat{x}(t) = \begin{bmatrix} \hat{i}(t) \\ \hat{v}(t) \end{bmatrix}$.

C. Case 3: DC-DC Buck-Boost Converter

Finally, the topology of the buck-boost converter is presented in Fig. 4a, and the voltage control responses by the RL and the cascaded PI controller are presented in Fig. 4b. The figure shows a voltage drop from an initial value of 200 V to 50 V with the input voltage to the buck-boost converter set at 100 V. Once again, the RL's performance is on par with the cascaded PI control strategy. The initial transient is consistent with the Q function's learning pattern to realize the converter's ability of stepping-up and stepping-down of voltage at the output.

The average state space model (over switch-on and switch-off) dynamics for a buck-boost converter is given by:

$$\begin{bmatrix} \frac{di(t)}{dt} \\ \frac{dv(t)}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1-D(t)}{L} \\ \frac{1-D(t)}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} i(t) \\ v(t) \end{bmatrix} + \begin{bmatrix} \frac{D}{L} \\ 0 \end{bmatrix} v_{in}(t). \quad (26)$$

Once again, we notice the existence of the control variable D being directly related to the state x, leading to the process of linearizing around a certain operation point. The linearized model for a buck-boost converter is given by:

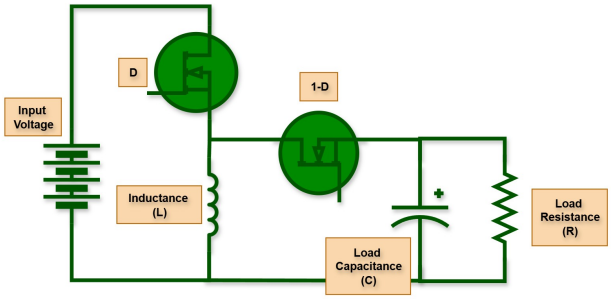
$$\begin{bmatrix} \frac{d\hat{i}(t)}{dt} \\ \frac{d\hat{v}(t)}{dt} \end{bmatrix} = \begin{bmatrix} 0 & -\frac{1-D}{L} \\ \frac{1-D}{C} & -\frac{1}{RC} \end{bmatrix} \begin{bmatrix} \hat{i}(t) \\ \hat{v}(t) \end{bmatrix} + \begin{bmatrix} \frac{D}{L} & \frac{v_{in}+v}{L} \\ 0 & -\frac{i}{C} \end{bmatrix} \begin{bmatrix} \hat{v}_{in}(t) \\ \hat{D}(t) \end{bmatrix} \quad (27)$$

D. Comparison review

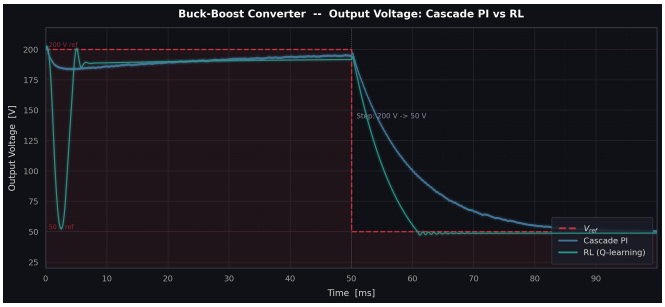
It is to be noted that the RL algorithm's Q-value is still adapting/learning during the run. Also, RL controller's Q-value acts as a soft adaptive gain rather than a saturating integrator which can make it produce an output (duty) with lesser chattering. The RL controller would improve with a longer training horizon or a pre-trained Q initialization rather than initializing it as $Q = 0$.

IV. Conclusion

In this technical report, the application of Reinforcement Learning (RL) is studied on dc-dc non-isolated power converters. Based on the simulation results, it looks clear that the RL strategy shows competitive results in comparison to the cascaded PI control strategy while attaining reference trajectories during transient and steady state conditions.



(a) Topology of Buck-Boost Converter.



(b) Change in output voltage reference from 200 V to 50 V.

Fig. 4. Case 3: Buck-Boost Converter.

REFERENCES

[1] D. Ernst, M. Glavic, and L. Wehenkel, "Power systems stability control: Reinforcement learning framework," *IEEE Transactions on Power Systems*, vol. 19, no. 1, pp. 427–435, 2004.

[2] M. Glavic, "(deep) reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annual Reviews in Control*, vol. 48, pp. 22–35, 2019.

[3] "A critical review of safe reinforcement learning strategies in power and energy systems," *Engineering Applications of Artificial Intelligence*, 2025.

[4] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. MIT Press, 2018. [Online]. Available: http://incompleteideas.net/book/the-book-2nd.html

[5] Y. Zhou, L. Zhou, Z. Yi, D. Shi, and M. Guo, "An in-depth review of model-free deep reinforcement learning in power system control," *IEEE Access*, 2024.

[6] B. L. Thayer and T. J. Overbye, "Deep reinforcement learning for electric transmission voltage control," *arXiv preprint*, 2020, arXiv:2006.06728. [Online]. Available: https://arxiv.org/abs/2006.06728

[7] D. Shi *et al.*, "Implementing deep reinforcement learning-based grid voltage control in real-world power systems: Challenges and insights," *arXiv preprint*, 2024, arXiv:2410.19880. [Online]. Available: https://arxiv.org/abs/2410.19880

[8] H. Wei, S. Chang, and J. Zhang, "Graph-based topological embedding and deep reinforcement learning for autonomous voltage control in power system," *Sensors*, vol. 25, no. 3, p. 733, 2025.

[9] "A comprehensive review of reinforcement learning-based voltage control in smart grids," *Renewable and Sustainable Energy Reviews*, 2025.

[10] "Deep reinforcement learning for volt-var control: Recent advances and future challenges," in *Proc. SPIE 13275, Sixth International Conference*, 2024.

[11] "A deep reinforcement learning approach to dc-dc power electronic converter control with practical considerations," *Energies*, vol. 17, no. 14, p. 3578, 2024.

[12] S. Shahria, A. Harun-Ur Rashid *et al.*, "Proximal policy optimization-based reinforcement learning approach for dc-dc boost converter control: A comparative evaluation against traditional control techniques," *Heliyon*, vol. 10, no. 18, p. e37823, 2024.

[13] D. Lee, B. Kim, S. Kwon, N.-D. Nguyen, M. K. Sim, and Y. I. Lee, "Reinforcement learning-based control of dc-dc buck converter considering controller time delay," *IEEE Access*, vol. 12, pp. 118 442–118 452, 2024, semantic Scholar CorpusID:272070413.

[14] "Voltage control of dc-dc converters through direct control of power switches using reinforcement learning," *Engineering Applications of Artificial Intelligence*, vol. 120, p. 105880, 2023.

[15] "Deep reinforcement learning for power converter control: A comprehensive review of applications and challenges," *IEEE Open Journal of Power Electronics*, 2025.

[16] A. Rajamallaiah, S. P. K. Karri, and Y. R. Shankar, "Deep reinforcement learning based control strategy for voltage regulation of dc-dc buck converter feeding cpls in dc microgrid," *IEEE Access*, vol. 12, pp. 17 419–17 430, 2024.

[17] K. Kumar, S. Kwon, and S. Bae, "Deep reinforcement learning-based control strategy for integration of a hybrid energy storage system in microgrids," *Journal of Energy Storage*, vol. 108, p. 114936, 2025.