TU Delft

Delft University of Technology

Document Version
Final published version

Citation (APA)
Kermansaravi, A., Vahedi, H., Alquennah, A. N., Trabelsi, M., & Lekić, A. (2025). Reinforcement Learning Controller Design for Full-Bridge Active Rectifier. In *Proceedings of the 2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe)* IEEE. https://doi.org/10.1109/ISGTEurope64741.2025.11305659

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Green Open Access added to TU Delft Institutional Repository
as part of the Taverne amendment.

More information about this copyright law amendment
can be found at https://www.openaccess.nl.

Otherwise as indicated in the copyright section:
the publisher is the copyright holder of this work and the
author uses the Dutch legislation to make this work public.

2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe) | 979-8-3315-2503-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/ISGTEurope64741.2025.11305659

Reinforcement Learning Controller Design for Full-Bridge Active Rectifier

Azadeh Kermansaravi[1*], Hani Vahedi[1], Alamera Nouran Alquennah[2], Mohamed Trabelsi[3], Aleksandra Lekić[1]
[1]Delft University of Technology, Delft, Netherlands
[2]Texas A&M University, College Station, TX, USA
[3]Electronics and Communications Engineering Department, Kuwait College of Science and Technology, Kuwait
* Corresponding author
{Z.Kermansaravi, A.Lekic, H.Vahedi}@tudelft.nl, alamera_nq@tamu.edu, m.trabelsi@kcst.edu.kw

***Abstract*—This paper presents a reinforcement learning controller (RLC) for a single-phase full-bridge rectifier as an interface for a battery energy storage system (BESS). A novel solution is presented that combines the traditional proportional-integral (PI) regulator with an RL-based control strategy using a proximal policy optimization (PPO) agent. In a high-fidelity Simulink-based digital twin setup, the agent learns to perform optimal switching actions for a single-phase full-bridge rectifier to achieve accurate current tracking and improved power quality. Simulation results show stable DC voltage regulation at 200 V, tracking response under 0.1 s, and harmonic compliance with THD equal to 2.38%. The hybrid control strategy guarantees robust dynamic performance and adaptability in the context of renewable energy and storage systems' varying source and load conditions. The findings demonstrate the potential of coupling AI-driven control with digital twins to empower the autonomy and resilience of future smart energy systems.**

***Index Terms*—Rectifier, Reinforcement Learning, Full-Bridge Rectifier, Battery Energy Storage Systems.**

I. Introduction

The growing presence of renewable energy systems like photovoltaic (PV) systems and battery energy storage systems (BESS) [1] adds dynamic and non-linear behaviour to the modern power grid due to converter switching actions (e.g., inverters, DC/DC converters, rectifiers) [2]. Such non-linearities challenge traditional linear controllers due to uncertain and time-varying source/load profiles in renewable systems [3]; however, they can also be leveraged for ancillary services [4].

To address these challenges, intelligent control methods that use reinforcement learning (RL) have emerged as promising alternatives [5]–[7]. Traditional linear control methods such as PI controllers and model predictive control (MPC) often require accurate mathematical models and are limited in handling the non-linearities and uncertainties of modern converter-dominated power grids. In contrast, RL methods are data-driven and model-free, capable of learning optimal control policies directly through system interaction without assuming a known model. This characteristic allows RL to adapt in real-time to operating conditions such as grid fluctuations and uncertainty. The data-driven approach used in RL allows for adaptable, robust, and autonomous control of complex systems for which accurate modelling is impossible or impractical. RL provides a natural fit for non-linear power electronic systems that may lack robustness under traditional model-based approaches. While RL may share some optimization goals with MPC-based methods, its fundamental advantage lies in experiential learning without requiring knowledge of plant dynamics or parameter assumptions.

RL has been applied in various areas of power electronics and systems [8], [9]. It has tuned neuro-fuzzy controllers in microgrids, stabilized DC/DC converters via learned duty cycles, and optimized the control parameters of dual active bridge converters for reduced reactive power, power loss minimization, and current stress mitigation. It has also contributed to predictive control approaches by automating the tuning of the weighting factors, meaning less dependence on explicit models [10], [11]. In particular, RL is applicable in finite-set model predictive control (FS-MPC), enabling both autonomous tuning of weighting factors and model-free replication of predictive policies. These approaches enhance the flexibility of control to dynamic objectives such as THD or switching frequency minimization, and highlight RL's growing role in high-performance converter control [12].

Recent research indicates RL's promise in real-world applications. In [13], an RL controller was applied to a 5-level Packed U-Cell (PUC) inverter, where the RL agent was trained using a PI-based reward function for both reference current tracking and current surge rejection. The authors proposed that separating capacitor voltage balancing from the RL policy may effectively reduce the dimensionality of the agent's observation while improving the training speed.

This paper presents a hybrid control framework combining a PI controller and a PPO-based RL agent. The PPO agent is trained to derive optimal switching actions for a single-phase full-bridge rectifier interfacing BESS to the grid. The learning process occurs in a high-fidelity digital twin in Simulink, which imitates the system dynamics of the physical grid and the dynamics of the power semiconductor converter. Although in the real case, the accent would be on implementing the RL control framework on the three-phase voltage source converter, this is the first step towards the single-phase control design for the same converter architecture.

The proposed method achieves accurate current follow-

979-8-3315-2503-3/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: TU Delft Library. Downloaded on January 08,2026 at 14:42:27 UTC from IEEE Xplore. Restrictions apply.

ing and enhanced power quality by integrating RL into the switching layer, even with changing operating conditions. The hybrid approach utilizes both classical control and AI-based control to be stable and maintain performance in real-time operation. The following sections describe the design of the RL controller, system architecture, and the training procedure in the Simulink-based environment.

The paper is organized as follows. Section II introduces the system and conventional control. Section III describes the proposed RLC design. Section IV summarizes the results, and Section V gives concluding remarks.

II. Full-Bridge Rectifier

Fig. 1 shows a full-bridge rectifier as one of the widely used rectifiers in power electronics systems using cascaded PI controllers [14].

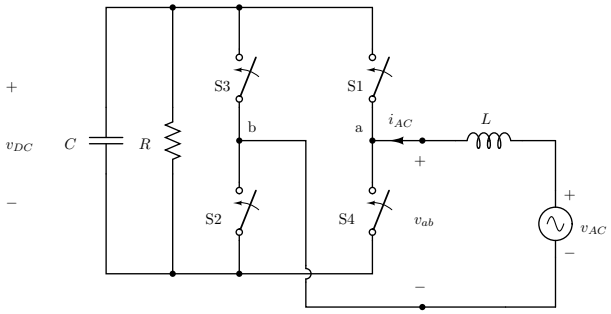

Fig. 1: Single-phase full-bridge rectifier.

To control this converter, the control strategy is depicted in Fig. 2. As observed, the current reference's amplitude is calculated through a voltage loop. As well, the phase of the current reference is obtained from the AC voltage, v_{AC}, with amplitude V_S, through the discrete PLL (created as a proportional-integral-derivative (PID)). Eventually, the current reference $i_{ref} = i_{ref,max}\sin(\omega t)$ is compared to the actual current, and the error is minimized using a PI block as the current loop controller. Action is an output of the PI controller, with input $i_{error} = i_{ref} - i_{AC}$. The action function provides switching pulses based on the logic:

$$S2 = \overline{S1}, \tag{1a}$$

$$\begin{cases} S1 = 1, & \text{if action} \geq \text{ carrier}, \\ S1 = 0, & \text{otherwise}, \end{cases} \tag{1b}$$

$$\begin{cases} S3 = 1, S4 = 0, & \text{if action} \leq -\text{carrier}, \\ S3 = 0, S4 = 0, & \text{otherwise}. \end{cases} \tag{2}$$

TABLE I: Full-Bridge Rectifier and PI Controller Parameters

Parameter	Value	Parameter	Value
$v_{DC,\text{ref}}$	200 V	V_S	170
L	2.5 mH	C	1 mF
R	40 Ω	K_p^{PLL}	180
K_i^{PLL}	3200	K_d^{PLL}	1
$K_p^{i_{AC}}$	0.18	$K_i^{i_{AC}}$	5
$K_p^{v_{DC}}$	0.01	$K_i^{v_{DC}}$	5

In table I, the values inside the rectifier and the parameters of the PI/PID control loops are listed. Please note that for controllers, with K_x^y are denoted gains for $x \in \{p, i, d\}$ being proportional, integral, and derivative gains, respectively, and $y \in \{v_{DC}, i_{AC}, PLL\}$ denoting that it represents controller for DC voltage, AC current and PLL, respectively.

III. Proposed RLC Design

This section presents the design and implementation of an RL-based controller for a single-phase full-bridge rectifier. The system creates an interface between a BESS and the grid. It aims to regulate the current i_{AC} to follow the grid voltage as a unity power factor mode of operation while keeping the capacitor voltage v_{DC} stable at a desired level.

To clarify terminology, the RL agent is defined as the decision-maker that determines a policy to control switching actions based on system observations, without requiring explicit system models. This architecture is particularly applicable in current smart grid applications, where rapid response to change and focused control of power flow are necessary [15]. The system operates in charging mode in the presented system, where the energy is stored in the BESS, but the controller design can be expanded to support bidirectional operation, discharging into the grid (i.e., in peak demand or in offering ancillary services). Unlike traditional applications, the suggested RL controller learns an optimal policy through interactions in the environment. It determines real-time switching actions to ensure accurate current tracking and voltage regulation, without relying on physical modelling or rule-based logic [13], [16].

A. System Architecture

The digital twin includes high-fidelity time-domain models of switching dynamics, control loops, and power components (e.g., parasitics, measurement noise, and feedback delay noises) with a sampling time of 20 μs, which guarantees accurate emulation of a physical s, switching during training. The RL agent interacts with the control block and switching logic to map signal measurements to control actions. The model structure supports closed-loop training, enabling continuous policy adaptation to system drift, disturbances, and operational uncertainties.

B. Observation and Action Spaces

The RL agent operates with a normalized observation space and a compact discrete action space, which helps to maintain numerical stability and promotes effective learning.

The observation space includes three normalized variables:

- i_{AC}: Measured AC current.
- i_{ref}: Reference current synchronized with the grid voltage.
- $i_{ref,max}$: Maximum expected reference current, used for normalization.

Observations are normalized to the range [-1, 1] to provide consistent training behavior under varying grid or load conditions.

Authorized licensed use limited to: TU Delft Library. Downloaded on January 08,2026 at 14:42:27 UTC from IEEE Xplore. Restrictions apply.

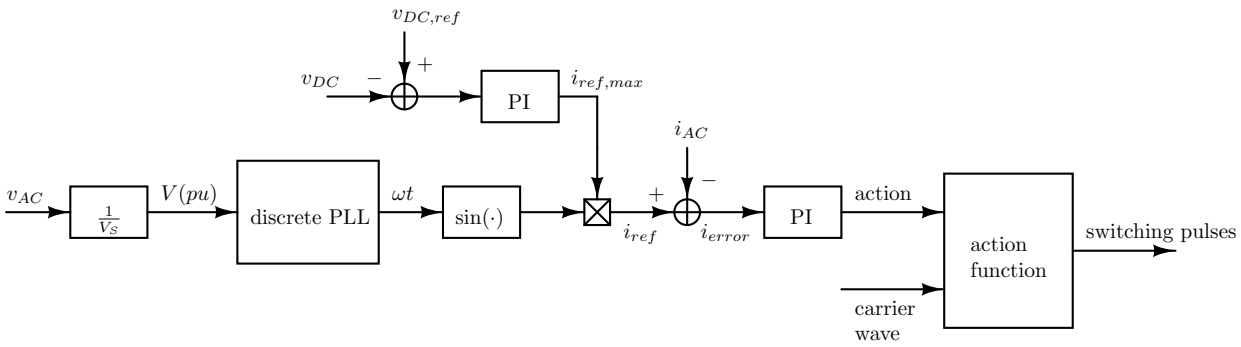

Fig. 2: Proposed Control of the single-phase full-bridge rectifier under study.

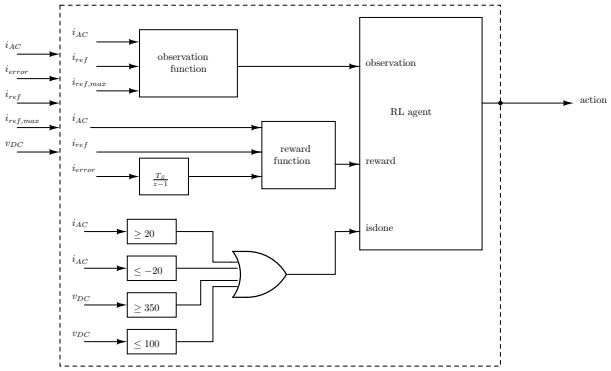

Fig. 3: RL design for the full-bridge control.

The action space has been simplified to consist of three discrete switching options, each linked to a specific control signal for the inverter:

$$\text{Action Space} = \{-1, 0, +1\}, \tag{3}$$

The values correspond to:

1) -1: Reverse polarity;
2) 0: Neutral or no switching;
3) +1: Forward polarity.

This configuration not only mitigates complexity in learning but also speeds up convergence while retaining sufficient expressiveness needed for effective control. The nature of this action space around zero enhances the learning dynamics in the PPO algorithm by promoting gradient symmetry.

C. Actor-Critic Network Architecture

The policy is trained using a PPO algorithm, which is effective in environments with discrete actions and continuous state spaces. The RL agent has an actor-critic structure:

- *Critic Network:* Estimates the state-value function V_c, providing a baseline to reduce variance in policy updates.
- *Actor Network:* Learns a stochastic policy $\pi(a \mid s)$ that maps observations to action probabilities, selecting actions that maximize long-term rewards.

D. Reward Function

The reward function is carefully designed to promote accurate current tracking, penalize deviations, and discourage instability.

Proportional Error: The immediate tracking error is defined as

$$e_p = i_{ref} - i_{AC} = i_{error}, \tag{4}$$

which encourages the agent to drive the actual current towards the reference value. Normalization of Errors: To handle varying magnitudes of i_{ref}, both the proportional and integral errors are normalized:

$$\text{normalizedError} = \begin{cases} 0, & \text{if } i_{ref} = 0, \\ \frac{e_p}{i_{ref}}, & \text{otherwise}, \end{cases} \tag{5}$$

$$\text{normalizedIntegralError} = \begin{cases} 0, & \text{if } i_{ref} = 0, \\ \frac{\text{integralError}}{i_{ref}}, & \text{otherwise}. \end{cases} \tag{6}$$

This normalization ensures numerical consistency and improves generalization across operating conditions.

Overshoot Penalty: To discourage instability, an additional penalty is imposed when the actual current exceeds the reference:

$$\text{overshootPenalty} = \begin{cases} |i_{ref} - i_{AC}|, & \text{if } |i_{ref}| > |i_{AC}|, \\ 0, & \text{otherwise}. \end{cases} \tag{7}$$

Reward Calculation: The reward at each step r_k is computed as a weighted sum of penalties:

$$\begin{aligned} r_k = & -|\text{normalizedError}| - \alpha|\text{normalizedIntegralError}| \\ & - \text{overshootPenalty} + \delta, \end{aligned} \tag{8}$$

where $\alpha = 0.01$ is the integral error weight, and $\delta = 0.01$ is a small positive bias encouraging exploration. This form promotes accurate current tracking, penalizes excessive error accumulation, and maintains system stability.

To ensure numerical stability, the reward is verified to be finite at each timestep. If any input (e.g., i_{ref}, i_{AC}, or integral error) results in a non-finite reward, diagnostic information is printed and training is halted for debugging purposes.

E. Training Process

The agent is trained using MATLAB's RL Toolbox using the PPO algorithm. The system model represents a single-phase full-bridge rectifier interfaced with BESS. The model includes sensors for AC i_{AC}, reference current i_{ref}, and associated error signals. PPO hyperparameters include a discount factor

Authorized licensed use limited to: TU Delft Library. Downloaded on January 08,2026 at 14:42:27 UTC from IEEE Xplore. Restrictions apply.

of 0.998, a clip factor of 0.2, entropy regularization, and an Adam optimizer for both networks.

Training runs for up to 10,000 episodes with early stopping based on average reward over a 50-episode window. The best-performing agent is saved, and training uses parallel workers to reduce runtime when hardware permits.

The objective is to learn a switching policy that ensures sinusoidal current tracking and DC voltage regulation under varying grid/load conditions.

IV. Results and Discussions

The proposed controller, as shown in Fig. 2 and 3, is implemented on the full-bridge rectifier and simulated in MATLAB/Simulink. System parameters are listed in Table II. Results include steady-state, transient, and learning performance.

TABLE II: Simulation Parameters

Grid Voltage (RMS) (v_{AC})	$120V$
DC Voltage (v_{DC})	$200V$
Inductor (L)	$2.5mH$
Auxiliary Capacitor	$1mF$
Sampling Time	$20\mu s$

Fig. 4 shows that the PI controller regulates the DC voltage at 200 V while the RL agent synchronizes grid current with voltage. The current waveform complies with harmonic standards with a THD of 2.4% (Fig.5).

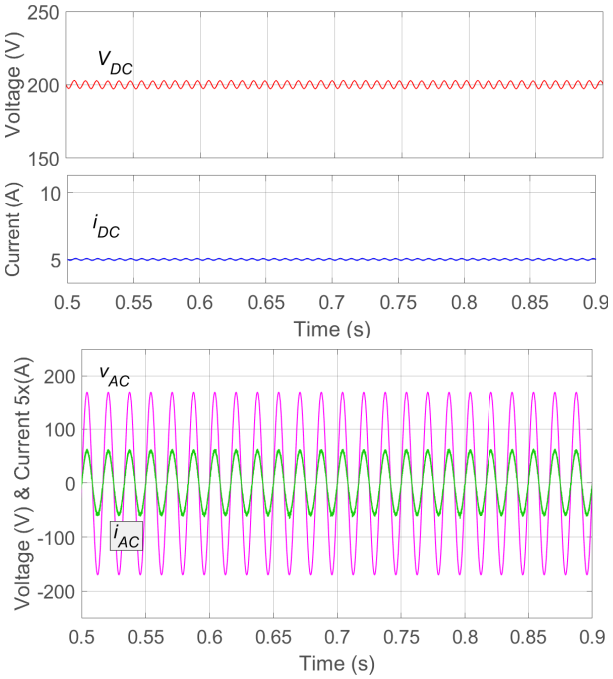

Fig. 4: Steady state results (Current scaled up by 5x)

Under the load changes from $40\,\Omega$ to $20\,\Omega$ (Fig. 6), the DC voltage is balanced within 0.1 s, and the AC current remains well-regulated and synchronized, confirming dynamic tracking performance.

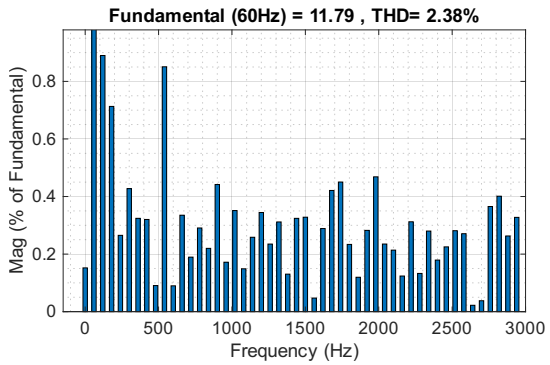

Fig. 5: FFT Analysis of the AC current

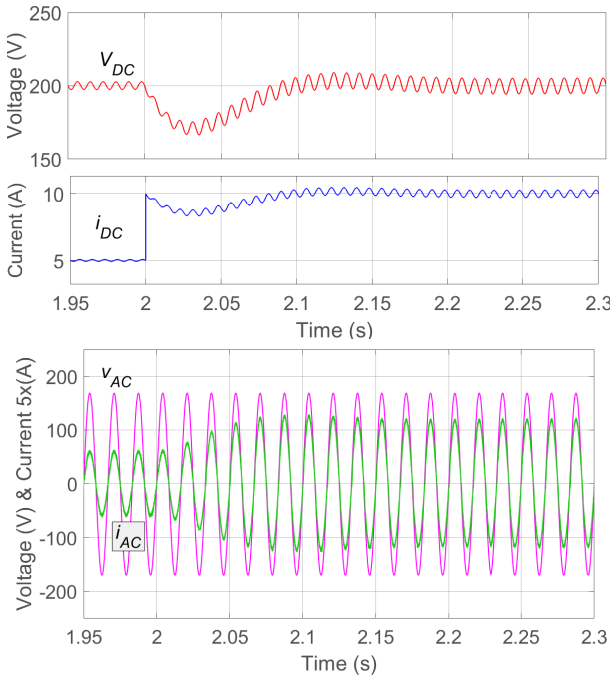

Fig. 6: Results during 100% increase in the load

During AC grid voltage fluctuation (170 V to 200 V peak), the AC current amplitude adjusts accordingly to maintain controlled energy flow (Fig. 7).

Fig. 8 illustrates key metrics from the RL agent's training process, including episode reward, average reward progression, action distribution during deployment, and tracking error over time. The steady increase in average reward indicates effective policy learning. The action distribution shows a preference for forward and neutral states, minimizing unnecessary switching. Tracking error decreases over episodes, reflecting improved control and adaptability.

In summary, the hybrid controller demonstrates robustness against grid and load variations, achieving fast and adaptive current tracking via reinforcement learning. While this work does not provide a full comparison against advanced control methods such as Model Predictive Control or fuzzy logic. Existing literature highlights the potential of RL-based strategies to outperform fixed-gain controllers in non-linear and highly dynamic cases. Compared to traditional PI control, the proposed method offers improved adaptability and power

Authorized licensed use limited to: TU Delft Library. Downloaded on January 08,2026 at 14:42:27 UTC from IEEE Xplore. Restrictions apply.

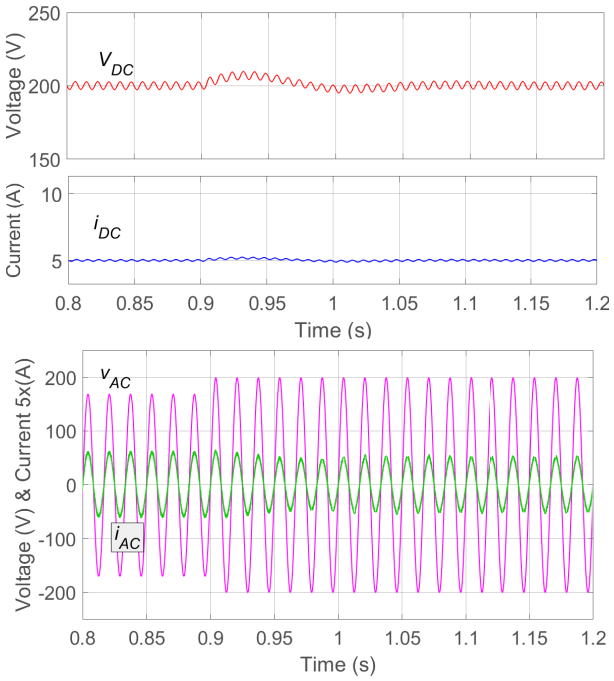

Fig. 7: Results during AC voltage fluctuation

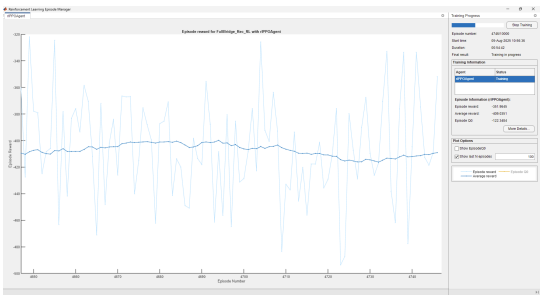

Fig. 8: RL agent training

quality under varying conditions. Future work includes a broader benchmarking with advanced control methods.

V. Conclusion

In this paper, an RLC has been designed and implemented on a single-phase full-bridge rectifier. The DC voltage was balanced with a PI controller, and its output was sent to the RLC as the amplitude of the reference current. The RL agent has been trained in an environment to regulate the AC side current to be in phase with the grid voltage, featuring power factor correction. As shown by simulation results, the RLC properly maintained the DC voltage at 200V despite grid and load disturbances, achieved current tracking within 0.1s, and met harmonic standards with a THD of 2.38%. Although this study was limited to simulation and a single-phase implementation, this framework provides significant potential for extending AI-based control strategies to three-phase systems and for real-time embedded applications in energy storage and power conversion. Future work includes experimental validation through hardware-in-the-loop and prototype testing to comprehensively assess controller performance and robustness.

VI. Acknowledgment

The project was in part funded by Kuwait Foundation for the Advancement of Sciences (KFAS) under project code: CN23-13EE-1882, and in part by the CRESYM project Harmony (https://cresym.eu/harmony/).

References

[1] A. Fu, A. Lekić, K.-N. D. Malamaki, G. C. Kryonidis, J. M. Mauricio, C. S. Demoulias, P. Palensky, and M. Cvetković, "A stochastic simulation-based approach for sizing dres penetration level and bess capacity in distribution grids," *International Journal of Electrical Power & Energy Systems*, vol. 161, p. 110164, 2024.

[2] K.-N. Malamaki, G. Kryonidis, S. Gkavanoudis, A. Fu, A. Lekić, J. M. Mauricio, E. Kontis, C. Demoulias, and M. Cvetković, "EASY-RES D6.5 report on the evaluation of project KPIs," 2021.

[3] M. Trabelsi, A. N. Alquennah, and H. Vahedi, "Review on single-dc-source multilevel inverters: Voltage balancing and control techniques," *IEEE Open Journal of the Industrial Electronics Society*, vol. 3, pp. 711–732, 2022.

[4] C. S. Demoulias, K.-N. D. Malamaki, S. Gkavanoudis, J. M. Mauricio, G. C. Kryonidis, K. O. Oureilidis, E. O. Kontis, and J. L. Martinez Ramos, "Ancillary services offered by distributed renewable energy sources at the distribution grid level: An attempt at proper definition and quantification," *Applied Sciences*, vol. 10, no. 20, 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/20/7106

[5] Y. Fujimoto, A. Kaneko, Y. Iino, H. Ishii, and Y. Hayashi, "Challenges in smartizing operational management of functionally-smart inverters for distributed energy resources: A review on machine learning aspects," *Energies*, vol. 16, no. 3, p. 1330, 2023.

[6] D. Alfred, D. Czarkowski, and J. Teng, "Reinforcement learning-based control of a power electronic converter," *Mathematics*, vol. 12, no. 5, p. 671, 2024.

[7] A. N. Alquennah, M. A. Saleh, H. Abu-Rub, A. Ghrayeb, S. Bayhan, and M. Trabelsi, "Double deep q networks reinforcement learning-based dynamic weighting factor of fcs-mpc for multilevel inverters," in *2024 IEEE 18th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG)*, 2024, pp. 1–7.

[8] 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)*, 2019, pp. 764–769.

[9] M. Gheisarnejad, H. Farsizadeh, and M. H. Khooban, "A novel nonlinear deep reinforcement learning controller for dc–dc power buck converters," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 8, pp. 6849–6858, 2020.

[10] S. Vazquez, D. Marino, E. Zafra, M. D. V. Peña, J. J. Rodríguez-Andina, L. G. Franquelo, and M. Manic, "An artificial intelligence approach for real-time tuning of weighting factors in fcs-mpc for power converters," *IEEE Transactions on Industrial Electronics*, vol. 69, no. 12, pp. 11 987–11 998, 2021.

[11] Z. Zhang, W. Tian, W. Xiong, and R. Kennel, "Predictive torque control of induction machines fed by 3l-npc converters with online weighting factor adjustment using fuzzy logic," in *2017 IEEE Transportation Electrification Conference and Expo (ITEC)*. IEEE, 2017, pp. 84–89.

[12] Y. Wan, Q. Xu, and T. Dragičević, "Reinforcement learning-based predictive control for power electronic converters," *IEEE Transactions on Industrial Electronics*, 2024.

[13] A. Kermansaravi, A. N. Alquennah, A. Lekić, M. Trabelsi, A. Ghrayeb, H. Abu-Rub, and H. Vahedi, "Reinforcement learning based control of grid-connected puc5 inverter," in *IECON 2024-50th Annual Conference of the IEEE Industrial Electronics Society*. IEEE, 2024, pp. 1–6.

[14] H. Vahedi, "5-level packed u-cell (puc5) active front-end rectifier," *IEEE Open Journal of Power Electronics*, vol. 6, pp. 1022–1027, 2025.

[15] A. Rajamallaiah, S. P. K. Karri, M. L. Alghaythi, and M. S. Alshammari, "Deep reinforcement learning based control of a grid connected inverter with lcl-filter for renewable solar applications," *IEEE Access*, vol. 12, pp. 22 278–22 295, 2024.

[16] R. kumar Mahto, P. Mishra, and A. Alam, "Design of controller for three-phase grid-connected inverter using reinforcement learning," in *2024 1st International Conference on Innovative Sustainable Technologies for Energy, Mechatronics, and Smart Systems (ISTEMS)*. IEEE, 2024, pp. 1–6.

Authorized licensed use limited to: TU Delft Library. Downloaded on January 08,2026 at 14:42:27 UTC from IEEE Xplore. Restrictions apply.