Electrical Power and Energy Systems 181 (2026) 112166

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Hierarchical multi-agent drl-based secondary control for real-time voltage and frequency regulation in renewable-integrated microgrids

Darioush Razmi [a], Peyman Razmi [b],*, Jose Rodriguez [c], Cristian Garcia [d], Zhenbin Zhang [a],*

[a] School of Electrical Engineering, Shandong University, Jinan 250061, China
[b] Smart Electric Power Group, Vaasa, Vaasa University, Finland
[c] Universidad San Sebastian, Bellavista 7, Santiago, Chile
[d] Department of Electrical Engineering, Faculty of Engineering, Universidad de Talca, Curico, Chile

ARTICLE INFO

Keywords:
Inverter-based microgrids
Multi-agent system
Reinforcement learning
Deep deterministic policy gradient
Secondary control
Total harmonic distortion (THD)

ABSTRACT

The integration of inverter-based microgrids (MGs) often induces variations in system characteristics and disturbances in the network. In islanded operation, voltage and frequency oscillations arise due to the lack of synchronization with the main grid. Ensuring reliable operation under such conditions requires advanced control techniques and proactive monitoring. This paper proposes an intelligent multi-agent secondary control architecture based on the Deep Deterministic Policy Gradient (DDPG) algorithm, where reinforcement learning (RL) agents dynamically adjust the parameters of proportional–integral (PI) controllers to mitigate disturbances and adapt to system uncertainties and varying operating conditions. The proposed framework employs a two-layer hierarchical structure. At the lower layer, two independent RL agents regulate frequency and voltage by tuning PI controller gains in real-time, ensuring stable operation under fluctuating loads and distributed generation. The upper layer incorporates a supervisory agent that monitors harmonic distortion and computes corrective signals sent to the droop controller, enabling coordinated adaptation and enhancing dynamic stability by reducing transient oscillations. As a proof-of-concept, the proposed multi-agent framework is implemented on a single DG unit within a four-converter microgrid, demonstrating the potential for improved dynamic performance under realistic operational scenarios. Validation in MATLAB/Simulink demonstrates substantial improvements in system performance. Frequency settling time decreased from 0.19 s (at $K_p = 0.3$) to 0.10 s (at $K_p = 0.1$), and voltage settling time decreased from 0.20 s (at $K_p = 0.3$) to 0.13 s (at $K_p = 0.1$). Total harmonic distortion (THD) significantly reduced from 2.07% (at $K_p = 0.3$) and 2.03% (at $K_p = 0.2$) to 0.68% (at $K_p = 0.1$). These results highlight the scalability and effectiveness of the proposed multi-agent DDPG-based secondary control, offering a reliable and intelligent solution for advanced MG energy management under realistic operating conditions.

1. Introduction

In recent decades, there has been an abrupt increase in the use of distributed generation (DG) systems. DG units have benefits such as reduced pollution among others as compared to conventional power generation systems [1]. In inverter-based microgrids, power electronic inverters connect distributed energy resources (DERs) such as solar cells, wind turbines or battery storages to the local AC network. These inverters convert the DC power generated by DERs into AC power compatible with the grid. However, the variability and intermittency associated with renewable energy sources introduce challenges in maintaining stable voltage and frequency levels [2]. Maintaining stable voltage and frequency within a microgrid is vital for the proper functioning of electrical equipment and the overall reliability of the power supply [3].

Over these years, researchers have developed various control techniques to ensure the stability of the grid voltage and frequency. Microgrids use traditional linear control techniques, most notably fixed-parameter PI (proportional integral) controllers and droop control for voltage and frequency regulation. These approaches, based on classical control theory, are regarded for their simplicity and ease of implementation. However, they do not possess all the features of traditional controls, particularly in modern microgrids with high renewable power penetration [4]. Droop control modifies the frequency and voltage output of inverters based on their power generation, replicating the

* Corresponding authors.
E-mail addresses: peyman.razmi@uwasa.fi (P. Razmi), zbz@sdu.edu.cn (Z. Zhang).

https://doi.org/10.1016/j.ijepes.2026.112166
Received 9 April 2026; Received in revised form 20 August 2026; Accepted 31 August 2026
Available online 9 September 2026

characteristics of synchronous generators. Droop control is useful for sharing loads and keeping the stability of the system in grid-tied operations. However, it lacks the precision and responsiveness needed to address large fluctuations in renewable energy output, load demand and is ineffective during isolated operation mode when the microgrid is disconnected from the main grid, requiring refined control strategies so as to maintain stability. In this context, a virtual impedance is introduced as a means of implementing this control, which operates on an existing system [5]. Despite the progress made in voltage-reactive power droop controllers, an ongoing challenge remains regarding the accurate sharing of reactive power, particularly in scenarios with substantial line impedances [6]. Moreover, its effectiveness in sharing current is diminished by voltage drop across transmission line impedances. Additionally, higher droop gains can lead to a reduction in bus voltage [7].

PI controllers have been extensively used for secondary control to bring voltage and frequency back to their nominal values upon disturbances. However, the fixed parameters limit their adaptability to changing conditions, resulting in sub-optimal performance and possible instability in dynamic environments. The inability of traditional PI controllers to adapt to real-time variations in generation and load profiles makes them less efficient in zones where renewable energy penetration is high. Nonetheless, in the recent past, researchers have turned their attention towards developing nonlinear controllers that can handle the complexities of nonlinear control theory in order to overcome the limitations of linear control methods.

2. State-of-the-art

2.1. Adaptive and nonlinear control

Adaptive and nonlinear control strategies have been extensively investigated in microgrids to overcome the limitations of conventional linear methods, particularly under high renewable penetration and varying load conditions. These methods aim to provide faster response, better stability, and multi-objective control capabilities. Non-linear control methods are used for DGs and microgrids that offer the advantage of rapid control and the ability to support multiple control objectives simultaneously. Model Predictive Control (MPC) has superior performance in comparison to linear controllers due to its handling of errors, stability, and transient responses [8]. Different variations of MPCs have been applied for V/f restoration. In [9], an MPC-based secondary V/f restoration method is proposed that estimates V/f deviation parameters to enable the restoration process. In [10], a model-based tracking consensus method is proposed for voltage regulation in which the secondary voltage controller acts as a distributed MPC (DMPC) synchronization process through non-linear feedback. A technique designed for restoring voltage and distributing power in medium-voltage DC networks is presented. The strategy employs a non-linear model predictive regulator that is embedded into a sophisticated droop control framework [11]. The technique of model-based control is capable of precisely classifying the system under its regulation. Nonetheless, these methods are incapable of recognizing non-linear systems in real-world scenarios, hence their applicability is limited. Sliding mode control systems, despite their efficacy, are often considered impractical due to their complexity and the occurrence of high-frequency chattering, which can lead to undesirable oscillations and increased stress on actuators [12]. Despite the theoretical advantages concerning stability and transients of MPC and sliding mode controllers, their practical implementation faces difficulties due to computational complexity, challenges arising from parameter tuning, and susceptibility to inaccuracies in model representations. A key trade-off thus develops between precision and implementability in the non-linear control of microgrids. Several adaptive control methods have been put forward for the regulation of voltage and frequency, such as an adaptive optimal output-feedback control algorithm using adaptive dynamic programming [13], an adaptive proportional-repetitive control (PRC) scheme [14], an adaptive model predictive control (AMPC) technique [15], an adaptive droop control and fuzzy PI control mechanisms [16]. Though adaptive control techniques ostensibly have advantages in such respects as their abilities to adapt to changing system conditions and uncertainty, their implementation may face serious challenges when it comes to droop controllers for microgrids. Hence, some researchers and practitioners might see them as unfeasible within present-day microgrid applications, preferring much simpler alternative controls that are considered more reliable. Theoretically adaptive methods can provide robustness against uncertainty and fluctuating conditions; however, they are usually afflicted by problems of convergence and heavy computational burden in real-time implementation, especially in multi-converter islanded microgrids. This draws attention to a gap in which mere adaptivity does not assure practical feasibility and motivates a quest for intelligent, learning-based alternatives.

2.2. DRL-based control in microgrids

Recent advances in microgrid control have increasingly leveraged deep reinforcement learning (DRL) techniques to address the limitations of conventional linear, nonlinear, and adaptive methods. DRL enables real-time adaptability, multi-objective optimization, and improved handling of uncertainties and non-linearities inherent in inverter-based microgrids. In [17], a self-learning data-driven method is proposed for achieving frequency synchronization autonomously and voltage restoration, especially suited for plug-and-play isolated microgrids. The work presented in [18] offers a voltage and frequency restoration approach via dynamic average consensus algorithm coupled with first-order consensus algorithm based on virtual impedance technique to achieve accurate reactive power sharing. These studies suggest a shift toward decentralized and data-driven methods that can leverage local measurement feedback when combined with consensus algorithms. Nevertheless, these techniques often assume precise communication and network topology constraints that may not hold true in real-world microgrid scenarios, limiting their scalability. Proportional–integral–derivative (PID) controllers can be adapted to manage changing system dynamics within non-linear system control, notably in adaptive control schemes. Conventional PID controllers are valid for linear systems with established parameters, while adaptive PID controllers are required in non-linear or systems having unknown dynamics. The non-linear, multi-input, and complex dynamic nature of microgrids does indeed present significant challenges for control and optimization. These complexities can make it difficult to ensure stable and efficient operation, which can, in turn, affect their practicality and commercial viability [19]. In spite of these challenges, significant advancements have been made in control technologies, such as model-free control algorithms like DDPG aimed at solving them. These efforts are helping reduce some of the limitations attributed to microgrid complexities [20]. Adaptive PID and model-free methods may operate independently of precise models; however, they heavily rely on the design of the reward function and the quality of training data, both of which influence convergence stability. The resulting combination highlights the subtle trade-off between learning ability and controllability. In contrast, DRL-based controllers dynamically adapt their control policies in real-time, offering improved robustness, scalability, and simultaneous voltage and frequency regulation under varying operational conditions. This capability addresses the limitations of traditional linear, nonlinear, and adaptive control methods, particularly in microgrids with high renewable penetration and variable loads. In [21], an adaptive distributed stochastic deep reinforcement learning (DSDRL) approach is proposed that enhances system performance under communication noise and variable delays. In [22], a transferable DRL-based control strategy combining deep Q-learning and transfer learning is proposed to accelerate training and enhance frequency stability in islanded microgrids. In [23], a distributed consensus-based voltage and frequency control method is proposed to effectively mitigate fault-induced delayed voltage recovery. These studies highlight the growing interest in multi-agent and DRL-based methods and provide a context for the novelty of the proposed intelligent multi-agent secondary control framework in this work. The DRL approach integrates features of real-time adaptability and multi-objective optimization. Due to its presumed advantages of improved generalization across operational scenarios, natural handling of nonlinearities, and reduced reliance on human intervention in tuning, DRL has emerged as an alternative to classical and adaptive control methods. However, challenges such as stability guarantees, interpretability, and communication delays remain, and the search for solutions continues as an active area of research. A DDPG algorithm framework is considered to acquire the optimum dispatch strategy through offline training and online decision-making. In [24], a DDPG algorithm is proposed to regulate the load voltage, and in [25], an actor-critic framework is explored to generate additional control actions. These actions prove beneficial for stabilizing frequency by adapting to the unpredictability of load disruptions and renewable energy sources (RESs). A virtual inertia control method based on reinforcement learning and deep deterministic policy gradients is proposed. The proposed approach is demonstrated using a standard microgrid architecture and compared to H-infinity and ideally tuned PI controllers [26]. Employing DDPG and actor-critic frameworks indicates a gradual transformation from single-objective, model-based controllers to multi-objective, model-free Reinforcement Learning controllers. The latter types can optimize voltage, frequency, and inertia emulation simultaneously—capabilities that classical methods cannot effectively achieve together. However, current limitations in deployment readiness are clearly reflected by the requirement for extensive simulation and careful reward shaping. Using the DDPG, an intelligent auxiliary controller for islanded microgrids is suggested. To ensure voltage and frequency consistency, the DDPG processor modifies the output power of the storage components [27]. Deep reinforcement learning regulates the load frequency of a hybrid power system. The complex structure of the interconnected power system is increased for a more practical strategy by adding RESs, variable loads, and electric vehicles. Using an enhanced twin-delayed DDPG-based reinforcement learning agent, the proposed approach adjusts the PID controller parameters [28]. In the distributed frequency control based on DRL, a data-driven strategy is proposed that adaptively determines the optimal cooperative control approach by integrating the traditional DRL framework with quantum machine learning. In this approach, each agent executes control actions using only local and nearby information within the distributed structure of the microgrid's secondary frequency control. The DDPG algorithm is developed to modify the agents' parameters to solve the DRL issue [29]. Such works indicate the trajectory toward distributed, hierarchical, and hybrid DRL systems; a combination of quantum machine learning or twin-delayed DDPG would greatly enhance convergence speed, coordination, and robustness. This is a trend reflecting the increasing complexity and sophistication needed from existing solutions to address microgrid applications in the real-world environment. Despite these promising advancements, several challenges remain unaddressed. Notably, the lack of formal guarantees on system stability and performance across diverse operational conditions, limited generalization to varying microgrid architectures, and insufficient robustness assessments in real-world environments hinder practical deployment [30–32]. Furthermore, coordination among multiple agents under communication constraints and uncertainties demands deeper investigation [33]. While emerging studies explore multi-agent frameworks and hybrid approaches integrating quantum machine learning or model predictive control, practical validation via hardware-in-the-loop (HIL) or field deployments is still rare [34]. Accordingly, further research is necessary to develop scalable, stable, and generalizable multi-agent DRL frameworks tailored for secondary voltage and frequency control in inverter-based microgrids. The persistent gaps in stability proofs, generalization, and real-world testing underline that current DRL methods, while promising, remain mostly at the simulation stage. Addressing these gaps is critical for transitioning DRL from theory to industrial practice. While prior studies have explored DRL-based control strategies for voltage or frequency regulation in microgrids, they often lack either generalization capabilities under varying dynamic conditions or suffer from centralized architectures with limited scalability. Moreover, most works focus on either voltage or frequency independently and do not investigate hierarchical multi-agent structures. In contrast to prior studies, this work introduces a distributed, hierarchical multi-agent framework that simultaneously addresses voltage and frequency control with improved coordination and adaptability under network disturbances, fluctuating loads, and dynamic DG scenarios. The proposed approach integrates a supervisory agent for harmonic monitoring and real-time PI tuning, ensuring system-wide stability while reducing transient oscillations. A comprehensive comparison of recent DRL-based secondary control methods in microgrids is summarized in Table 1.

Research Gap: Despite the substantial advancements in linear, nonlinear, adaptive, and DRL-based control methods for microgrids, several gaps still remain unaddressed. Most of the existing studies focus either on voltage or frequency regulation independently, often relying on centralized frameworks with limited scalability. Nevertheless, very few works have explored distributed and hierarchical multi-agent structures that can simultaneously regulate both voltage and frequency under time-varying load variations, DG fluctuations, and communication constraints. In addition, practical validation of such strategies through realistic scenarios, including plug-and-play operations and transient disturbances, is still lacking.

Problem Statement: The main challenge in regulating microgrids is to keep stable voltage and frequency values in the presence of fluctuating renewable energy sources and dynamic loads. Traditional control methods, such as fixed-parameter PI controllers and droop control, are often inadequate for dealing with these variations, particularly in island mode operation. There is a need for a more adaptive and intelligent control strategy that can dynamically adjust to changing conditions and ensure reliable operation of microgrids. The results of this study demonstrate that the proposed hierarchical multi-agent DRL-based control strategy effectively closes this gap by ensuring simultaneous voltage and frequency regulation, reducing transient oscillations, and enhancing power quality in a four-converter microgrid system under realistic operational scenarios.

The main contributions of this paper are given as follows:

- Development of a hierarchical multi-agent DRL framework with three specialized agents (voltage, frequency, and harmonic control) for real-time, simultaneous voltage and frequency regulation in microgrids.
- Implementation of adaptive control under load fluctuations and intermittent distributed generation, validated as a proof-of-concept on a single DG unit within a four-converter microgrid, demonstrating improved dynamic performance and adaptability under realistic operating conditions.
- Integration of DDPG to dynamically tune PI controller parameters, enabling real-time response to network disturbances and improving overall system stability.
- Introduction of a supervisory agent for harmonic monitoring, reducing transient oscillations and enhancing power quality.
- Comprehensive MATLAB/Simulink validation confirming superior dynamic performance, improved voltage and frequency stability, and effective handling of renewable energy variability.

The rest of the paper is organized as follows. In Section 2, deep reinforcement learning is presented and DDPG training and its execution process are explained. In Section 3, the proposed control strategy is introduced and the structure and equations of the multi-agent secondary control are described in detail. In Section 4, the local stability analysis

Table 1
Comparison of recent DRL-based secondary control methods in microgrids.

DRL algorithm	Control objective	Key features	Summary	Training strategy	Scalability addressed	Reference
DDPG	Optimal dispatch	Reduces control complexity; single-agent design	Offline + online learning applied to inverter control	Both	✗	[20]
DDPG	Voltage regulation	Accurate load control; lacks large-scale validation	Centralized voltage regulation framework	Offline	✗	[24]
Actor-Critic	Frequency stabilization	Adapts to renewable variability; centralized control logic	Actor-critic structure for robustness	Offline	✗	[25]
DDPG + Virtual Inertia	Frequency and stability	Enhances inertia emulation; compared with H-infinity	Virtual inertia-based DRL strategy	Both	✗	[26]
DDPG	Islanded MG control	Controls storage power; simulation-only results	Simulation-driven frequency and voltage consistency	Offline	✗	[27]
TD3 (Twin Delayed DDPG)	PID tuning	Handles RES/EVs; improves controller tuning	Complex DRL-PID hybrid for hybrid MGs	Both	✓	[28]
Multi-agent DDPG + Quantum ML	Distributed frequency control	Local-only data; integrated quantum ML; comm. constraints	Quantum-enhanced multi-agent frequency regulation	Online	✓	[29]
DQN	Voltage + frequency control	Formal stability proofs; single-agent; small-scale	DQN for stability enhancement in MGs	Offline	✗	[30]
Robust PG DRL	Adaptive control	Robust under uncertainties; limited generalization	Robust PG-based secondary controller	Both	✗	[31]
Multi-agent DDPG	Distributed secondary control	Theoretical guarantees; comm. delay issues	Multi-agent cooperative learning	Both	✓	[32]
Distributed MARL	Frequency control	Explicit comm. modeling; discusses scalability	MARL with comm. constraints for secondary control	Online	✓	[33]
Hybrid MPC + DRL	Frequency regulation	HIL-tested; complex hybrid structure	Combines model-based MPC and DRL	Both	✓	[34]
DDPG-based multi-agent hierarchical control	**Voltage and Frequency**	**Hierarchical two-layer RL agents tuning PI controllers; validated as proof-of-concept on a single DG unit within a 4-converter microgrid; objective-oriented multi-agent design**	**Distributed, scalable, real-time adaptation with supervisory harmonic control layer**	**Online**	✗	**Proposed method**

Note: For the proposed method, scalability is claimed theoretically based on the decentralized architecture; empirical validation on larger systems is left for future work.

of the voltage and frequency control loops is described. In Section 5, the simulation results are presented and the performance of the proposed multi-agent system is evaluated under scenarios. Finally, the conclusion is presented in Section 6.

3. Deep reinforcement learning

Reinforcement Learning (RL) is a machine learning technique in which an agent makes decisions in a given environment to maximize cumulative rewards over time. DRL blends RL and deep learning, allowing agents to solve complicated decision-making issues. The agent interacts with its surroundings, monitors states, takes decisions, and receives feedback in the form of rewards or penalties. The DDPG method, which combines Deep Q-Learning with Policy Gradients, is developed for continuous reinforcement learning. Fig. 1 depicts the three primary components of the DDPG process: observation, training, and execution.

To begin with, the agent engages with the environment, recording its encounters in the form of (s_t, a_t, r_t, s_{t+1}). In these tuples, s_t represents state at time t; a_t denotes action taken; r_t indicates reward; and s_{t+1} is the next state. Outcomes from those experiences go into a buffer called experience replay, which is further used in future training. This allows for more steady progress by having different events before it each time. During the training process, randomly chosen mini-batches from this buffer are employed to update the networks. The architecture of the network consists of two essential parts: an online network and a target network, where for example current decision-making actor neural networks are part of the online network. It has two sections: online network and target network. The online network consists of two different neural networks, namely, the actor that selects actions according to current state s_t, while the critic calculates the Q-value $Q(s_t, a_t)$, which is an estimate of the expected return if one takes action a_t in state s_t. On the other hand, the target network is made up of time-delayed versions of the actor and critic networks that stabilize training with constant targets for Q-value updates. In the training process, the actor network is updated using policy gradients, aiming to maximize the expected return J. Therefore, determining this gradient can be done as follows:

$$\nabla_{\theta^\mu} J \approx \mathbb{E}\left[\nabla_a Q(s, a|\theta^Q)\nabla_{\theta^\mu}\mu(s|\theta^\mu)\right] \tag{1}$$

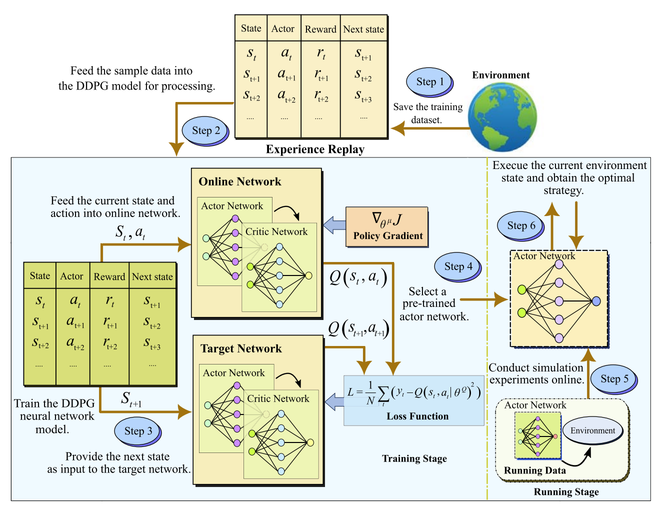

Fig. 1. DDPG training and execution process for reinforcement learning-based control.

The critic network is updated by minimizing the loss function:

$$L = \frac{1}{N}\sum_i \left(y_i - Q(s_i, a_i|\theta^Q)\right)^2 \tag{2}$$

where $y_i = r_i + \gamma Q(s_{i+1}, \mu(s_{i+1}|\theta^{\mu'})|\theta^{Q'})$. Once the networks are trained, the actor network is used to execute actions in the environment, aiming to achieve the optimal strategy. The actor network's performance is validated and refined through simulation experiments, ensuring robust and effective policy execution. The key concepts involved in DDPG are the actor network, which proposes actions based on current states; the critic network, which evaluates actions by estimating Q-values; experience replay, which stores and samples past experiences to break correlations and stabilize learning; and the target network, which provides stable target values for the critic network updates.

4. Proposed control strategy

This section introduces a novel hierarchical multi-agent secondary control framework designed to enhance the dynamic performance

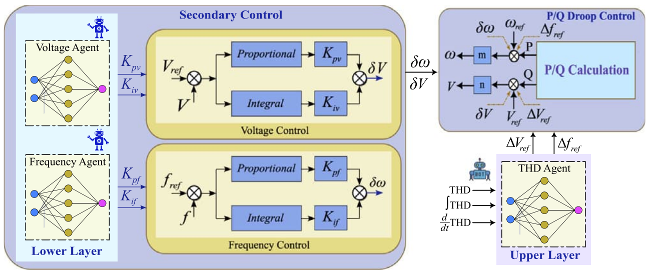

Fig. 2. General overview of the proposed hierarchical control strategy.

and transient stability of inverter-based microgrids, particularly during transitions from grid-connected to islanded modes. The proposed architecture utilizes a two-level control scheme with three dedicated reinforcement learning agents operating at different levels of the hierarchy, as shown in Figs. 2 and 3. At the lower layer, two independent reinforcement learning agents are responsible for real-time voltage and frequency regulation. Specifically, a voltage control agent and a frequency control agent are designed to dynamically adjust the gains of conventional PI controllers by observing local voltage and frequency deviations. These agents ensure localized control stability and responsiveness under varying load and generation conditions. At the upper supervisory layer, a harmonic control agent is introduced. Unlike the conventional power quality methods, this supervisory agent focuses on enhancing the overall system dynamic performance and transient stability. It monitors harmonic distortions in voltage and current waveforms across the microgrid and generates corrective modulation signals. These corrective terms are transmitted to the lower-layer voltage and frequency agents, effectively altering their control adaptation behaviors. This supervisory mechanism enables coordinated action among agents, improving both steady-state accuracy and transient response under network disturbances, nonlinear loads, and inverter-dominated scenarios. The detailed design and mathematical formulation of each agent are presented in the following subsections. The hierarchical design ensures minimal communication, with lower-layer agents relying mainly on local measurements and the upper-layer sending only essential corrective signals, resulting in low communication cost and scalable operation.

4.1. Problem formulation

Consider a microgrid comprising N distributed energy resources (DERs), M loads, and inverter-based interfaces. The objective of the proposed secondary control strategy is to maintain voltage (V) and frequency (f) within permissible thresholds while minimizing transient deviations and enhancing overall dynamic stability. Additionally, power quality improvement through harmonic mitigation is integrated via hierarchical coordination. The Multi-Agent System (MAS) framework shown in Fig. 4 facilitates distributed, scalable, and adaptive control suitable for dynamic and uncertain operating environments.

4.1.1. Voltage control agent

The voltage control agent operates at the lower layer, aiming to minimize voltage deviation at each bus. Let V_i denote the voltage at bus i and V_{ref} the reference voltage. The voltage deviation is defined as:

$$\Delta V_i = V_{\mathrm{ref}} - V_i \tag{3}$$

A PI controller is employed to regulate inverter power injection based on ΔV_i, formulated as:

$$\Delta V = K_{pV} \cdot \Delta V_i(t) + K_{iV} \int \Delta V_i(t)dt \tag{4}$$

Here, K_{pV} and K_{iV} represent proportional and integral gains, respectively. These gains are not fixed but dynamically adjusted in real-time via the DDPG algorithm based on evolving voltage deviations.

4.1.2. Frequency control agent

The frequency control agent, also positioned at the lower layer, focuses on system frequency stabilization. Let f be the measured frequency and f_{ref} the nominal reference. The frequency deviation is calculated as:

$$\Delta f_i = f_{\mathrm{ref}} - f_i \tag{5}$$

Its control action modulates active power via:

$$\Delta\omega = K_{pf} \cdot \Delta f_i(t) + K_{if} \int \Delta f_i(t)dt \tag{6}$$

Here, K_{pf} and K_{if} are adaptively tuned by a dedicated DDPG agent to enhance frequency stability.

4.1.3. Harmonic control agent

At the upper supervisory level, the harmonic control agent plays a pivotal role in enhancing system-wide dynamic stability rather than solely focusing on conventional steady-state harmonic minimization. It continuously monitors THD and generates corrective feedback signals:

$$\Delta V_{ref}, \Delta f_{ref} = f_{\mathrm{Harmonic}}(\mathrm{THD}_V, \mathrm{THD}_I) \tag{7}$$

These feedback terms, ΔV_{ref} and Δf_{ref}, are communicated to the voltage and frequency agents to fine-tune their adaptive gain mechanisms, promoting better damping of oscillations and minimizing overshoots during dynamic transitions. This supervisory influence significantly enhances the system's dynamic performance under high penetration of inverter-based resources and nonlinear load conditions. Overall, this hierarchical multi-agent approach enables intelligent, adaptive, and coordinated secondary control capable of delivering superior transient response, steady-state regulation, and power quality enhancement for modern microgrids.

Actions of the harmonic control agent (a_3). The harmonic control agent generates actions that primarily influence the other two agents or system parameters, rather than directly controlling inverters. These actions include:

- ΔV_{ref}: A numerical correction applied to the voltage reference signal (V_{ref}) for the voltage control agent. This correction updates $V_{ref}^{new} = V_0 + \Delta V_{ref}$.
- Δf_{ref}: A numerical correction applied to the frequency reference signal (f_{ref}) for the frequency control agent. This correction updates $f_{ref}^{new} = f_0 + \Delta f_{ref}$.

$$a_3(t) = \left[\Delta V_{ref}, \Delta f_{ref}\right] \tag{8}$$

Reward function for the harmonic control agent (r_3). To guide its learning, the harmonic control agent's reward function is based on minimizing harmonic distortion and maintaining voltage quality. A virtual error for THD can be defined as $e_{THD_V} = THD_V - THD_{target}$, where THD_{target} is the desired maximum THD. The reward function can be formulated as:

$$r_t = -\alpha \cdot e^2_{\mathrm{THD}_V} \tag{9}$$

Here, α is a positive weighting factor that determines the relative importance of minimizing voltage THD. This negative quadratic form ensures that the agent is heavily penalized for large deviations, encouraging actions that drive the THD towards its desired values. In this study, the weighting coefficient α in the harmonic control agent's reward function is set to 1 based on preliminary simulations to balance harmonic mitigation and overall system stability, ensuring that the agent effectively reduces harmonic distortion without adversely affecting real-time voltage and frequency regulation under varying load conditions and renewable energy fluctuations.

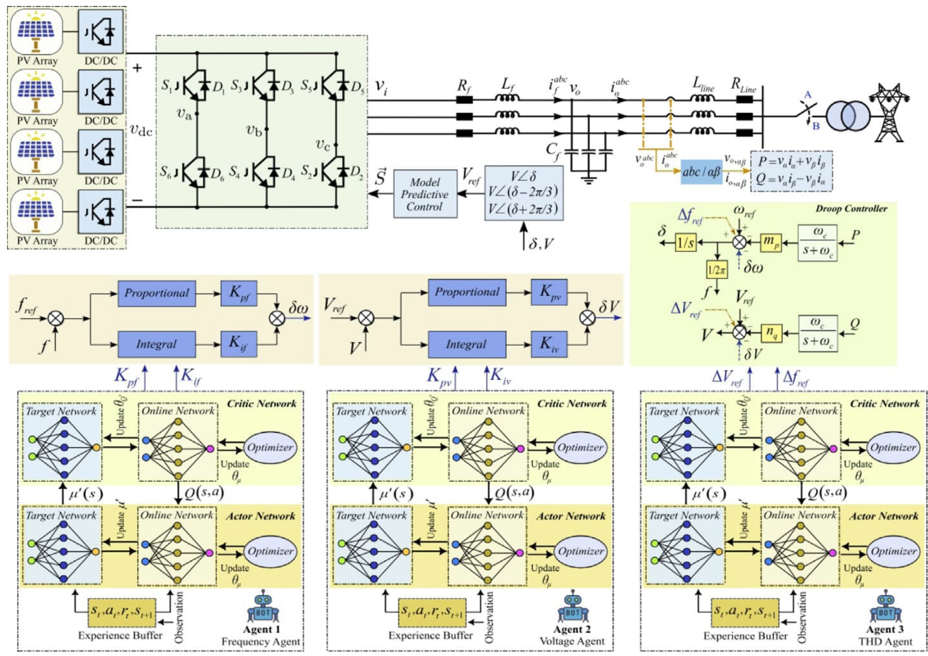

Fig. 3. Block diagram of the proposed control strategy for the multi-agent microgrid system.

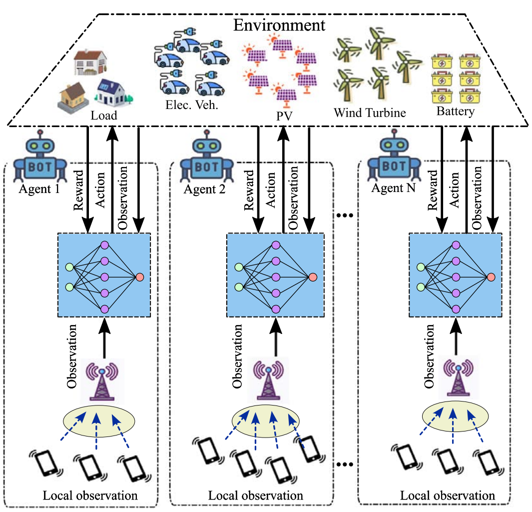

Fig. 4. Deep MARL diagram, where the agents receive local observations from their associated user equipment devices (UEs).

4.2. Integration with droop control

Droop control is a primary control method used in microgrids to share load among multiple DERs based on their respective power capacities. The droop characteristics for frequency and voltage are given by:

$$f = f_0 - k_f \cdot P \tag{10}$$

$$V = V_0 - k_v \cdot Q \tag{11}$$

where f_0 and V_0 are the nominal frequency and voltage, P and Q are the active and reactive power outputs, and k_f and k_v are the droop coefficients for frequency and voltage, respectively. The secondary control layer aims to restore the frequency and voltage to their nominal values after the initial droop response, while also considering power quality objectives.

4.2.1. Secondary voltage control

The secondary voltage control adjusts the reference voltage V_{ref} dynamically to compensate for voltage deviations. With the influence of the harmonic control agent, the updated reference voltage becomes $V_{ref}^{new} = V_0 + \Delta V_{ref}$. The combined control law with droop and secondary control can be expressed as:

$$\begin{aligned} V_i(t) = {} & V_0 - k_v \cdot Q + K_{pV} \cdot \Delta V_i(t) \\ & + K_{iV} \int \Delta V_i(t)\, dt + \Delta V_{ref} \end{aligned} \tag{12}$$

4.2.2. Secondary frequency control

Similarly, the secondary frequency control adjusts the reference frequency f_{ref} to compensate for frequency deviations. With the influence of the harmonic control agent, the updated reference frequency becomes $f_{ref}^{new} = f_0 + \Delta f_{ref}$. The combined control law with droop and secondary control is:

$$\begin{aligned} f_i(t) = {} & f_0 - k_f \cdot P + K_{pf} \cdot \Delta f_i(t) \\ & + K_{if} \int \Delta f_i(t)\, dt + \Delta f_{ref} \end{aligned} \tag{13}$$

4.3. Multi-agent coordination

The voltage, frequency, and harmonic control agents operate within a hierarchical framework, where each agent interacts indirectly through the environment to achieve overall system stability. The agents achieve coordination through indirect interaction: the harmonic agent modifies the reference signals ($\Delta V_{ref}, \Delta f_{ref}$) that influence the state space of the voltage and frequency agents. This structure allows for coordinated behavior without the complexity of shared learning objectives. This decentralized approach enhances the system's resilience to disturbances and faults. The interaction and coordination among agents can be modeled as a multi-agent reinforcement learning problem, where agents share relevant state information, such as voltage levels, frequency measurements, THD metrics, and control actions, to optimize their policies jointly. Agents in the MAS framework work semi-independently towards the common objective of ensuring system stability and achieving voltage and frequency restoration to nominal levels. These agents can be conceptualized as individual decision makers who in turn sense the local environment and act upon it in order to manipulate their own variables. This particular arrangement permits a scalable and adaptable control structure which is capable of meeting different conditions and configurations of the microgrid. The harmonic control agent, in particular, coordinates with the voltage and frequency agents by adjusting their reference signals ($\Delta V_{ref}, \Delta f_{ref}$). This allows the harmonic agent, without directly controlling power output, to influence the other two agents in a way that enhances dynamic performance and accelerates voltage and frequency restoration following disturbances, thereby contributing to overall system stability.

4.4. RL-based control problem formulation using MDP

Each agent's control problem can be expressed as a Markov Decision Process (MDP), where they learn the best possible policy π that minimizes the variations and maximizes system performance. The components of the MDP for an agent i are defined as follows:

1. **State** s_i: The voltage-controlling agent has its current voltage V_i as the state, while the frequency-controlling agent uses its current frequency f as the state. The harmonic control agent's state includes THD metrics, RMS voltage/current, and the errors/actions of the other two agents. Other general state variables can consist of voltage and frequency deviations' historical data, current load demand, and generation levels.

$$s_t = \Big[\, e_f(t),\ e_v(t),\ \dot{e}_f(t),\ \dot{e}_v(t),\ \int e_f(t)\,dt,\ \int e_v(t)\,dt,\ f(t),\ V(t),\ P(t),\ Q(t),\ t\,\Big], \quad s_t \in S \tag{14}$$

Remark: The state variables in the MDP formulation can be classified into two categories. The history-independent states include the instantaneous frequency error $e_f(t)$, voltage error $e_v(t)$, their derivatives $\dot{e}_f(t)$ and $\dot{e}_v(t)$, the measured frequency $f(t)$ and voltage $V(t)$, active power $P(t)$, reactive power $Q(t)$, and current time t. These variables depend only on current measurements and not on past actions. The history-dependent states are the integral terms $\int e_f(t)dt$ and $\int e_v(t)dt$, which accumulate past errors and thus depend on the agent's previous actions through the closed-loop control history. The inclusion of these integral terms is essential for PI control to eliminate steady-state error. The augmented state vector (including integrals) satisfies the Markov property because it fully captures all information necessary to predict future system evolution.

2. **Action** a_i: The action represents the control input. For the voltage and frequency agents, this involves adjusting the PI controller gains K_{pV}, K_{iV}, K_{pf}, and K_{if}. For the harmonic control agent, actions include ΔV_{ref}, Δf_{ref}. The action space is continuous, reflecting the continuous nature of control inputs.

$$a_t = [\Delta V_{\mathrm{ref}}, \Delta f_{\mathrm{ref}}, K_{pf}, K_{if}, K_{pv}, K_{iv}], \quad a_t \in A(s_t) \tag{15}$$

3. **Reward** r_i: The reward function evaluates the performance of the control action.
   - For the voltage control agent, the reward function could be:

   $$r_v(t) = -(\Delta V_i(t)^2) + 10 \tag{16}$$

   - Similarly, for the frequency control agent, the reward function can be:

   $$r_f(t) = -(\Delta f_i(t)^2) + 10 \tag{17}$$

   - For the harmonic control agent, the reward function is designed to penalize harmonic distortion:

   $$r_h(t) = -(\mathrm{THD}_V(t) - \mathrm{THD}_{\mathrm{target}})^2 \tag{18}$$

   The negative sign ensures that the agent is penalized for deviations, promoting actions that minimize these deviations.
   Remark: The value of 10 is arbitrarily chosen as a fixed positive reward and can be replaced by any positive number, depending on system requirements and design criteria, to encourage maintaining optimal performance. This term helps ensure cumulative rewards are often positive, aiding stable learning.

4. **Policy** π_i: The policy is the mapping from states to actions, which the agent learns using the DDPG algorithm. The policy aims to maximize the cumulative reward over time, guiding the agent to perform actions that lead to better control performance across all objectives.

4.5. Comparative analysis of control agents

To provide a clear overview of the proposed multi-agent control strategy, Table 2 summarizes the key characteristics, objectives, inputs, outputs, and reward functions for each of the three DRL agents. This table highlights their distinct roles and how they collectively contribute to the overall stability and power quality of the microgrid.
Remark1: Each of the three DRL agents in the proposed hierarchical framework employs an individual reward function rather than a shared common reward. This design choice prevents conflicting objectives and allows each agent to specialize in its dedicated control task. Although the reward functions are independent, the agents interact indirectly through the environment: the harmonic agent modifies the reference signals ($\Delta V_{\mathrm{ref}}, \Delta f_{\mathrm{ref}}$) that influence the state space of the voltage and frequency agents, thereby achieving cooperative behavior without a shared reward.
Communication Structure: The three agents communicate in a hierarchical manner as follows: The voltage and frequency agents (lower layer) send their local state information specifically voltage deviation (ΔV), frequency deviation (Δf), and their current PI gains ($K_{pv}, K_{iv}, K_{pf}, K_{if}$) to the harmonic control agent at the supervisory layer. The harmonic agent (supervisory layer) processes this information along with system-wide THD measurements and computes corrective reference signals ($\Delta V_{\mathrm{ref}}, \Delta f_{\mathrm{ref}}$). These signals are then transmitted back to the voltage and frequency agents, which update their reference setpoints as $V_{ref}^{new} = V_0 + \Delta V_{ref}$ and $f_{ref}^{new} = f_0 + \Delta f_{ref}$.
No direct communication occurs between the voltage and frequency agents; their coordination is achieved indirectly through the shared environment and the supervisory harmonic agent. This minimal communication architecture reduces latency and enhances scalability.

4.6. Implementation with DDPG

In this study, each agent in the hierarchical multi-agent framework is trained using the Deep Deterministic Policy Gradient (DDPG) algorithm, which combines the actor–critic structure with deterministic policy gradients to handle continuous action spaces. The algorithm enables each agent to learn optimal control actions for voltage and frequency regulation based on local state observations and received rewards. Unlike conventional DDPG implementations, the proposed approach introduces task-specific adaptations for hierarchical coordination. The lower-layer agents employ DDPG to dynamically tune the PI controller parameters responsible for local voltage and frequency regulation, while the upper-layer supervisory agent learns to coordinate harmonic compensation and inter-agent synchronization. The learning process incorporates experience replay and target networks to ensure training stability, and the reward functions are explicitly designed to minimize voltage/frequency deviations and harmonic distortions. The overall training objective is to achieve coordinated adaptive control among all agents within the hierarchical framework through sequential training, enabling real-time system stability under fluctuating load and renewable generation conditions. It is important to note that while this work presents a hierarchical multi-agent design with three specialized

Table 2
Comparative analysis of reinforcement learning-based control agents in microgrids under the proposed approach.

Control attribute	Voltage control agent	Frequency control agent	Harmonic control agent
Main objective	Minimize voltage deviation $\Delta V = V_{ref} - V_i$	Minimize frequency deviation $\Delta f = f_{ref} - f_i$	Reduce THD of voltage and current signals
Observations	$V_i, V_{ref}, \Delta V, \int \Delta V\,dt, \frac{d\Delta V}{dt}, Q, t$	$f_i, f_{ref}, \Delta f, \int \Delta f\,dt, \frac{d\Delta f}{dt}, P, t$	$THD_V, V_{rms}, I_{rms}, e_f, e_v, a_f, a_v$
Error metric	$\Delta V_i = V_{ref} - V_i$	$\Delta f_i = f_{ref} - f_i$	$e_{THD_V}, e_{V_{rms}}$
Actions	K_{pV}, K_{iV}	K_{pf}, K_{if}	$\Delta V_{ref}, \Delta f_{ref}$
Reward function	$r_v(t) = -\Delta V_i(t)^2 + 10$	$r_f(t) = -\Delta f_i(t)^2 + 10$	$r_i(t) = -e^2_{THD_V}$
RL algorithm	DDPG	DDPG	DDPG
Control scope	Secondary, local voltage regulation	Secondary, local frequency regulation	Secondary, system-wide power quality coordination
Mode of impact	Direct inverter output adjustment	Direct DER active power control	Indirect tuning of other agents' references
Key innovation	Online PI gain tuning via deep RL	Online PI gain tuning via deep RL	Coordinated multi-agent THD mitigation strategy

agents, the current implementation utilizes a sequential training procedure to ensure training stability. The sequential approach mitigates the non-stationarity problem commonly encountered in multi-agent learning. The more general problem of true cooperative multi-agent learning with joint training (e.g., using MADDPG, MAPPO, or centralized training with decentralized execution) is beyond the scope of this study and is left for future research. There were various considerations taken into account in selecting DDPG as the major methodology. Even though SAC and PPO, other deep reinforcement learning agents, were on the table, DDPG was chosen due to the following reasons:

- DDPG is the most appropriate continuous-action-space manager, as it has shown great performance in tasks requiring smooth and precise adjustments. This property is crucial for real-time tuning of PI controller gains in microgrids, ensuring stability and accuracy in both voltage and frequency regulation.
- Great consideration was given to DDPG's stability and ease of implementation. Its relatively simple architecture compared to SAC and PPO made it easier to implement, debug, and integrate into the multi-agent hierarchical control framework, allowing us to focus on problem-specific challenges rather than technical complexities.
- DDPG supports both offline training and online execution, enabling the system to adapt to dynamic changes in load, distributed generation, and network disturbances in real-time.
- Its actor-critic structure allows each agent to efficiently learn local policies while coordinating with other agents, which is essential for the hierarchical multi-agent secondary control proposed in this work.
- DDPG has demonstrated reliable convergence and performance under uncertain and stochastic conditions, making it particularly suitable for islanded microgrids with variable renewable energy sources and fluctuating loads.
- The DDPG algorithm employs an actor-critic architecture. Both the actor and critic networks consist of two hidden layers with 500 neurons each, using ReLU activation functions. The actor network outputs deterministic actions (PI gains) using a tanh activation function scaled to the action bounds. The action bounds for each agent are: frequency agent ($K_p \in [0.01, 6]$, $K_i \in [0.01, 0.8]$), voltage agent ($K_p \in [0.01, 0.9]$, $K_i \in [0.01, 0.9]$), and harmonic agent ($K_p \in [0.01, 0.9]$, $K_i \in [0.01, 0.9]$). The complete training hyperparameters are summarized in Table 3.

5. Data-driven local stability analysis of frequency and voltage control loops

In nonlinear power electronic systems, especially grid-connected inverters in microgrids, frequency and voltage regulation are important for operational stability. Because of the intermittent nature of renewable energy and the time-varying load, the control problem becomes more and more complex. Despite the nonlinear dynamics exhibited by the overall system, local small-signal behavior around steady-state operating points can be well-characterized by means of a linear approximation. Such an approach is conventionally accepted in power electronics to conduct stability analyses that remain tractable. Data-driven system identification provides a powerful mechanism to model the frequency and voltage loops based on experimentally collected input–output datasets. An ARX model is applied here to capture local linear dynamics and, therefore, allows the use of well-known classical control tools such as root locus and Routh–Hurwitz criteria for stability analysis. This section presents the mathematical formulation for constructing transfer functions from data, deriving closed-loop characteristic equations, and defining the stability criteria under reinforcement learning (RL)-based PI controllers.

5.1. Frequency loop: Data-driven modeling

Consider the input–output dataset collected for the frequency control loop:

$$\mathcal{D}_f = \{e_f(t), f(t)\}_{t=1}^{N} \tag{19}$$

where $e_f(t)$ denotes the frequency deviation (input) and $f(t)$ is the measured system frequency (output). The system is modeled using an ARX structure:

$$A_f(q^{-1})f(t) = B_f(q^{-1})e_f(t) + e_f(t) \tag{20}$$

with

$$\begin{aligned} A_f(q^{-1}) &= 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \\ B_f(q^{-1}) &= b_1 q^{-1} + \cdots + b_m q^{-m} \end{aligned} \tag{21}$$

Applying a bilinear transformation to convert the identified discrete-time model to the continuous-time domain yields the transfer function:

$$G_f(s) = \frac{\beta_{f0}s^m + \beta_{f1}s^{m-1} + \cdots + \beta_{fm}}{s^n + \alpha_{f1}s^{n-1} + \cdots + \alpha_{fn}} \tag{22}$$

5.2. Voltage loop: Data-driven modeling

Similarly, the input–output dataset for the voltage loop is given by:

$$\mathcal{D}_v = \{e_v(t), v(t)\}_{t=1}^{N} \tag{23}$$

where $e_v(t)$ is the voltage error input and $v(t)$ represents the output voltage. The ARX model is formulated as:

$$A_v(q^{-1})v(t) = B_v(q^{-1})e_v(t) + e_v(t) \tag{24}$$

which leads to the continuous-time transfer function:

$$G_v(s) = \frac{\beta_{v0}s^m + \beta_{v1}s^{m-1} + \cdots + \beta_{vm}}{s^n + \alpha_{v1}s^{n-1} + \cdots + \alpha_{vn}} \tag{25}$$

5.3. PI controller and closed-loop characteristic equations

The PI controllers are implemented for both loops as follows:

$$C_f(s) = K_{pf} + \frac{K_{if}}{s} = \frac{K_{pf}s + K_{if}}{s} \tag{26}$$

$$C_v(s) = K_{pv} + \frac{K_{iv}}{s} = \frac{K_{pv}s + K_{iv}}{s} \tag{27}$$

The characteristic equation of the closed-loop frequency control system is obtained by:

$$D_f(s) = s\left(s^n + \sum_{i=1}^{n}\alpha_{fi}s^{n-i}\right) + (K_{pf}s + K_{if})\left(\sum_{j=0}^{m}\beta_{fj}s^{m-j}\right) = 0 \tag{28}$$

Similarly, for the voltage loop:

$$D_v(s) = s\left(s^n + \sum_{i=1}^{n}\alpha_{vi}s^{n-i}\right) + (K_{pv}s + K_{iv})\left(\sum_{j=0}^{m}\beta_{vj}s^{m-j}\right) = 0 \tag{29}$$

5.4. Root locus-based stability assessment

The characteristic polynomials can be expanded as:

$$D_f(s) = \sum_{k=0}^{n+m+1} d_k^{(f)}s^k, \quad D_v(s) = \sum_{k=0}^{n+m+1} d_k^{(v)}s^k \tag{30}$$

Stability is ensured if all roots lie strictly in the open left-half complex plane:

$$\Re(s_i) < 0, \quad \forall s_i \in \text{Roots of } D(s) \tag{31}$$

Applying the Routh–Hurwitz criterion to the characteristic equations, the sufficient condition for stability is:

$$d_k(K_p, K_i) > 0, \quad \forall k \in \{0, 1, \dots, n+m+1\} \tag{32}$$

which implies all elements of the first column in the Routh table must be strictly positive for both frequency and voltage loops. In this framework, RL policies output the PI controller gains (K_{pf}, K_{if}) and (K_{pv}, K_{iv}). To guarantee closed-loop stability, the admissible action spaces for RL agents are constrained by:

$$S_f = \{(K_{pf}, K_{if}) : D_f(s) \text{ is stable}\} \tag{33}$$

$$S_v = \{(K_{pv}, K_{iv}) : D_v(s) \text{ is stable}\} \tag{34}$$

This constraint formulation ensures that RL exploration and exploitation remain confined within the safe operational boundaries determined by the local linearized stability conditions derived from data.
Remark on DRL Stability: The local stability analysis presented in this section provides necessary conditions for stability of the closed-loop system when the PI gains are fixed. However, the proposed DRL controller continuously adapts the PI gains in real-time based on the learned policy. Therefore, rigorous stability guarantees for the closed-loop system with adaptive DRL gains require additional theoretical tools such as Lyapunov-based verification or constrained action spaces that restrict the agent's exploration to provably stable regions. This limitation is acknowledged, and future work will address formal stability certification for the DRL-based controller.

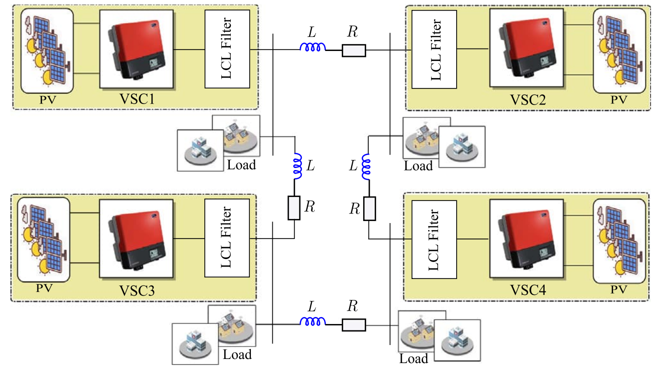

Fig. 5. Schematic diagram of the interconnected VSC-based system with LCL filters and loads.

Table 3
Microgrid, RL training, and droop parameters used in simulations.

System configuration parameters	
Nominal voltage	$V_n = 380$ V
Nominal frequency	$f_n = 50$ Hz
Line impedance	$R_l = 0.25\ \Omega$, $X_l = 1$ mH
Filter inductance	$L_f = 1.4$ mH
Capacitance filter	$C_f = 50\ \mu$F
Filter resistor	$R_f = 0.1\ \Omega$
Sampling time	$T_s = 50\ \mu$s
Load1	$P = 10$ kW, $Q = 4$ kVar
Load2	$P = 15$ kW, $Q = 5$ kVar
Load3	$P = 8$ kW, $Q = 3$ kVar
Load4	$P = 15$ kW, $Q = 7$ kVar
Droop coefficients	
DG1	$m_p = 4.5 \times 10^{-5}$, $n_p = 3.4 \times 10^{-4}$
DG2	$m_p = 3.6 \times 10^{-5}$, $n_p = 5.5 \times 10^{-4}$
DG3	$m_p = 4.1 \times 10^{-5}$, $n_p = 4.1 \times 10^{-4}$
DG4	$m_p = 3.4 \times 10^{-5}$, $n_p = 6.7 \times 10^{-4}$
RL training and reward parameters	
Actor learning rate	1×10^{-4}
Critic learning rate	1×10^{-4}
Replay buffer size	1×10^{6} transitions
Mini-batch size	64
Target network update rate	$\tau = 0.001$
Discount factor	$\gamma = 0.99$
Exploration noise variance	0.05 (p.u.)^2
Noise variance decay rate	1×10^{-6} per episode
Gradient threshold	1.0
Reward weight (THD)	$\alpha = 1$

6. Simulation results

To evaluate the performance and effectiveness of the proposed multi-agent secondary control strategy, a comprehensive case study is conducted on a microgrid composed of four DG units (see Fig. 5). The system is subjected to dynamic disturbances, including DG unit switching and load variation events. The parameters of the test systems are summarized in Table 3.
Computational considerations: The proposed hierarchical multi-agent framework employs three RL agents (voltage, frequency, and harmonic control) operating on a single DG unit (DG1). This design is objective-oriented, where each agent is responsible for a specific control objective (voltage regulation, frequency regulation, and harmonic mitigation) within a hierarchical structure, rather than being distributed across multiple DG units. This is a key distinction from conventional multi-agent systems where each DG has its own set of

agents. While the framework is designed to be inherently scalable to multiple DG units, implementing three DRL agents on all four DGs would result in twelve agents running simultaneously, exceeding the computational capacity of the available simulation platform. In preliminary tests, simulation time increased by approximately 300% when attempting to apply RL agents to multiple DGs. Therefore, the proposed method is applied to DG1 as a proof-of-concept, and its performance is compared against conventional PI and fuzzy controllers on the same DG. This setup ensures a fair and clear performance comparison between RL-based and traditional PI-based control strategies. However, the proposed hierarchical multi-agent framework is inherently scalable and adaptable to microgrids with different topologies, higher numbers of DGs, or hybrid AC/DC systems. With sufficient computational resources, the framework can be extended to multiple DGs without structural modifications. Full distributed implementation across all DGs is left for future work with more powerful hardware.

In this section, two distinct scenarios are conceived and executed for a full evaluation of the effectiveness of the proposed RL-based secondary control:

(i) Scenario 1 – DG Switching Events

This test involves the disconnection and reconnection of DGs in islanded mode:

- **Stage 1 (0.5–1 s):** The microgrid is islanded at $t = 0.5$ s.
- **Stage 2 (1–1.5 s):** The secondary controller is activated at $t = 1$ s.
- **Stage 3 (1.5–2 s):** DG3 is disconnected at $t = 1.5$ s.
- **Stage 4 (2–2.5 s):** DG4 is disconnected at $t = 2$ s.
- **Stage 5 (2.5–3 s):** DG3 is reconnected at $t = 2.5$ s.
- **Stage 6 (3–3.5 s):** DG4 is reconnected at $t = 3$ s.

(ii) Scenario 2 – Load Variation Events

In this test, the system undergoes load disturbances while operating in islanded mode to assess disturbance rejection performance:

- **Stage 1 (0.5–1 s):** The system is islanded at $t = 0.5$ s.
- **Stage 2 (1–1.5 s):** The secondary controller is activated at $t = 1$ s.
- **Stage 3 (1.5–2 s):** A load is added to DG1 at $t = 1.5$ s.
- **Stage 4 (2–2.5 s):** A load is added to DG4 at $t = 2$ s.
- **Stage 5 (2.5–3 s):** The load is disconnected from DG1 at $t = 2.5$ s.
- **Stage 6 (3–3.5 s):** The load is disconnected from DG4 at $t = 3$ s.

This dual-scenario testing allows for a fair and comprehensive evaluation of the proposed control strategy in both generator-switching and load-disturbance conditions. **Training scheme:** The three agents are trained sequentially in the following order to mitigate non-stationarity:

- **Step 1:** The frequency control agent is trained alone for 30 episodes until its reward converges to a stable value (Fig. 13b). Once converged, the agent maintains its optimal policy naturally.
- **Step 2:** The voltage control agent is trained alone for 30 episodes while the frequency agent's policy remains stable due to its convergence (Fig. 13c).
- **Step 3:** The harmonic control agent is trained for 30 episodes while both frequency and voltage agents' policies are stable due to their prior convergence (Fig. 13a).

This sequential training approach provides a stable environment for each agent to converge to its optimal policy. Once an agent converges, its policy remains stable, providing a stationary environment for the subsequent agent's training. While this sequential approach is practical for hierarchical control, true cooperative multi-agent learning with joint training (e.g., using MADDPG or MAPPO) is left for future work. Fig. 6 shows how the four-DG system operates under scenario (i) with different control strategies. The conventional PI control method shows that the frequencies of DG1 to DG4 are 49.9444 Hz, 49.9459 Hz, 49.9457 Hz, and 49.9467 Hz, respectively. In contrast, the fuzzy control method demonstrates slightly improved performance compared to the conventional PI controller, with frequencies 49.9450 Hz, 49.9502 Hz, 49.950 Hz, and 49.955 Hz for DG1 to DG4, respectively. However, under the proposed control strategy, with the application of the proposed approach to DG1, the frequency of DG1 significantly improves and effectively returns to its nominal value, reaching about 49.998 Hz. Moreover, at 1.5 s and 2 s, when DGs experience switching events, the system exhibits substantially better frequency deviation (Δf) performance compared to the conventional method. Specifically, the Δf values are 0.0096 Hz and 0.022 Hz for the proposed approach, 0.02 Hz and 0.0468 Hz for the conventional PI controller, and 0.019 Hz and 0.0460 Hz for the fuzzy control, respectively. In terms of the voltage amplitude waveform, after the activation of the secondary controller, the voltage amplitudes of the DGs under the conventional PI control method reach 309.06 V, 308.925 V, 308.83 V, and 308.743 V for DG1 to DG4, respectively. Under the fuzzy control method, these values slightly improve to 309.36 V, 309.20 V, 308.945 V, and 308.697 V, respectively. However, with the application of the proposed control strategy to DG1, the voltage amplitude of DG1 reaches its reference value of 311 V. The voltage amplitudes of DG2, DG3, and DG4 reach 308.807 V, 308.725 V, and 308.55 V, respectively. Therefore, the proposed control strategy outperforms both conventional PI and fuzzy control in terms of frequency stability and voltage regulation. Fig. 7 illustrates the dynamic performance of Scenario (ii) under different secondary control strategies. In this scenario, by applying the proposed control strategy to DG1, it is observed that the dynamic response of DG1 shows superior performance compared to the other control methods. Specifically, during the two time intervals at 1.5 s and 2 s, the frequency deviation (Δf) in the conventional PI controller reaches 0.0115 Hz and 0.0132 Hz, respectively. For the fuzzy control method, these values are 0.0115 Hz and 0.0130 Hz, while under the proposed strategy, they are significantly reduced to 0.0059 Hz and 0.0068 Hz, respectively, indicating an optimal frequency regulation performance. At 1.5 s and 2 s, when DG1 and DG4 are integrated into the system, respectively, the superiority of the proposed strategy is clearly demonstrated. In these moments, the frequency deviation under the proposed approach is limited to only 0.0059 Hz and 0.0068 Hz, whereas the conventional PI and fuzzy control methods exhibit similar but higher values of 0.0115 Hz and 0.0132 Hz, respectively. These results demonstrate the substantial improvement in transient performance and the capability of the proposed strategy to effectively reduce frequency deviation during transient conditions. Fig. 8 shows frequency and voltage deviations for Scenarios (i) and (ii) under the proposed approach applied to DG1. The figure demonstrates the system behavior during different stages, including disconnection from the grid, activation of the secondary controller, and the tripping and reconnection of DG3 and DG4. It can be observed that the proposed strategy ensures stable frequency and voltage regulation, effectively minimizing deviations during these transient events. Fig. 9 illustrates the performance of the studied system under the proposed control strategy for various values of K_p, representing the gain of the harmonic agent. This agent plays a crucial role in damping oscillations, and its appropriate tuning through K_p contributes to enhanced system stability. The figure presents two scenarios, (i) and (ii). In each scenario, the upper plots depict the frequency response, while the lower plots illustrate the voltage response over time. For a K_p value of 0.3, the frequency and voltage plots exhibit noticeable oscillations in response to the events occurring in the system. These oscillations indicate a less stable response with reduced damping performance compared to the other K_p values. For a K_p value of 0.2, by carefully adjusting the gain to 0.2, the oscillations in both frequency and voltage are visibly reduced. This observation suggests that this value of K_p offers improved damping characteristics and a more stable dynamic response to system disturbances. For a K_p value of 0.1, further reduction of the gain to 0.1 results in the most optimal performance among the tested values. The frequency and voltage responses demonstrate minimal oscillations and exhibit a faster settling time after each event. This indicates that a K_p value of 0.1

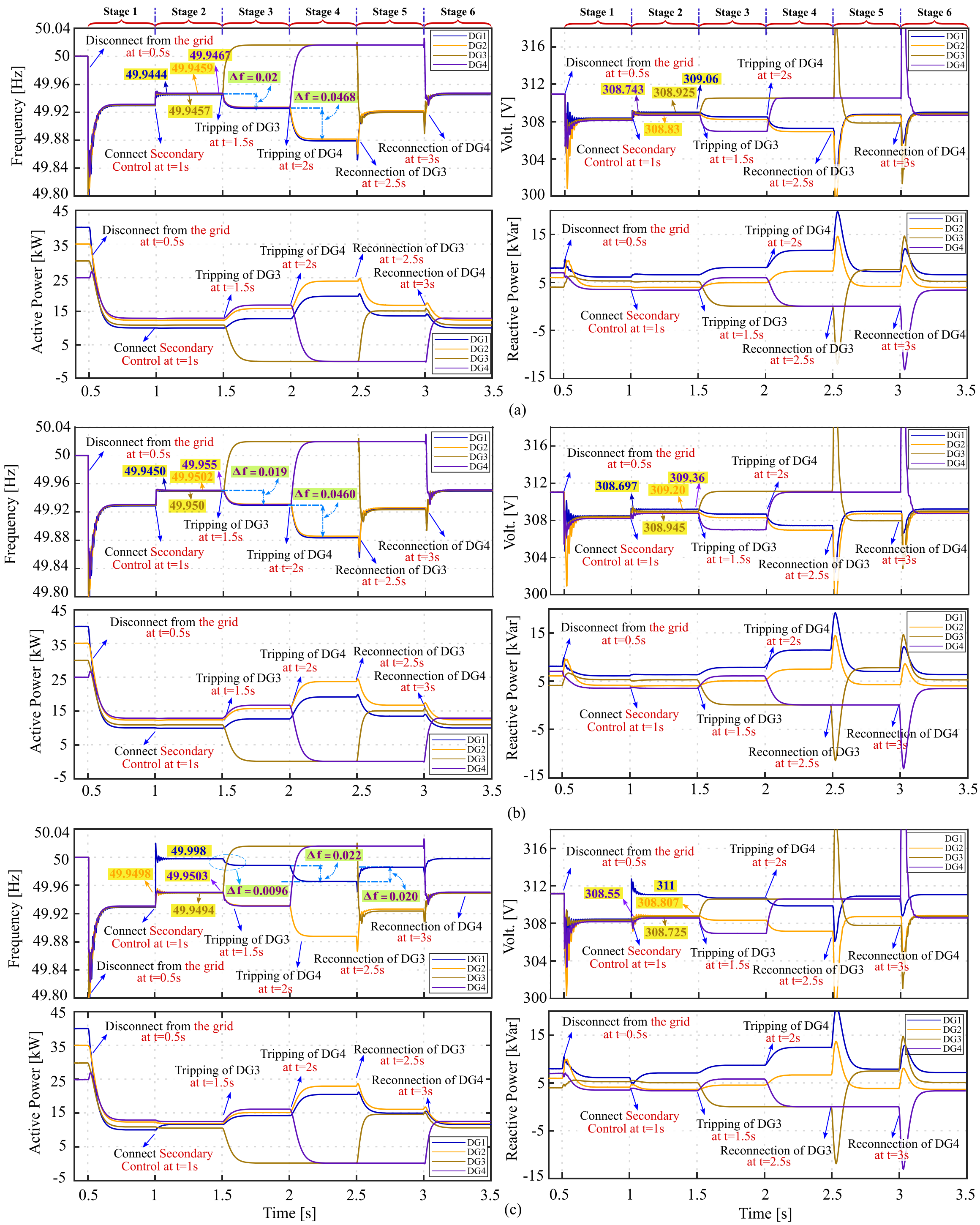

Fig. 6. Dynamic performance of the microgrid under Scenario (i) considering different secondary control strategies: (a) Conventional PI controller, (b) Fuzzy controller, and (c) The proposed controller.

provides the best overall stability and damping performance under the studied operating conditions and disturbance scenarios. Fig. 10 presents the comparison of the THD under the proposed control strategy for DG1 with three different proportional gain values K_p while keeping $K_i = 0.08$ constant. The results indicate that for $K_p = 0.1$, the THD is minimized at 0.68%. Increasing K_p to 0.2 leads to a THD of 2.03%, and further increasing K_p to 0.3 results in a THD of 2.07%. These outcomes reveal that increasing K_p causes a deterioration in harmonic performance, with higher THD values. Compared to $K_p = 0.2$, the THD at $K_p = 0.1$ is reduced by approximately 66.5%, and compared to $K_p = 0.3$, the improvement reaches around 67.1%. Therefore, selecting a lower K_p value leads to a significant improvement in power quality and harmonic distortion reduction. As shown in Tables 4 and 5, the proposed MA-DDPG controller outperforms both conventional PI and fuzzy controllers under different scenarios. For DG switching events (Table 4), the MA-DDPG controller achieves a voltage settling time of 0.13 ± 0.01 s and a frequency settling time of 0.10 ± 0.01 s, which are approximately 35% and 55% faster than the conventional PI controller, respectively. Total harmonic distortion (THD) is reduced to 0.68 ± 0.02%, and the maximum frequency deviation at 1.5 s and 2 s

Table 4
Quantitative comparison of control strategies for DG1 (DG switching events).

Method	Voltage settling (s)	Frequency settling (s)	THD (%)	Max. frequency deviation (Hz)	
				At 1.5 s	At 2 s
Conventional PI	0.20	0.22	3.07	0.020	0.0468
Fuzzy control	0.18	0.17	3.03	0.019	0.0460
Conventional DRL (DDPG)	0.25	0.16	2.52	0.010	0.026
Proposed MA-DDPG (Mean ± Std. Dev. over 5 runs)	**0.13** ± **0.01**	**0.10** ± **0.01**	**0.68** ± **0.02**	**0.0096** ± **0.001**	**0.022** ± **0.001**

Table 5
Quantitative comparison of control strategies for DG1 (Load variation events).

Method	Voltage settling (s)	Frequency settling (s)	THD (%)	Max. frequency deviation (Hz)	
				At 1.5 s	At 2 s
Conventional PI	0.25	0.22	3.42	0.0115	0.0132
Fuzzy control	0.22	0.20	3.31	0.0115	0.0130
Conventional DRL (DDPG)	0.25	0.18	2.50	0.0068	0.0077
Proposed MA-DDPG (Mean ± Std. Dev. over 5 runs)	**0.13** ± **0.01**	**0.10** ± **0.01**	**0.70** ± **0.02**	**0.0059** ± **0.001**	**0.0068** ± **0.001**

decreases to 0.0096 ± 0.001 Hz and 0.022 ± 0.001 Hz, representing more than 50% improvement compared to PI control. Under load variation events (Table 5), similar improvements are observed: the MA-DDPG controller reduces voltage and frequency settling times to 0.13 ± 0.01 s and 0.10 ± 0.01 s, respectively, while THD decreases to 0.70 ± 0.02%, and maximum frequency deviations are reduced to 0.0059 ± 0.001 Hz and 0.0068 ± 0.001 Hz at 1.5 s and 2 s, roughly halving the deviations seen in conventional controllers. These results quantitatively confirm the superior dynamic performance, harmonic mitigation, and transient stability of the proposed multi-agent controller across multiple operational scenarios, as evaluated over five independent runs.

Throughout this section, the term Conventional DRL refers to a standard single-agent DDPG controller that is distinct from our proposed multi-agent framework. This baseline uses a single DDPG agent that directly outputs frequency control actions (PI gains) on DG1, trained with the reward function $r = -(\Delta f)^2 + 10$. It does not include separate agents for voltage regulation or harmonic mitigation. This baseline is included to provide a fair comparison against conventional DRL approaches and to highlight the advantages of the proposed hierarchical multi-agent design.

Fig. 11 shows the root locus analysis of the loops controlling frequency and voltage for the two different scenarios of operation, operating with varying proportional gains of $K_p = 0.1, 0.2, 0.3$ and a fixed integral gain of $K_i = 0.08$. The analysis is undertaken to evaluate the stability margins and damping behavior of the concrete control structure. For frequency control, at $K_p = 0.1$, the poles are located in the left-half complex plane with a considerable distance from the imaginary axis, indicating strong stability and adequate damping. As K_p increases toward 0.2, poles tend to approach the imaginary axis increasingly, so there is less damping and the transient response is weak. At $K_p = 0.3$, some of the complex-conjugate poles are at least nearing or crossing the imaginary axis. This means the periodical behavior sets in and the system could potentially be dynamically unstable. A similar trend is observed in voltage control. At lower K_p values, the system exhibits stable dynamics with proper damping. As K_p increases, the poles gradually move toward the imaginary axis, indicating diminished damping and increased oscillatory tendencies. Fig. 12 presents the performance assessment of the proposed control strategy in two different scenarios. The results show the output voltage and the RMS voltage (Vrms) of DG1 and DG4 across different stages of operation. The output voltage waveforms of the DGs, highlight the voltage stability and quality of the control strategy. The RMS voltage values indicate how well the proposed strategy maintains voltage levels close to the desired reference throughout the different stages. Fig. 13 illustrates the episode reward trends during the training process for three different control agents, including the THD reduction agent, the frequency control agent, and the voltage control agent. This figure demonstrates the performance of each agent over consecutive training episodes and highlights the improvement in stability and control quality. For the THD agent, the episode rewards in the initial episodes are within a negative range approximately -90, indicating poor performance in reducing harmonic distortion at the beginning of the training. As training progresses, the reward rapidly increases and reaches a positive range (around 10) after approximately 7 to 8 episodes, eventually converging to a stable value. This behavior indicates the success of the THD agent in learning effective control policies for harmonic distortion reduction. The frequency agent initially records much lower rewards around -1800, reflecting greater challenges in frequency regulation. During the early episodes, significant fluctuations in rewards are observed; however, after around 10 to 12 episodes, the rewards gradually increase and eventually stabilize at a positive value (approximately 250), indicating the gradual improvement and success of the agent in achieving stable frequency control. For the voltage agent, the initial rewards are also negative (around -800), but this agent exhibits the fastest improvement trend. Within the first 5 to 7 episodes, the rewards quickly increase to very high positive values (around 1000) and then remain stable at that level, demonstrating the high effectiveness of the voltage agent in fast learning and achieving stable voltage regulation in the system. To further validate the choice of DDPG, three state-of-the-art DRL algorithms DDPG, SAC, and TD3 are implemented and compared under identical test scenarios (DG switching and load variation events). Tables 6 and 7 summarize the quantitative comparison for both scenarios. While SAC and TD3 achieve marginally better harmonic mitigation (approximately 4%–5% lower THD), DDPG offers approximately 35% faster training convergence and 40% lower memory requirements. Given the real-time constraints of microgrid control and the marginal performance difference, DDPG remains a practical and computationally efficient choice.

7. Conclusion

This paper presented a deep reinforcement learning-based advanced secondary control strategy for frequency and voltage regulation in inverter-based distributed generation units in microgrids. The proposed multi-agent system employs reinforcement learning algorithms to dynamically optimize control parameters, significantly enhancing system stability. Unlike traditional methods with fixed parameters, the proposed approach adapts control parameters under varying operating conditions, increasing resilience to load variations and renewable intermittency.

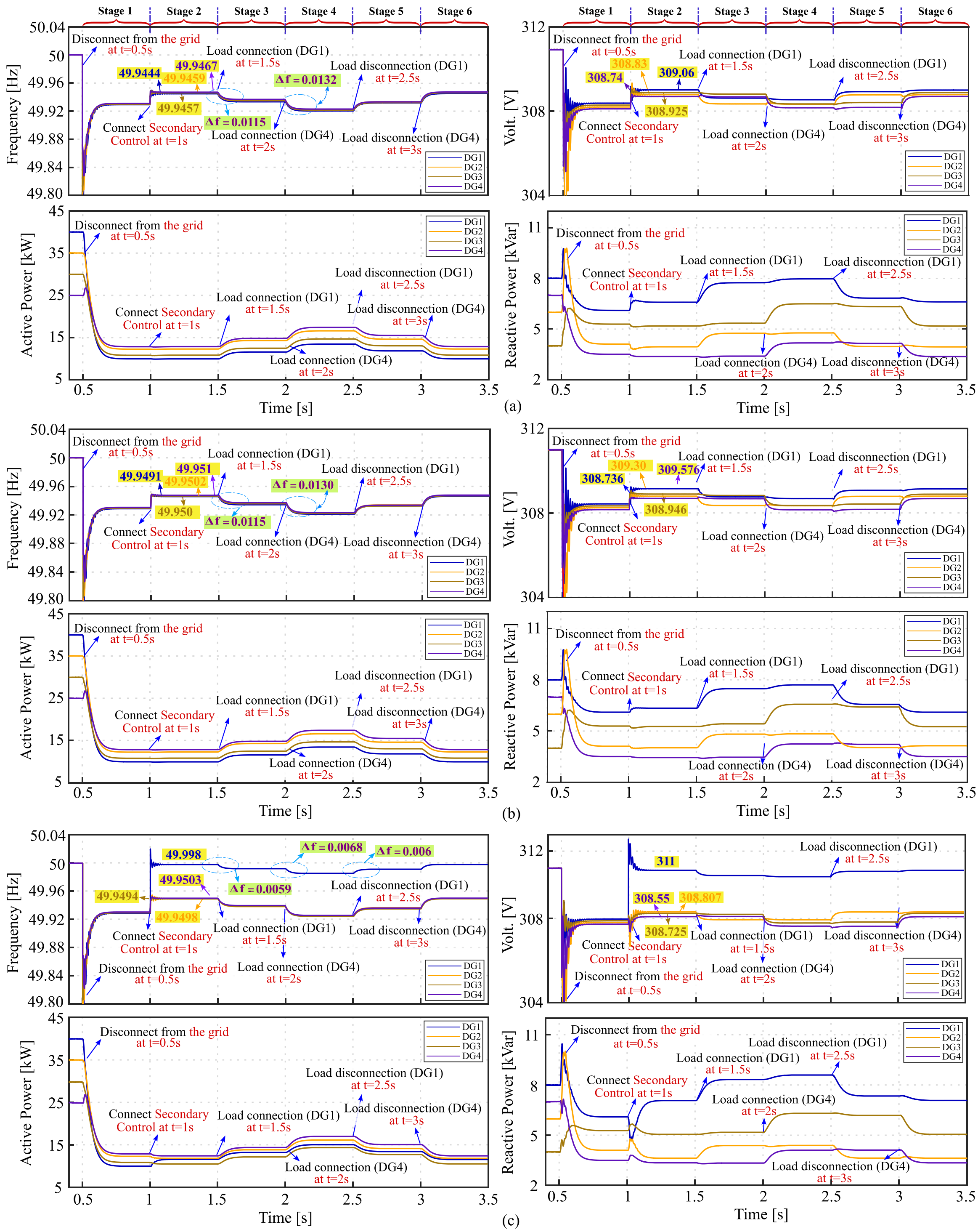

Fig. 7. Dynamic performance of the microgrid under Scenario (ii) considering different secondary control strategies: (a) Conventional PI controller, (b) Fuzzy controller, and (c) The proposed controller.

The control framework is structured in a two-layer hierarchical architecture. At the lower layer, two independent RL agents regulate system frequency and voltage by dynamically tuning the PI controller gains, thereby maintaining nominal operation despite fluctuations in demand and generation. At the upper layer, a supervisory agent monitors harmonic distortion and computes corrective signals for the droop controller, enabling coordinated adaptation and enhancing dynamic stability by effectively reducing transient oscillations. The proposed framework, implemented as a proof-of-concept on a single DG unit, demonstrates the potential of objective-oriented multi-agent DRL for microgrid secondary control. The hierarchical structure with three specialized agents (voltage, frequency, and harmonic control) working cooperatively on a single DG showcases the feasibility of this approach, while the framework's inherent scalability paves the way for future extensions to multiple DG units. MATLAB/Simulink simulations demonstrate that the DDPG-based approach outperforms classical PI and fuzzy control methods in minimizing voltage and frequency deviations. Quantitative results show that the proposed method reduces frequency settling time from 0.22 s to 0.10 s, voltage settling time from 0.20 s to 0.13 s, and THD from 3.07% to 0.68%. Comparative

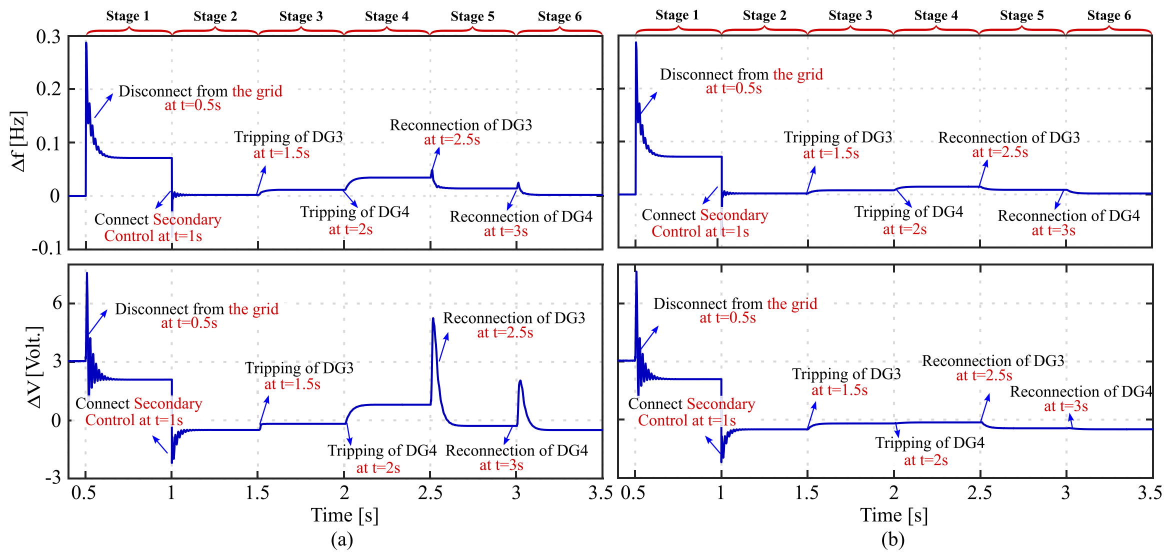

Fig. 8. Voltage and frequency deviation responses under the proposed control: (a) Scenario (i), (b) Scenario (ii).

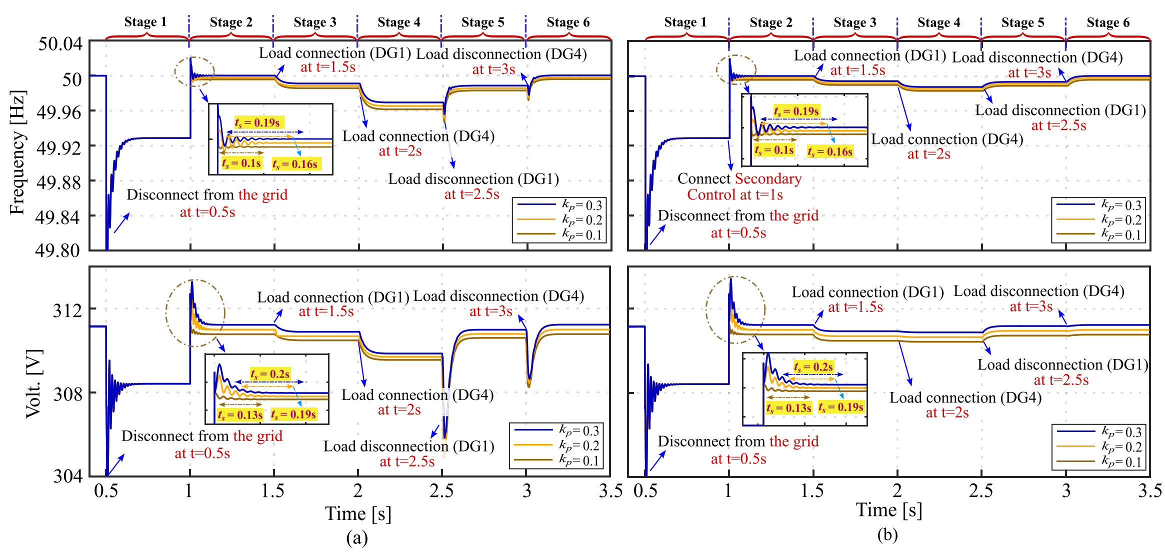

Fig. 9. System performance under the proposed DDPG-based control strategy for different converged values of the proportional gain K_p generated by the agents, in (a) Scenario (i) and (b) Scenario (ii).

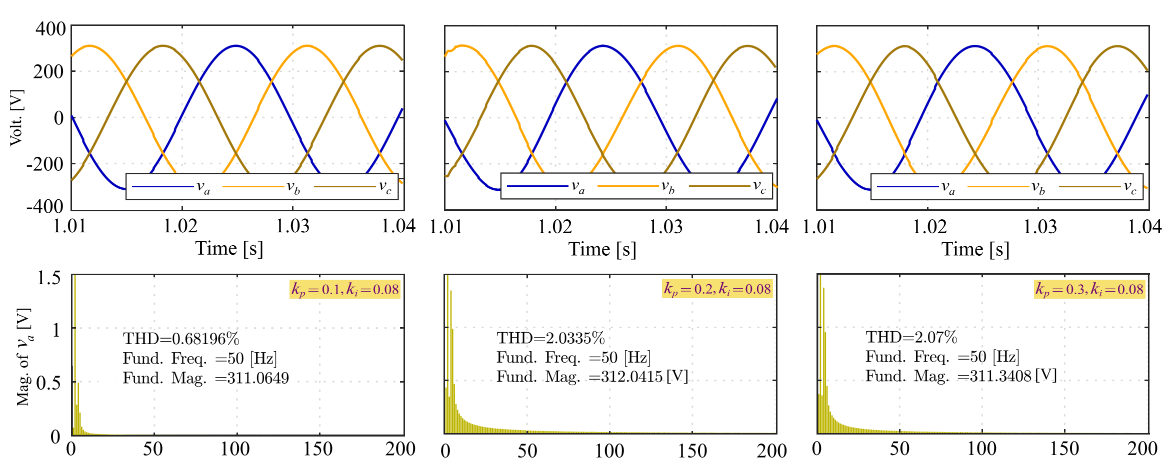

Fig. 10. Comparison of the total harmonic distortion under the proposed DDPG-based control strategy for DG1 in Scenario (i) for different converged gain values generated by the agents: (a) $K_p = 0.1$, (b) $K_p = 0.2$, and (c) $K_p = 0.3$.

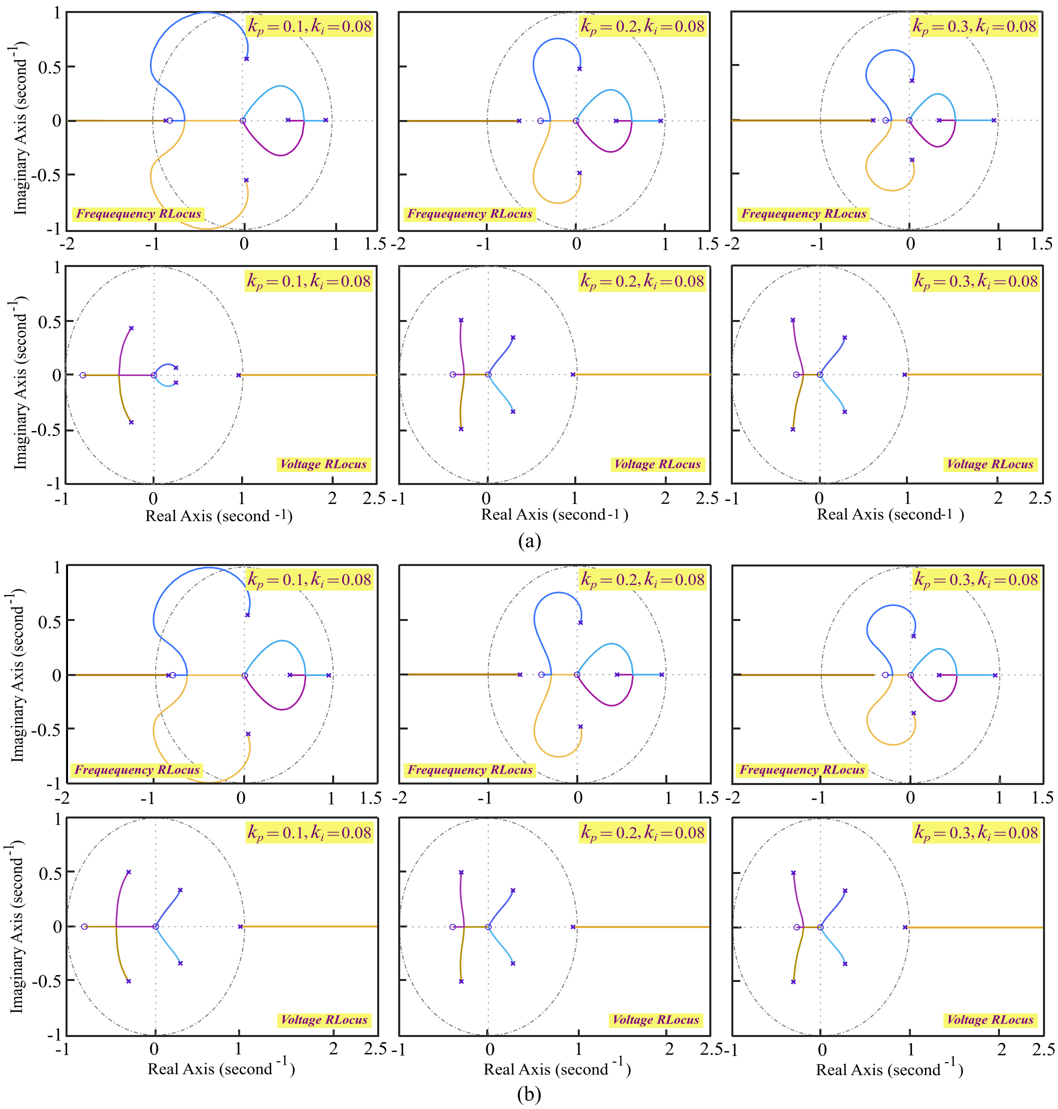

Fig. 11. Root locus of the system for the two studied scenarios: (a) DG switching, (b) Load switching, with $K_i = 0.08$ and $K_p = 0.1, 0.2, 0.3$.

Table 6
Performance comparison under Scenario (i) – DG switching events.

Metric	DDPG	SAC	TD3
Frequency settling time (s)	0.10	0.11	0.11
Voltage settling time (s)	0.13	0.14	0.14
THD (%)	0.68	0.65	0.66
Training time (hours)	2.5	3.8	3.9
Memory usage (GB)	1.2	2.0	2.1

Training time and memory usage are reported under identical training configurations; measured on an Intel Core i7-8700K @ 3.70 GHz, 32 GB RAM, and NVIDIA RTX 2080 Ti.

Table 7
Performance comparison under Scenario (i) – DG switching events.

Metric	DDPG	SAC	TD3
Frequency settling time (s)	0.10	0.11	0.11
Voltage settling time (s)	0.13	0.14	0.14
THD (%)	0.70	0.67	0.68
Training time (hours)	2.5	3.8	3.9
Memory usage (GB)	1.2	2.0	2.1

Training time and memory usage are reported under identical training configurations; measured on an Intel Core i7-8700K @ 3.70 GHz, 32 GB RAM, and NVIDIA RTX 2080 Ti.

analysis with SAC and TD3 confirms that while advanced algorithms offer marginal improvements in harmonic mitigation (≈4%–5%), DDPG achieves superior computational efficiency (35% faster training, 40% less memory), making it more suitable for real-time implementation on embedded hardware.

Limitations and future work: This study is limited to simulation-based validation without hardware experiments. For the proposed method, scalability is claimed theoretically based on the decentralized architecture; empirical validation on larger systems is left for future work. Furthermore, the current study evaluates robustness only under the same operational scenarios used during training (in-distribution evaluation). Generalization of the proposed controller to unseen scenarios (e.g., different types of disturbances, varying DG topologies, or communication failures) has not been validated and is left for future work. Furthermore, the coupled nature of the three agents introduces non-stationarity from each agent's perspective. While sequential training (frequency agent first, then voltage agent, then harmonic agent) partially mitigates this issue by allowing each agent to converge to its optimal policy before the next agent begins training, it is important to note that this sequential procedure does not constitute true cooperative

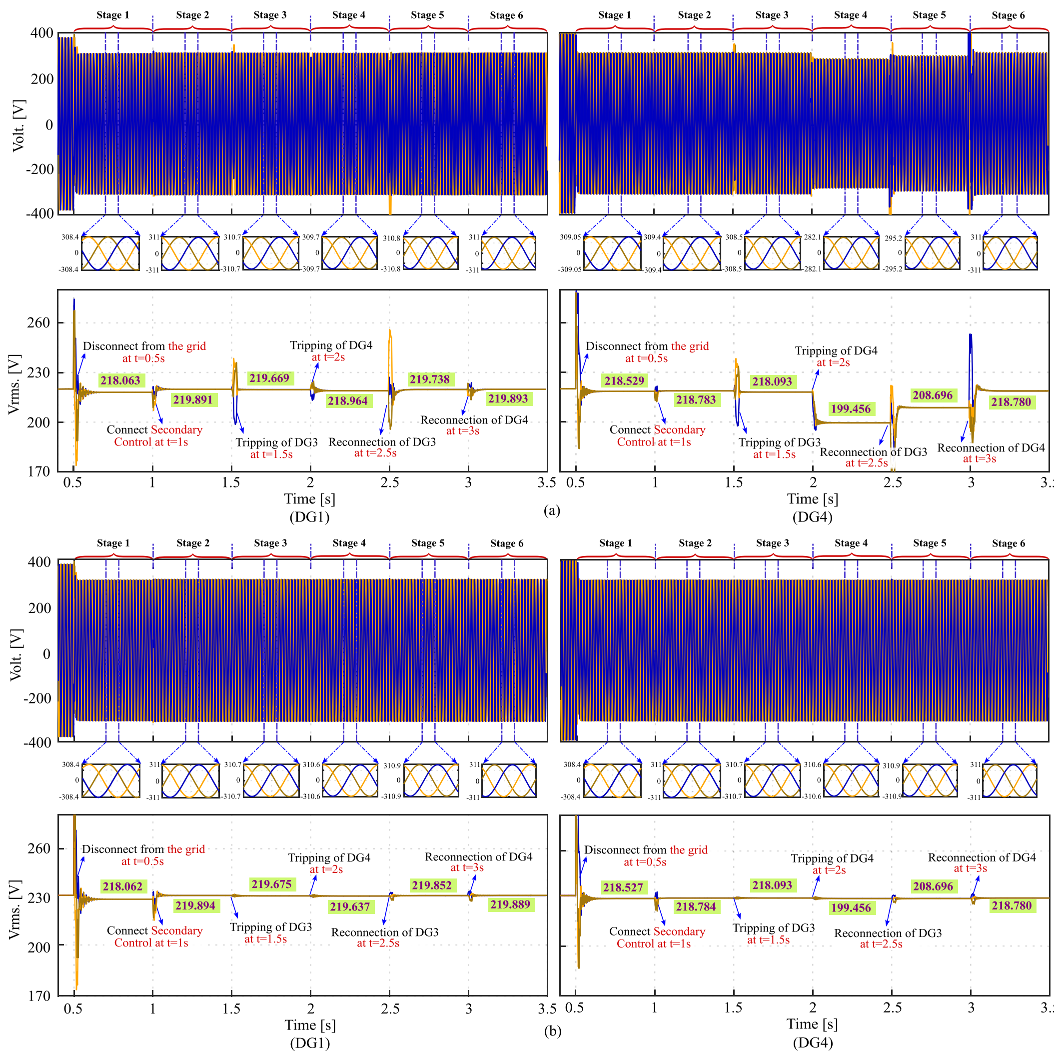

Fig. 12. Performance evaluation of the proposed control strategy in Case III: output voltage and V_{rms} of DG1 and DG4.

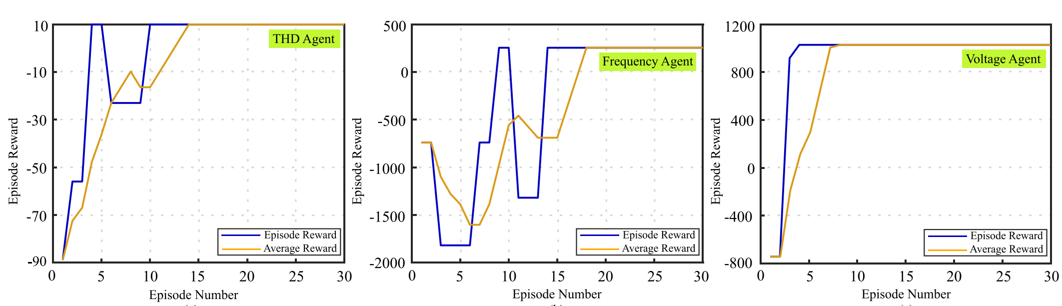

Fig. 13. Episode reward during training process for (a) THD agent, (b) Frequency agent, and (c) Voltage agent.

multi-agent learning. Joint training with proper handling of non-stationarity (e.g., using multi-agent PPO or centralized training with decentralized execution) is left for future work. Credit assignment in this hierarchical structure also requires further investigation. Future research will focus on real-time implementation in laboratory-scale or hardware-in-the-loop microgrids, investigating plug-and-play operation for multiple distributed generation units, and evaluating the effects of communication delays and network latency on control performance. Moreover, the framework can be extended to larger and more complex microgrids, including 33-bus AC/DC hybrid systems. This study demonstrates that deep reinforcement learning provides a powerful and practical tool for intelligent microgrid management, advancing the state of the art in secondary control and paving the way toward scalable, stable, and optimized solutions for next-generation power systems.

CRediT authorship contribution statement

Darioush Razmi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peyman Razmi:** Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Conceptualization. **Jose Rodriguez:** Supervision. **Cristian Garcia:** Supervision. **Zhenbin Zhang:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

C. Garcia and J. Rodriguez acknowledge the support of ANID/FONDECYT through Regular Research Projects 1241099 and Basal Project CIA250006 (AC3E).

Data availability

The data that has been used is confidential.

References

[1] Moradi M, Heydari M, Zarei SF. An overview on consensus-based distributed secondary control schemes in DC microgrids. Electr Power Syst Res 2023;225:109870.
[2] Ortega-Romero I, Serrano-Guerrero X, Barragán-Escandón A, Ochoa-Malhaber C. Optimal integration of distributed generation in long medium-voltage electrical networks. Energy Rep 2023;10:2865–79.
[3] Toularoud MS, Rudposhti MK, Bagheri S, Salemi AH. A hierarchical control approach to improve the voltage and frequency stability for hybrid microgrids-based distributed energy resources. Energy Rep 2023;10:2693–709.
[4] Leveraging blockchain technology for resilient and robust frequency control in a renewable-based hybrid power system with hydrogen and battery storage integration. Energy Convers Manage 2023;283:116888. http://dx.doi.org/10.1016/j.enconman.2023.116888.
[5] Sevostyanov NA, Gorbunov RL. Control strategy to mitigate voltage ripples in droop-controlled DC microgrids. IEEE Trans Power Electron 2023.
[6] Mohammed N, Callegaro L, Ciobotaru M, Guerrero JM. Accurate power sharing for islanded DC microgrids considering mismatched feeder resistances. Appl Energy 2023;340:121060.
[7] Hussaini H, Yang T, Bai G, Urrutia-Ortiz M, Bozhko S. A low-complexity artificial neural network-based optimal droop gain design strategy for DC microgrids onboard the more electric aircraft. IEEE Trans Transp Electrification 2023.
[8] Samanta S, Lagoa C, Chaudhuri NR. Nonlinear model predictive control for droop-based grid forming converters providing fast frequency support. IEEE Trans Power Deliv 2023.
[9] Poonahela I, Krama A, Bayhan S, Abu-Rub H, Begovic M, Shadmand MB. Predictive voltage and frequency restoration for decentralized fcs-mpc based droop controlled dgs in ac microgrids. In: 2022 3rd international conference on smart grid and renewable energy. SGRE, IEEE; 2022, p. 1–6.
[10] Lou G, Gu W, Xu Y, Cheng M, Liu W. Distributed MPC-based secondary voltage control scheme for autonomous droop-controlled microgrids. IEEE Trans Sustain Energy 2016;8(2):792–804.
[11] Putri S, Hosseinipour A, Ge X, Moazeni F, Khazaei J. Voltage restoration in MVDC shipboard microgrids with economic nonlinear model predictive control. 2023, arXiv preprint arXiv:2312.03603.
[12] Li B, Liang Y, Wang G, Li H, Ding J. A control strategy for soft open points based on adaptive voltage droop outer-loop control and sliding mode inner-loop control with feedback linearization. Int J Electr Power Energy Syst 2020;122:106205.
[13] Davari M, Gao W, Jiang Z-P, Lewis FL. An optimal primary frequency control based on adaptive dynamic programming for islanded modernized microgrids. IEEE Trans Autom Sci Eng 2020;18(3):1109–21.
[14] Ye J, Liu L, Xu J, Shen A. Frequency adaptive proportional-repetitive control for grid-connected inverters. IEEE Trans Ind Electron 2020;68(9):7965–74.
[15] Gbadega PA, Saha AK. Load frequency control of a two-area power system with a stand-alone microgrid based on adaptive model predictive control. IEEE J Emerg Sel Top Power Electron 2020;9(6):7253–63.
[16] Jan MU, Xin A, Abdelbaky MA, Rehman HU, Iqbal S. Adaptive and fuzzy PI controllers design for frequency regulation of isolated microgrid integrated with electric vehicles. IEEE Access 2020;8:87621–32.
[17] Lin S-W, Chu C-C, Tung C-F. Data-driven distributed Q-learning droop control for frequency synchronization and voltage restoration in isolated AC micro-grids. IEEE Trans Ind Appl 2023.
[18] Hou S, Chen J, Chen G. Distributed control strategy for voltage and frequency restoration and accurate reactive power-sharing for islanded microgrid. Energy Rep 2023;9:742–51.
[19] Uddin M, Mo H, Dong D, Elsawah S, Zhu J, Guerrero JM. Microgrids: A review, outstanding issues and future trends. Energy Strat Rev 2023;49:101127.
[20] Chen W, Wu N, Huang Y. Real-time optimal dispatch of microgrid based on deep deterministic policy gradient algorithm. In: 2021 international conference on big data and intelligent decision making. BDIDM, IEEE; 2021, p. 24–8.
[21] Dehkordi NM, Nekoukar V. Adaptive distributed stochastic deep reinforcement learning control for voltage and frequency restoration in islanded AC microgrids with communication noise and delay. Sci Rep 2025;15(1):27315. http://dx.doi.org/10.1038/s41598-025-13010-6.
[22] Li S, Blaabjerg F, Anvari-Moghaddam A. A transferable DRL-based intelligent secondary frequency control for islanded microgrids. Electronics 2025;14(14):2826.
[23] Alghamdi B. Distributed consensus-based voltage and frequency control for isolated microgrids with fault-induced delayed voltage recovery mitigation. Front Energy Res 2025;12:1468496.
[24] Xie L, Li Y, Xiao J, Yang J, Xu B, Ye Y. Research on autonomous operation control of microgrid based on deep reinforcement learning. In: 2021 IEEE 5th conference on energy internet and energy system integration (EI2). IEEE; 2021, p. 2503–7.
[25] Khooban MH, Gheisarnejad M. A novel deep reinforcement learning controller based type-II fuzzy system: Frequency regulation in microgrids. IEEE Trans Emerg Top Comput Intell 2020;5(4):689–99.
[26] Skiparev V, Belikov J, Petlenkov E. Reinforcement learning based approach for virtual inertia control in microgrids with renewable energy sources. In: 2020 IEEE PES innovative smart grid technologies europe (ISGT-europe). IEEE; 2020, p. 1020–4.
[27] Barbalho PId, Lacerda V, Fernandes R, Coury DV. Deep reinforcement learning-based secondary control for microgrids in islanded mode. Electr Power Syst Res 2022;212:108315.
[28] Khalid J, Ramli MA, Khan MS, Hidayat T. Efficient load frequency control of renewable integrated power system: A twin delayed DDPG-based deep reinforcement learning approach. IEEE Access 2022;10:51561–74.
[29] Yan R, Wang Y, Xu Y, Dai J. A multiagent quantum deep reinforcement learning method for distributed frequency control of islanded microgrids. IEEE Trans Control Netw Syst 2022;9(4):1622–32.
[30] Li W, Chen R, Xu J. Stability analysis of deep reinforcement learning based control for inverter-dominated microgrids. IEEE Trans Smart Grid 2022;13(4):2518–29.
[31] Wang Y, Zhang L, Ma J. Robust deep reinforcement learning for adaptive secondary control of islanded microgrids. IEEE Trans Ind Electron 2023;70(2):1541–50.
[32] Zhang H, Liu M, Song K. Multi-agent deep reinforcement learning for distributed secondary control of microgrids with stability guarantees. Electr Power Syst Res 2024;215:108543.
[33] Liu C, Guo Y, Huang F. Distributed multi-agent reinforcement learning with communication constraints for microgrid control. IEEE Trans Cybern 2023.
[34] Gao J, Sun W, Zhao L. Hybrid model predictive and reinforcement learning control for microgrid frequency regulation: A hardware-in-the-loop study. Control Eng Pract 2023;135:105497.