Q-learning based event triggered reactive power sharing in AC microgrids

Energy Reports 15 (2026) 109290

Contents lists available at ScienceDirect

Energy Reports

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

Research paper

Q-learning based event triggered reactive power sharing in AC microgrids

Salam Nikravesh [a], Amin Karimi [a,*], Karvan Karimizadeh [b], Yousef Khayat [c,*], Saeed Golestan [d]

[a] Department of Electrical Engineering, Sa.C, Islamic Azad University, Sanandaj, Iran
[b] Department of Electrical Engineering, Mari.C, Islamic Azad University, Marivan, Iran
[c] Department of Electrical Engineering, Chalmers University of Technology, Göteborg, Sweden
[d] Energy Department, Aalborg University, Aalborg, Denmark

ARTICLE INFO

Keywords:
Autonomous microgrids
Reactive power-sharing
Secondary control
Voltage regulation

ABSTRACT

Reactive power-sharing (RPS) and voltage containment within grid-code limits are critical yet challenging objectives in islanded low-inertia AC microgrids (MG); because of unknown feeder impedances, droop-induced deviations, and the high communication burden of conventional distributed secondary control. To address these challenges, this paper proposes a distributed Q-learning-based linear quadratic Gaussian (LQG) policy scheme, where the value function (Q-values) adaptively determines the optimal control policy over time. Unlike conventional LQG methods that rely on deterministic feedback laws derived from system models, the Q-learning-based LQG utilizes a learned policy that adapts to observed performance. The proposed scheme is further enhanced by an event-triggered mechanism (ETM) to simultaneously achieve operational objectives in the physical layer and reduce data exchange in the cyber layer of MGs. Simulation and hardware-in-the-loop experiments demonstrate: (i) voltage restoration, (ii) exact proportional RPS, (iii) considerable reduction in communication events compared with continuous-time benchmarks, and (iv) robust performance under communication delays while preserving stability and power-sharing accuracy. These results enable cost-effective, resilient deployment of islanded MGs with low-bandwidth communication infrastructure, facilitating high renewable-energy penetration in remote communities while ensuring compliance with grid codes.

1. Introduction

Microgrids (MGs) can operate in grid-connected and islanded modes, with hierarchical control structures comprising primary, secondary, and tertiary layers to stabilize frequency and voltage and coordinate power allocation among distributed generators (DGs). Despite advancements, challenges in reactive power-sharing (RPS) and voltage containment persist. Techniques such as virtual impedance and distributed secondary control strategies have shown promise in addressing these issues (Mahmood et al., 2024; Islam et al., 2024; Falope et al., 2024). Event-triggered mechanisms (ETMs) have emerged as efficient solutions to reduce communication overhead in distributed SC systems. Building on these challenges, our research integrates Q-learning-based LQG control to enhance RPS and voltage regulation while reducing communication data exchange. Reactive power-sharing imbalances due to variations in feeder impedance and distributed generators (DGs) characteristics, and difficulties in voltage containment within grid-code limits under dynamic load and generation scenarios are considered as the main reactive power sharing challenges.

To address these challenges, a hierarchical control structure has been developed, which comprises three layers: primary, secondary, and tertiary (Bevrani et al., 2017). The primary control layer stabilizes frequency and voltage and coordinates power allocation among DGs through droop control. However, droop control introduces steady-state errors in frequency and voltage regulation. To mitigate these issues, the secondary control (SC) layer is employed. SC schemes are classified into centralized, decentralized, and distributed approaches (Khayat et al., 2019), with centralized SC suffering from single-point-of-failure risks, decentralized SC being ineffective for synchronization, and distributed SC offering a more flexible and robust data exchange mechanism to achieve system objectives, though it is highly reliant on communication networks (Ahmed et al., 2020).

Unlike significant advancements in hierarchical control structures for MGs, reactive power-sharing (RPS) and voltage containment persist as formidable challenges in MG operation. Primary and secondary control layers are intended to stabilize frequency and voltage while managing power allocation among DGs; however, variations in feeder impedances and DG characteristics frequently obstruct balanced RPS, resulting in power imbalances that precipitate voltage instability and diminish overall system efficacy (Jafari et al., 2019). Furthermore, upholding voltage within prescribed grid-code limits is vital to avert

* Corresponding authors.
E-mail addresses: Amin.karimi1981@iau.ac.ir (A. Karimi), yousef.kh@chalmers.se, ykh@energy.aau.dk (Y. Khayat).

https://doi.org/10.1016/j.egyr.2026.109290
Received 20 May 2025; Received in revised form 5 March 2026; Accepted 11 April 2026
Available online 12 May 2026

overvoltage or undervoltage scenarios, yet this task grows progressively difficult among fluctuating load and generation conditions. To mitigate these issues, techniques such as virtual impedance have demonstrated potential by offsetting line impedance disparities, thereby facilitating precise RPS and bolstering voltage stability. Traditional droop control, while foundational, induces steady-state errors that SC strategies address through centralized, decentralized, or distributed paradigms. ETMs enhance communication efficiency by activating data exchanges only upon notable deviations, surpassing time-triggered approaches that incur redundant overhead (Dehkordi and Nekoukar, 2025). The proposed Q-learning-based LQG scheme, integrated with ETM, offers a model-free adaptation to uncertainties, ensuring robust RPS, voltage regulation, and reduced cyber-layer interactions, as validated through simulations and laboratory experiments.

Conversely, voltage regulation guarantees that DG output voltages stay inside designated boundaries, averting excessive or insufficient voltage states. SC tactics, including distributed variants, proficiently oversee voltage magnitudes while facilitating reactive power allocation. Such techniques sustain voltages in permissible bounds despite fluctuations in demand or connectivity disruptions. A further key hurdle in distributed SC lies in its substantial dependence on connectivity infrastructures for DG information interchange. To counter this, primary methods examined include time-triggered mechanisms (TTMs) and ETMs (Li et al., 2025). TTMs align data transfers at fixed periods or defined rates, yet this frequently causes surplus exchanges in stable phases, regardless of unaltered system variables. Conversely, ETMs commence transfers only upon meaningful shifts or occurrences, rendering them more resourceful by curtailing superfluous interactions. Within TTMs, the standard method for sending/receiving data employs a recurring timer at a fixed rate (Ge et al., 2020). In contrast, ETMs rely on a predefined limit for discrepancies between present and prior samples to formulate a standard exchange plan. Diverse ETMs variants have been put forward for dynamic setups.

In recent years, various types of ETMs have been proposed for dynamical systems, including applications in microgrid control, where they have been successfully employed for secondary frequency and voltage control (Dehkordi et al., 2018; Chen et al., 2017; Qian et al., 2019). The control structures of Dehkordi et al. (2018) and Chen et al. (2017) are respectively distributed and decentralized, and Qian et al. (2019) investigates both of them for the frequency and voltage regulation in AC microgrids. Specifically for voltage control in islanded AC microgrids, Chen and Xiao (2018) and Guo et al. (2017) introduce a novel event-based approach by selecting a new ETM that is based on distributed consensus cooperative scheme. In Weng et al. (2018) and Fan et al. (2016), a cooperative frequency and voltage restoration control is studied with the units correctly participating in reactive power sharing. Ding et al. (2018) and Mohammadi et al. (2021) investigate active power sharing in the presence of the secondary frequency control. Most studies introduce triggering conditions based on state variables. Nevertheless, it is possible to design an ETM in which the triggering condition is independent of the system's state(s) (Wan et al., 2021). On the other hand, some studies have considered the effects of uncertainty on ETM (Ge et al., 2020). In Yang et al. (2019), for instance, a dynamic event-triggered robust secondary frequency control approach is designed for an islanded AC microgrid. All previous research on the ETM of AC microgrids has utilized a unified mechanism for all dynamic modes. While their SC structures can be centralized, decentralized, or distributed.

Despite these advancements, ETM-based SC faces several unresolved challenges. Time delays in communication networks, for example, can degrade system performance, particularly in power grids with high penetration of renewable energy sources (Yan et al., 2022). Research has explored solutions such as H_∞-weighted policies and fuzzy-based memory controllers to address these issues and enable more resilient secondary control strategies (Yan et al., 2023a,b). Furthermore, uncertainty in communication conditions and the saturation effects of linear systems pose additional hurdles that require sophisticated control designs.

Given the challenges of RPS, voltage containment, and communication efficiency in MGs, this research pursues two main objectives: (1) to regulate the voltage of the MG while ensuring correct participation of DGs in RPS, and (2) to reduce the amount of communication data exchanged among DGs. These objectives are achieved through a Q-learning-based LQG control scheme implemented at the SC level. Unlike traditional LQG control, which combines an LQR with a Kalman filter and relies on a known system model, the proposed Q-learning-based approach integrates reinforcement learning techniques into the LQG framework. Q-learning, a model-free method, enables the control strategy to learn optimal actions through trial and error, making it particularly effective for systems with dynamic and partially observable conditions (see Table 1).

This paper makes the following important contributions:

- Model-Free Learning Capability: Unlike conventional LQG control, which requires a known and accurate system model, the proposed Q-learning-based LQG learns from interactions with the environment. It updates Q-values based on a quadratic performance index, allowing it to adaptively find an optimal control policy over time. This feature is particularly beneficial for reactive power sharing in MGs where line impedances are often unknown.
- Dynamic Policy Adaptation: While traditional LQG employs a deterministic feedback law derived from the system model, the Q-learning-based LQG adapts dynamically based on observed performance. This approach effectively addresses uncertainties in grid impedance and load variations, which are critical for voltage stability and reactive power sharing.
- Robustness and Communication Efficiency: The Q-learning-based LQG exhibits robustness in reactive power sharing and voltage regulation, even under challenging conditions such as system noise and communication delays. Additionally, the number of communication exchanges is significantly reduced in the presence of the ETM compared to conventional ETM-based and time-triggered SC approaches, enhancing its practicality for real-world MG operations.

The rest of this paper is organized as follows: Section 2 provides an overview of the DG topology, primary control, and the design of the averaging-based distributed secondary control. Section 3 presents the multi-event triggered mechanism, detailing its design procedure and integration within the control strategy. In Section 4, the proposed learning-based LQG policy scheme, combined with the ETM, is applied to an islanded AC microgrid, and its effectiveness is evaluated through comprehensive simulation and experimental results. Finally, Section 5 concludes the paper with summarizing the key findings.

2. System modelling and dynamics

Fig. 1 illustrates the hierarchical control architecture of autonomous MGs, comprising three main layers. At the top, the Global Control manages coordination among multiple MGs and integrates with the main grid via an SCADA system. Each MG has a Microgrid Central Controller (MGCC) that oversees local operations and ensures communication with the global controller. Within the MG, the configuration of control layers is further divided into Primary Control (PC), which stabilizes individual DGs units, and the SC, which coordinates DG units to correct voltage and frequency deviations at the microgrid level with appropriate active and reactive power sharing. A DG unit's physical and PC configuration including the inner control loops of the PC level is demonstrated in Fig. 2. The DG unit governed as a droop-based grid forming VSC is connected to the electrical network and includes three main control loops. The Voltage Control Loop regulates the output voltage of the DG unit, while the Current Control Loop ensures

Table 1
LQR-based solutions to improve reactive power sharing in islanded MGs.

Ref.	Controller type	Communication	Complexity	Voltage control	Reactive power-sharing	Experimental verification
Mahmud et al. (2014)	Robust LQR	Continuous	Low	✔	✘	✘
Khayat et al. (2018)	Optimal LQR	Continuous	Medium	✔	✘	✔
Sen and Kumar (2020)	LQR	Continuous	Medium	✔	✘	✘
Suresh et al. (2020)	LQR	Continuous	Low	✔	✘	✘
Feng and Wang (2021)	PSO-based LQG	ETM	Medium	✔	✘	✘
Mejia-Ruiz et al. (2021)	LQG	Continuous	Medium	✔	✘	✘
Mejia-Ruiz et al. (2022)	LQG	Continuous	High	✔	✔	✘
Wang et al. (2021)	Reinforcement learning	Continuous	High	✔	✘	✘
Lu et al. (2021)	PI	ETM	Low	✔	✔	✔
Kang et al. (2024)	Nonlinear Lyapunov	ETM	Low	✔	✔	✘

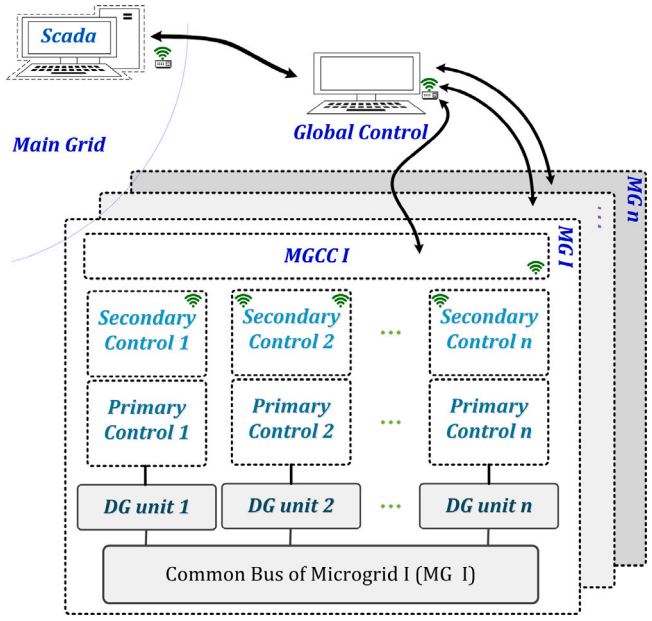

Fig. 1. Hierarchical control of autonomous MGs.

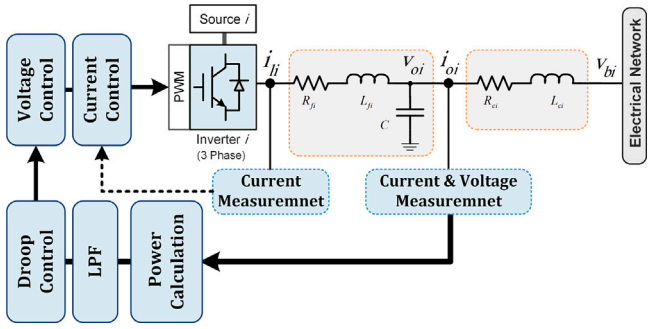

Fig. 2. Physical part of a DG unit connected to the network with the configuration of the inner control loops in the primary control level of MGs.

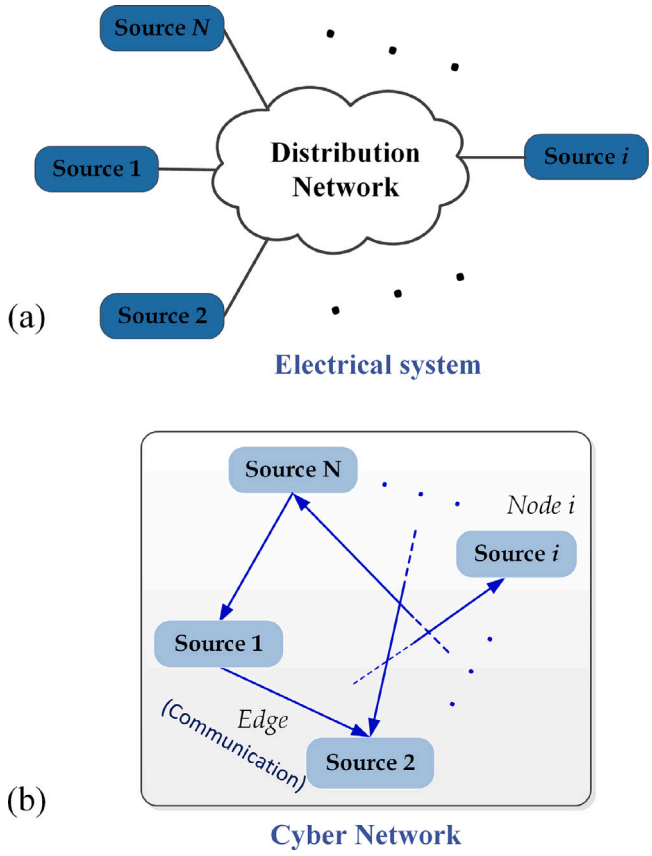

Fig. 3. Cyber–physical representation of MG system: (a) physical layer for electrical connection of the sources, and (b) cyber layer for communication interconnection of the sources.

stable current injection into the grid to maintain system stability and desired power output. Additionally, an inner Low-Pass Filter (LPF) smoothens signals, and a Droop Control mechanism allows the DG unit to share loads with other units proportionally according to the active and reactive power's droop gains.

Fig. 3 represents the cyber–physical architecture of a MG system, illustrating the interaction between the physical and cyber layers. In the physical layer of Fig. 3(a), the distribution network electrically connects multiple energy sources to enable power-sharing and delivery across the system. In the cyber layer of Fig. 3(b), a CN interlinks these sources (nodes) via communication edges, enabling the exchange of real-time information. This integration allows for coordinated control, monitoring, and decision-making to ensure efficient and stable micro-grid operation. The diagram highlights how the physical and cyber systems work in tandem, with the electrical connections delivering power and the communication network supporting advanced control strategies.

2.1. Graph theory

A bidirectional CN offers that the DGs can send in and receive information from each other; in this CN, DGs are considered as graph network nodes, and the communication links as graph edges. In this way, the MG can be introduced as a set of DGs which shares information via a communication network (CN) links. therefore, an undirected graph $\mathcal{G}_r$ can be employed as

$$\mathcal{G}_r = (V_\mathcal{G}, E_\mathcal{G}, A_\mathcal{G}) \tag{1}$$

to describe the graph topology of the MG with $N > 0$ for showing the number of the connected DGs; such that

$$V_{\mathcal{G}} = (v_1, v_2, \ldots, v_N), E_{\mathcal{G}} \subset V_{\mathcal{G}} \times V_{\mathcal{G}} \tag{2}$$

where $V_{\mathcal{G}}$ is for the MG nodes and $E_{\mathcal{G}}$ is for the MG edges. Then, the symmetric matrix $A_{\mathcal{G}}$ in form of

$$A_{\mathcal{G}} = [a_{ij}] \in R^N \times R^N$$

is stood for as the adjacency matrix of $\mathcal{G}_r$. In case of data exchange from DG_i to DG_j, $(j,i) \in E_{\mathcal{G}} \Leftrightarrow a_{ij} = a_{ji} = 1$ is always held ; otherwise, $a_{ij} = a_{ji} = 0$. $L = [l_{ij}] \in R^N \times R^N$ expresses the Laplacian matrix of the graph $\mathcal{G}_r$, with the elements $l_{ii} = \sum_{j=1, j\neq i}^{N} a_{ij}$ and $l_{ij} = -a_{ij}$.

Let $x_i \in R$ stand for the value of state variables of interest at bus i; which in our context, it refers to dq voltage and currents, active, and reactive powers. According to the consensus algorithm, the variables x_i can achieve consensus in case of $x_i(t) - x_j(t) \to 0$ as $t \to \infty$. If the A matrix has a spanning tree, i.e., there is a path from any single node to any other one in the communication graph, consensus can be achieved via the following algorithm (Khayat et al., 2018):

$$\dot{x}_i = \left(x_{ref} - x_i\right) - \sum_{j \in \mathbb{N}(i)} a_{ij}\left(x_i - x_j\right)$$

LBCLs are required to implement the distributed SC strategy proposed in this work. During the simulations in Section 5 and the experimental tests, the employed CN is updated at a frequency of 100 Hz. The distance between DGUs is considered at the weighting factors a_{ij}, and, IEC 61850 as the communication protocol (specifically developed for MG applications) is recommended for data exchange among DGUs. A data packet with a size of 27 bits is shared to carry out control actions in accordance with IEC 61850.

2.2. Control objectives

The active and reactive powers (P and Q) of the ith DGU injected toward the node j can be simply defined as follows:

$$P_i = \alpha\left(V_i^2 - V_i V_j \cos\theta\right) + \beta V_i V_j \sin\theta \tag{3}$$

$$Q_i = \beta\left(V_i^2 - V_i V_j \cos\theta\right) - \alpha V_i V_j \sin\theta \tag{4}$$

where

$$\alpha = \frac{R_{ij}}{\left|Z_{ij}\right|^2}, \quad \beta = \frac{X_{ij}}{\left|Z_{ij}\right|^2} \tag{5}$$

$$\left|Z_{ij}\right| = \sqrt{R_{ij}^2 + X_{ij}^2}. \tag{6}$$

where V_i stands for the voltage magnitude of the ith DG, and $|Y_{ij}|$ represents the magnitude of the admittance Y_{ij}. On the other hand, the voltage droop control can be stated as follows

$$v_{odi} = \mathcal{V}_i^* - n_{Qi} Q_i(t), \tag{7}$$

$$v_{oqi} = 0, \tag{8}$$

where $\mathcal{V}_i^*$ is the MG's nominal voltage value given as the set point to the droop control mechanism, $Q_i(t)$ represents the measured reactive power for the i^th DG, and n_{Qi} is the droop gain. Eqs. (7) and (8) describe the voltage reference generation in the primary control level of each distributed generator (DG) using a droop control mechanism. In Eq. (7), v_{od_i} is the d-axis output voltage of the ith DG unit, computed by subtracting a scaled value of the measured reactive power $Q_i(t)$ from the nominal voltage V_i^*. The droop gain n_{Qi} governs the sensitivity of voltage drop relative to reactive power injection, enabling proportional reactive power sharing among DGs. Eq. (8) fixes the q-axis voltage reference v_{oq_i} to zero to ensure decoupled control and voltage-oriented alignment in a synchronous reference frame.

In order to reduce voltage deviations from nominal magnitudes and to share reactive power appropriately, an SC must be considered and designed. The controller for both voltage and reactive power is given by a derivation of (7) as a function of time.

$$\begin{aligned}\dot{\mathcal{V}}_i^* &= \dot{v}_{odi} + n_{Qi}\dot{Q}_i \\ &= u_i^{\delta v} + n_{Qi} u_i^{\delta Q}\end{aligned} \tag{9}$$

where $u_i^{\delta v}$ and $u_i^{\delta Q}$ present the SC's voltage and RPS, respectively.

The secondary voltage control adjusts $u_i^{\delta v}$ and $u_i^{\delta Q}$ such that the SC goals are met as

$$\lim_{t \to t_f} v_i(t) \approx v^*, \tag{10}$$

$$\lim_{t \to \infty}\left(n_i Q_i(t) - n_j Q_j(t)\right) = 0. \tag{11}$$

In this way, the following definition can be provided:

Definition 1. In an MG, the voltage restoration procedure is done when (10) is met, and the reactive power sharing is reached if (11) is met.

Considering that the PC operates with a much higher bandwidth in contrast to the SC layer, the inner loop's dynamics can be neglected for the SC designs, and only the voltage and RPS dynamics are included in the Q-LbC to be developed.

3. Q-learning based reactive power control

In order to develop the proposed Q-LbC control scheme for meeting the SC goals mentioned above, first let us recall the required state-space representation of a DGU, as formulated in Bidram et al. (2013).

$$\dot{\delta}_i = \omega_i^{nom} - m_i^P P_i - \omega_{com} \tag{12a}$$

$$\dot{P}_i = \omega_{ci}\left(v_{odi} i_{odi} + v_{oqi} i_{oqi} - P_i\right) \tag{12b}$$

$$\dot{Q}_i = \omega_{ci}\left(v_{odi} i_{oqi} - v_{oqi} i_{odi} - Q_i\right) \tag{12c}$$

$$\dot{i}_{ldi} = \frac{-R_{fi}}{L_{fi}} i_{ldi} + \omega_{com} i_{lqi} + \frac{v_i^{nom} - n_i^Q Q_i - v_{odi}}{L_{fi}} \tag{12d}$$

$$\dot{i}_{lqi} = \frac{-R_{fi}}{L_{fi}} i_{lqi} - \omega_{com} i_{ldi} - \frac{v_{oqi}}{L_{fi}} \tag{12e}$$

$$\dot{v}_{odi} = \omega_{com} v_{oqi} + \frac{i_{ldi} - i_{odi}}{C_{fi}} \tag{12f}$$

$$\dot{v}_{oqi} = -\omega_{com} v_{odi} + \frac{i_{lqi} - i_{oqi}}{C_{fi}} \tag{12g}$$

$$\dot{i}_{odi} = \frac{-R_{fi}}{L_{ci}} i_{odi} + \omega_{com} i_{oqi} + \frac{v_{odi} - v_{bdi}}{L_{ci}} \tag{12h}$$

$$\dot{i}_{oqi} = \frac{-R_{fi}}{L_{ci}} i_{oqi} - \omega_{com} i_{odi} + \frac{v_{oqi} - v_{bqi}}{L_{ci}} \tag{12i}$$

These equations describe the internal dynamics of each DGU, including active/reactive power, currents, voltages, and internal states, represent the nonlinear dynamic model of each DGU in the microgrid, derived based on dq-frame modeling and consistent with previous studies, such that (12a) describes the frequency deviation dynamics based on droop response to active power P_i, (12b)–(12c) show the dynamics of active and reactive powers, respectively, governed by current–voltage products, (12d)–(12e) present the d- and q-axis current dynamics of the local filter inductance, affected by line resistance and cross-coupling terms. (12f)–(12g) present the voltage dynamics of the output capacitor in the dq-frame, reflecting the influence of filter currents. And, (12h)–(12i) show the current dynamics of the output LCL inductor (or load side), including disturbances from bus voltages v_{bd_i}, v_{bq_i}.

The dynamical model of DGU_i can be written in the following form Bidram et al. (2013):

$$\dot{x}_i^{DGU} = \mathbf{f}_i(x_i^{DGU}) + \mathbf{g}_i u_{i1}^{v} + \mathbf{h} u_{i2}^{\omega} + \mathbf{k}_i(x_i^{DGU}) D_i \tag{13a}$$

$$y_i = \mathbf{c}_i x_i^{DGU} \tag{13b}$$

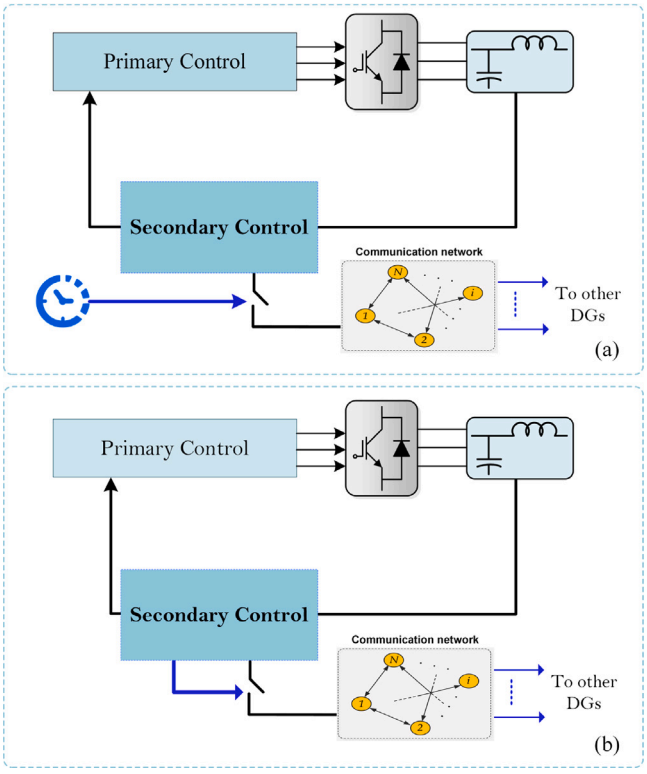

Fig. 4. Data exchange among DG units: (a) Continuous time trigger, and (b) ETM policy.

where

$$x_i^{DGU} = [\delta_i \;\; P_i \;\; Q_i \;\; i_{ldi} \;\; i_{lqi} \;\; v_{odi} \;\; v_{oqi} \;\; i_{odi} \;\; i_{oqi}]^T.$$

The term $D_i = [\omega_{com} \;\; v_{bdi} \;\; v_{bqi}]^T$ is introduced as a known disturbance. The detailed expressions for $\mathbf{f}_i$, $\mathbf{g}_i$, and $\mathbf{k}_i$ can be extracted from Bidram et al. (2013) (see Fig. 4).

Before deriving the discrete-time state-space model used in the learning-based LQG controller, it is essential to theoretically justify the modeling framework and its consistency with the established literature on inverter-based MGs. The dynamic equations in (12a) represent the standard dq-frame model of a VSC equipped with inner current and voltage control loops and an outer droop controller (Bidram et al., 2013), where the fast inner loops allow the SC to treat the system as approximately linear around its operating point. Following the hierarchical control paradigm, the SC layer operates at a bandwidth significantly slower than the primary voltage and current loops. Because of this timescale separation, prior studies consistently show that the nonlinear dq-frame converter dynamics can be linearized around an equilibrium and approximated by a discrete-time linear time-invariant (LTI) system for SC design (Khayat et al., 2019). This linearization captures the dominant dynamics relevant to voltage restoration and reactive power sharing while enabling controllers with tractable quadratic performance objectives. Therefore, the discrete-time model (15) used in the proposed Q-learning–based LQG design, provides the structural form required to define the quadratic cost function and the policy parameterization, consistent with classical LQR/LQG formulations employed in MG's SC. Similar linearized models underpin other widely used voltage and reactive-power secondary controllers, including distributed LQR methods (Sen and Kumar, 2020; Khayat et al., 2018) and reinforcement-learning-based control (Wang et al., 2021).

In the context of Q-learning, only the state–input structure of the LTI model is utilized during policy evaluation, whereas the exact values of matrices A and B are not required. This is a key distinction from

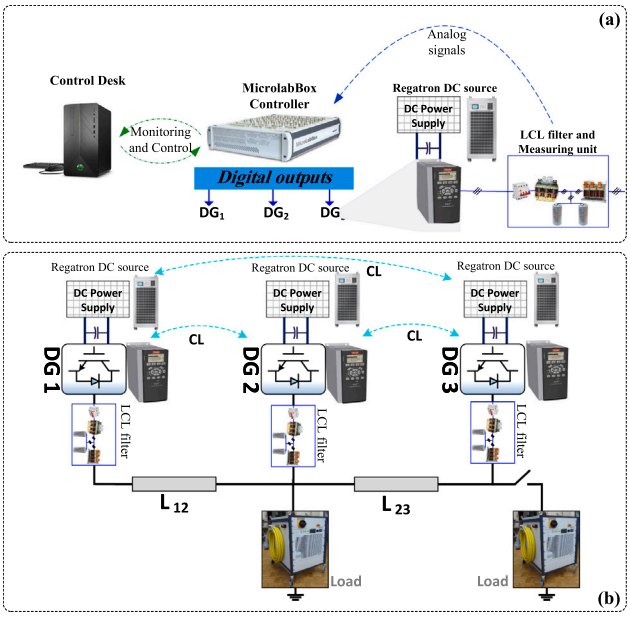

Fig. 5. AC microgrid network case study and the experimental set up configuration. (a) experimental set up, and (b) MG's physical and cyber network configuration.

conventional LQG, whose optimal control law depends explicitly on the solution of the algebraic Riccati equation. Here, the Q-learning approach learns the value function and policy directly from system trajectories, making it suitable for MGs where line impedances and converter parameters may be uncertain or time-varying (Guo et al., 2017; Fan et al., 2016). Consequently, the adopted model structure is theoretically justified for enabling model-free learning while preserving the LQG control architecture.

3.1. Control design

Considering a discrete-time linear time-invariant form of the dynamic systems of (13) is given by the following state-space representation:

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ y_k &= Cx_k \end{aligned} \tag{14}$$

where $x_k \in \mathcal{R}_n$ stands for the state variables, $u_k \in \mathcal{R}_m$ refers to the inputs, and $y_k \in \mathcal{R}_p$ refers to the outputs of the system. Assuming that the controllability and observability requirements on the pairs of (A, B) and (A, C) are met, then, to find the feedback control law $u_k = h(x_k)$ that minimizes the long-term cost

$$\mathcal{F}_h(x_k) = \sum_{i=k}^{\infty} r(x_i, u_i) \tag{15}$$

where $r(x_i, u_i)$ refers to the kth step cost result in the control input u_i in the state x_i and $\mathcal{F}_h(x_k)$ stands for the objective cost of the policy $u_k = h(x_k)$. In this way, the optimal value for the objective function $\mathcal{F}_h$ is given as

$$\mathcal{F}^*(x_k) = \min_h \sum_{i=k}^{\infty} r(x_i, h(x_i)). \tag{16}$$

considering a quadratic form for $r(x_k, u_k)$ as

$$r(x_k, u_k) = x_k^T Q x_k + u_k^T R u_k \tag{17}$$

for an LQR problem, where $Q \in \mathcal{R}_{n\times n>0}$ and $R \in \mathcal{R}_{m\times m>0}$ are referred to the performance matrices. By this way, the state-feedback control forms

$h(x_k) = \mathbf{K}x_k$ with the quadratic cost function (15), with the optimal state feedback control $\mathbf{K}^*$ that minimizes (17) for (13) is

$$u_k^* = \mathbf{K}^* x_k = -(R + B^T P^* B)^{-1} B^T P^* A x_k. \tag{18}$$

Algorithm 1 State Feedback learning-based LQG Policy

Require: Provide discrete-time LTI form (14) and (15)
Ensure: Check observability for (A, C)
Ensure: Check controllability for (A, B)
  Set the discounting factor as $0 < \gamma < 1$
  Set states performance matrix: $Q \in \mathcal{R}_{n\times n>0}$
  Set inputs performance matrix: $R \in \mathcal{R}_{m\times m>0}$
  while (19) is met **do**
    if (21) is held **then**
      Calculate $\mathbf{Q}_K(x_k, u_k)$
      update Control law (27)
      update ETM law (28)
    else
      Set a new value for states performance matrix
      Set a new value for inputs performance matrix
    end if
  end while

Considering that the controllability of (A, B) and observability of (A, C) are met, and $P^* = P^{*T}$ is the unique positive definite solution to the ARE

$$A^T PA - P + Q - A^T PB(R + B^T PB)^{-1} B^T PA = 0. \tag{19}$$

A straightforward solution to solve the LQR problem is to employ the Bellman learning approach, however, it suffers from excitation noise. To address this drawback, a discounting factor such as $0 < \gamma < 1$ is typically employed in the cost function, which leads to reform the ARE as follows,

$$V(x_k,\ u) = \sum_{i=k}^{\infty} \gamma^{i-k} r(x_i,\ u_i). \tag{20}$$

$$-P_\gamma + Q + \gamma(A^T P_\gamma A - A^T P_\gamma B(R/\gamma + B^T P_\gamma B)^{-1} B^T P_\gamma A) = 0 \tag{21}$$

and the corresponding discounted control is given as

$$u_k^* = \mathbf{K}_\gamma^* x_k = -(R/\gamma + B^T P_\gamma B)^{-1} B^T P_\gamma A x_k. \tag{22}$$

It is worth noting that for ensuring the stability of the control gain $\mathbf{K}^*$ in the presence of the discounted factor γ, the cost function should be bounded as

$$\sum_{i=k}^{\infty} \gamma^{i-k} x_i^T \left(Q + K_\gamma^T R K_\gamma\right) x_i < \infty. \tag{23}$$

However, due to the discounting factor γ, the boundedness of the cost does not ensure the convergence of x_k to 0 as k goes to ∞. To address this, a Bellman-based Q-function is defined as

$$\mathbf{Q}_K(x_k, u_k) = \gamma^{i-k} x_i^T \left(Q + K_\gamma^T R K_\gamma\right) x_i + \mathcal{F}_h(x_{k+1}) \tag{24}$$

which is the sum of the single-step cost of executing an arbitrary control u_k from state x_k together with the cost of executing policy K from x_{k+1} and all future states. For the case of LQR, the $\mathbf{Q}_K$-function can be represented as

$$\begin{aligned}\mathbf{Q}_K(x_k, u_k) &= \gamma x_i^T \left(Q + K_\gamma^T R K_\gamma\right) x_i + x_{k+1}^T Q x_{k+1} \\ &= \gamma x_i^T \left(Q + K_\gamma^T R K_\gamma\right) x_i \\ &\quad + (Ax_k + Bu_k)^T Q (Ax_k + Bu_k)\end{aligned} \tag{25}$$

or equivalently as

$$\mathbf{Q}_K(x_k, u_k) = \tag{26}$$

Table 2
Electrical and control parameters of the test MG.

Electrical parameters		
Parameters	Symbol	Value
Output voltage of rectifier	V_{DC}	650 V
Nominal voltage magnitude	V_i	325 V
Nominal frequency	f	50 Hz
Switching frequency	f_s	10 kHz
Capacitance of LCL filter	C_f	25 μF
Input/output inductance of LCL filter	L_i/L_o	1.8 mH
Load 1	Z_1	43 Ω, 0.3 H
Load 2	Z_2	124 Ω, 0.1 H
Inner loop coefficients and other control parameters		
Control Parameters	DGU: 2 and 4	DGU: 1 and 3
$P-\omega$ droop coefficient	0.001 rad/W s	0.002 rad/W s
$Q-v$ droop coefficient	0.005 V/VAr	0.01 V/VAr
Current integral/proportional terms	1000/0.5	1000/0.5
Voltage integral/proportional terms	120/0.05	120/0.05
Control design gain c	200	200

$$\begin{bmatrix} x_k \\ u_k \end{bmatrix}^T \begin{bmatrix} \gamma\left(Q + K_\gamma^T R K_\gamma\right) + A^T PA & A^T PB \\ B^T PA & R + B^T PB \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix} = \begin{bmatrix} x_k \\ u_k \end{bmatrix}^T H \begin{bmatrix} x_k \\ u_k \end{bmatrix}$$

which offers a quadratic form for both the states and inputs. Considering the cost $\mathbf{F}_h^*$ associated with policy K, we can find an improved policy $\mathcal{K}^*$ by selecting

$$\begin{aligned} u_k^* &= \mathcal{K}^* x_k \\ &= \arg\min_u (Q_K(x_k, u_k)). \end{aligned} \tag{27}$$

The minimizing u provides the considered performance index and can be obtained through the solution of $(\partial/\partial u_k)\mathbf{Q}_K = 0$, where its convergence can be guaranteed when (21) is met. In addition, the updating times instants t_k^i for implementation of the ETM through the Q-function mechanism can be calculated as

$$\begin{aligned} t_{k+1}^i = \\ \inf\{t > t_k^i : \|\mathbf{Q}_K(x_k, u_k) - \mathbf{Q}_{K-1}(x_{k-1}, u_{k-1})\| + \gamma_z > \varepsilon_i\} \end{aligned} \tag{28}$$

where γ_z is a positive value for keeping the zeno behaviour challenges.

Remark I. It can be observed that a balance between control update and signal transmission burden can be provided by an appropriate tuning of ε_i in the ETM scheme. And, due to its always positive value of γ_z, it can avoid Zeno phenomenon.

It is worth to highlight that the discrete LTI model is used only offline for initialization and stability verification. Online learning part is data-driven. Indispensable modeling elements for implementation can be itemized as

- Offline subspace identification of nominal (A,B) only for verifying controllability/observability and for providing the initial stabilizing gain;
- No online matrix inversion or explicit A/B is required after the learning phase.

3.2. Stability and convergence

To prove the policy iteration converges to an ε -near-optimal policy and that the closed-loop system remains asymptotically stable for any fixed (or slowly adapting) policy under the Q-value-deviation triggering condition (28), let us first consider the following assumptions:

- (A1) The communication graph $\mathcal{G}$ is undirected and connected (contains a spanning tree).

- (A2) The initial policy K^0 is stabilizing (obtained offline from nominal subspace-identified model).
- (A3) States are measured continuously locally; only neighbor consensus errors $\bar{e}_{v,i}$, $\bar{e}_{Q,i}$ are transmitted at event instants (ZOH).
- (A4) Inter-event times are bounded below by $\tau_{\min} > 0$ (enforced by $\gamma_z > 0$; verified experimentally as $\tau_{\min} \geq 35$ ms even under 200 ms delay).

Now, for the closed-loop asymptotic stability under sparse asynchronous updates, let us consider the augmented local state of DG_i:

$$z_i = \begin{bmatrix} x_{\text{local},i}^T & \bar{e}_{v,i}^T & \bar{e}_{Q,i}^T \end{bmatrix}^T \in \mathbb{R}^5, \tag{29}$$

with secondary input $u_i = K_i \hat{z}_i$, where $\hat{z}_i(t)$ uses the last received neighbor values (ZOH). Let us define the holding (measurement) error

$$e_i(t) = \hat{z}_i(t) - z_i(t), \quad t \in [t_k^i, t_{k+1}^i). \tag{30}$$

The local closed-loop dynamics (after fast inner loops) become

$$\dot{z}_i(t) = A_i z_i(t) + B_i K_i (z_i(t) + e_i(t)) = A_{\text{cl},i} z_i(t) + B_i K_i e_i(t), \tag{31}$$

where $A_{\text{cl},i} = A_i + B_i K_i$, and K_i be the converged (near-optimal) gain from policy iteration. By construction, there exists a unique positive-definite $P_i = P_i^T > 0$ solving the discrete/continuous ARE

$$A_{\text{cl},i}^T P_i + P_i A_{\text{cl},i} + Q_i + K_i^T R_i K_i = 0 \tag{32}$$

(with $Q_i > 0$, $R_i > 0$). Let us choose the quadratic Lyapunov function

$$V_i(z_i) = z_i^T P_i z_i. \tag{33}$$

Its derivative along the trajectories is

$$\begin{aligned} \dot{V}_i &= z_i^T (A_{\text{cl},i}^T P_i + P_i A_{\text{cl},i}) z_i + 2 z_i^T P_i B_i K_i e_i \\ &= - z_i^T (Q_i + K_i^T R_i K_i) z_i + 2 z_i^T P_i B_i K_i e_i. \end{aligned} \tag{34}$$

Applying the inequality $|2a^T b| \leq \varepsilon_1 \|a\|^2 + \frac{1}{\varepsilon_1}\|b\|^2$ with $\varepsilon_1 = \frac{\lambda_{\min}(Q_i + K_i^T R_i K_i)}{2}$, leads to:

$$\dot{V}_i \leq -\frac{\lambda_{\min}(Q_i + K_i^T R_i K_i)}{2} \|z_i\|^2 + \beta_i \|e_i\|^2, \tag{35}$$

where $\beta_i = \frac{2\|P_i B_i K_i\|^2}{\lambda_{\min}(Q_i + K_i^T R_i K_i)}$. The Q-value-deviation trigger (29) directly bounds the error. Because the parameterized Q-function is quadratic, $Q_K(z,u) = [z^T \ u^T] \bar{H} [z^T \ u^T]^T$, a small Frobenius change $\|\Delta \bar{H}\|_F$ (or value deviation) implies (by continuity and positive-definiteness of $\bar{H}_{uu}$) that the state deviation satisfies

$$\|e_i(t)\| \leq \sigma_i \|z_i(t)\|, \quad \sigma_i < \sqrt{\frac{\lambda_{\min}(Q_i + K_i^T R_i K_i)}{2\beta_i}}, \tag{36}$$

whenever the trigger has not fired (the threshold ε_i and γ_z are chosen exactly to enforce this inequality). Substituting yields

$$\dot{V}_i \leq -\alpha_i \|z_i\|^2, \quad \alpha_i = \frac{\lambda_{\min}(Q_i + K_i^T R_i K_i)}{2}(1 - \sigma_i^2) > 0. \tag{37}$$

Hence $\dot{V}_i < 0$ for $z_i \neq 0$, proving local asymptotic stability. ■For the overall MG (stacked vector $Z = \text{col}\{z_i\}$), the composite Lyapunov function $V(Z) = \sum_i V_i(z_i)$ satisfies $\dot{V} < 0$ because the graph is connected and each local trigger is independent (asynchronous updates do not destabilize the composite system by the small-gain theorem for interconnected systems). Zeno behavior is excluded by $\gamma_z > 0$ (minimum inter-event time $\tau_{\min} \geq \gamma_z / L$, where L is the Lipschitz constant of the Q-dynamics).

To address the convergence of Q-learning policy iteration under sparse updates, let us recall the policy iteration proceeds to solve least squares for $\bar{H}^{l+1}$ from collected $(z_l, u_l = K^l z_l + e_l, z_{l+1})$ (data collected locally at every sample; neighbor $\bar{e}$ held via ZOH). Under sparse neighbor updates, the collected data matrix Q contains a bounded perturbation ΔQ proportional to the holding error e_i. Because the trigger fires precisely when $\|\Delta H\|$ (i.e., policy sub-optimality) exceeds ε_i, the perturbation remains bounded: $\|\Delta Q\| \leq c\varepsilon_i$, where $c > 0$. Let us $J^*(z)$ be the optimal infinite-horizon cost. For the learned policy K^ϵ, the performance satisfies

$$J^\epsilon(z) \leq J^*(z) + \delta(\varepsilon), \tag{38}$$

where the sub-optimality gap is $\delta(\varepsilon) \propto \varepsilon_i/(1-\gamma)$, thestandard result for approximate policy iteration with Bellman residual bounded by ε. Because the trigger (29) stops updates exactly when the Q-deviation drops below ε_i, the learned policy is guaranteed ε to be near-optimal even under asynchronous sparse observations.

4. Results and discussion

In this section, the proposed event-triggered LQG control policy scheme is applied to an islanded AC microgrid.

To this end, the experimental validation of the proposed Q-learning-based ETM scheme was conducted using a real-time hardware platform to verify its practical feasibility. The testbed consists of three Danfoss power converters, each configured as a DC distributed generation unit. The DC prime movers of these converters are energized by a Regatron programmable DC power supply, which emulates dynamic source behavior under load changes. These converters are connected to corresponding DC loads to mimic the structure of a microgrid.

The entire system is controlled via a dSPACE MicrolabBox real-time controller, which executes the event-triggered Q-learning control logic. The control algorithm was implemented in MATLAB/Simulink and compiled to run in real-time on the MicrolabBox platform, enabling closed-loop operation and fast control cycle execution. Communication between units was emulated within the controller to allow precise triggering and synchronization of data exchange events.

Through a number of scenarios executed in a real-time experimental setup shown in Fig. 5(a), the performance of the proposed event-triggered LQG is evaluated for an AC MG with three DGUs is taken into account. The PC and SC layers of each DGU are implemented as discussed in Section 2. The electrical and control parameters for the system are listed in Table 2.

4.1. Secondary control activation

Fig. 6 illustrates the simulation and experimental results of the MG studied in a scenario where the SC is activated at $t = 2$ s, showcasing the enhanced performance due to the proposed learning-based LQG policy scheme. The figure is divided into four subplots, each demonstrating a different aspect of the system's performance. In subplot (a), the voltage response of three DG units is depicted. Before SC activation at $t = 2$ s, the voltages show some deviation. However, once the SC is activated, the voltages converge rapidly and stabilize around the nominal value. This indicates that the SC successfully corrects the voltage deviation and maintains voltage stability among the DG units. Subplot (b) presents the sharing of reactive power between the DG units, such that before the activation of the SC, the sharing of reactive power is unbalanced, but after the activation of the secondary control at $t = 2$, the sharing of reactive power becomes completely balanced with respect to the coefficients of reactive power droop. This balance is critical for maintaining the overall stability and efficiency of the power system, particularly in systems with multiple DG units. Subplot (c) confirms the reactive power sharing through the use of specific indices $n_i Q_i = n_j Q_j$ for all DG units, where the indices represent the proportional sharing of reactive power. After the SC is activated, the indices demonstrate convergence, signifying that all units are contributing proportionally to the reactive power requirements of the system. Subplot (d) highlights the active power sharing before and after the SC activation. The indice $m_i P_i = m_j P_j$ shows that the system maintains

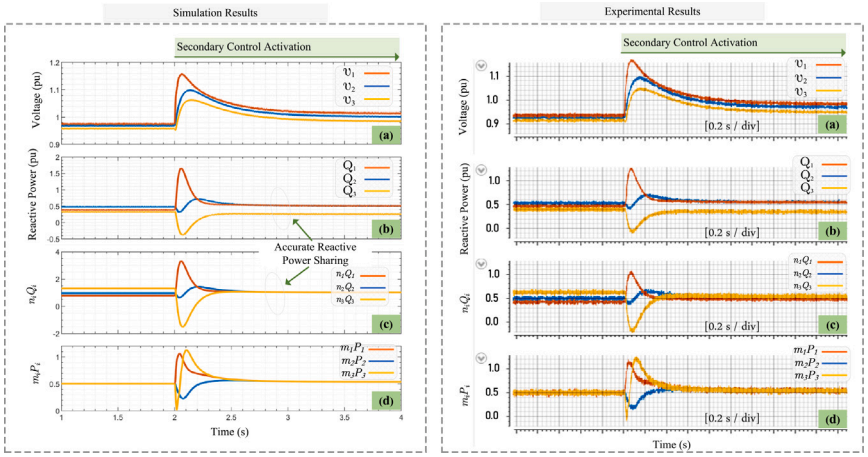

Fig. 6. Simulation and experimental results to show the performance of the proposed learning-based LQG policy scheme, when the SC is activated at $t = 2$ s: (a) voltage of the DG units, (b) reactive power-sharing among DG units (c) reactive power sharing confirmation by depicting the indices $n_i Q_i = n_j Q_j$ among all the DG units, and (d) keeping the active power sharing before and after activating the proposed SC by depicting the indices $m_i P_i = m_j P_j$ among all the DG units.

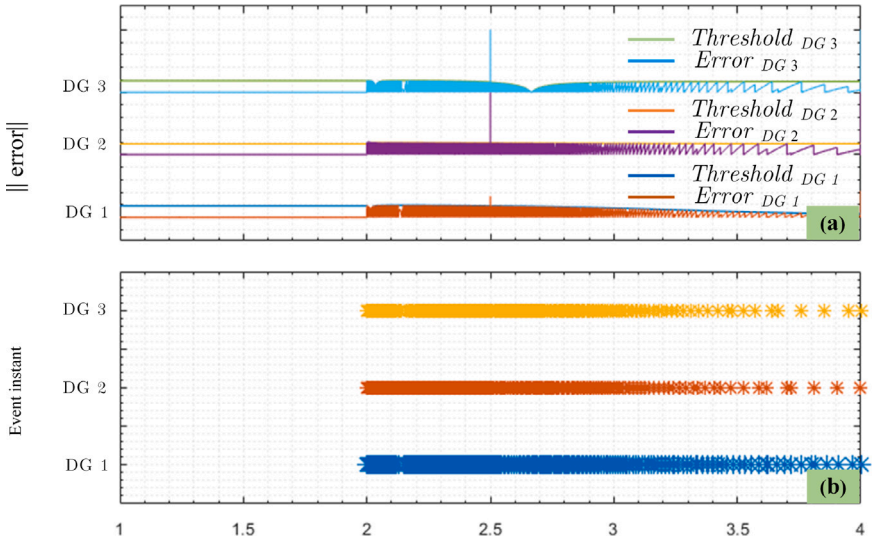

Fig. 7. Error and threshold of the proposed Q-learning-based ETM secondary control: (a) thresholds and errors for each DG unit, and (b) event instants for communication data exchange.

appropriate active power sharing even after the SC is triggered. This is essential for ensuring that the load demand is shared equally across the DG units, preventing overloading or under-utilization of any specific unit. Additionally, as it can be seen from Fig. 7 threshold (28) helps reduce the communication load by only triggering data exchange when necessary, thus mitigating the overuse of CNs without compromising system stability.

Fig. 7 consists of two subplots: in subplot (a), the error and threshold for each of the three DG units are shown over time. Before the activation of the SC, the errors for all DG units are relatively high. The thresholds, which are adaptively regulated based on the system's conditions, also adjust over time. Once SC is activated, the errors quickly decrease and converge to zero, while the thresholds remain stable and maintain values appropriate for each DG unit's performance. Subplot (b) shows the event instants of communication data exchange for each DG unit over time. The proposed learning-based LQG policy scheme reduces unnecessary communication by only triggering data exchange when the error exceeds the threshold for any particular DG unit. As illustrated in this subplot, the frequency of communication events reduces significantly after SC activation, highlighting the efficiency of the proposed scheme in minimizing communication data exchange, which is especially beneficial for systems with limited communication bandwidth. In steady-state conditions, fewer event instants are observed, indicating that the system requires less data transmission once the errors are sufficiently small and remain within the allowable thresholds. The similar dynamic performance seen from the simulation and experimental results confirm how much the provided modeling is accurate for the cases.

4.2. Load change

The experimental results shown in Fig. 8 illustrate the performance of the proposed learning-based LQG policy scheme power sharing strategy in the studied MG. The two provided plots represent the dynamic behavior of the system in terms of voltage regulation and reactive power-sharing during a load increase event, between 6.5 s and 8.5 s. In plot (a), it can be seen that the voltage profiles of the three DG units all initially stabilized close to 1 pu, signifying steady-state system operation. When the load increases at around 6.5 s, the voltages drop momentarily. These deviations are due to the system's immediate reaction to the new load demand, but the ETM mechanism, governed by Q-LbC, quickly activates to restore the voltage levels.

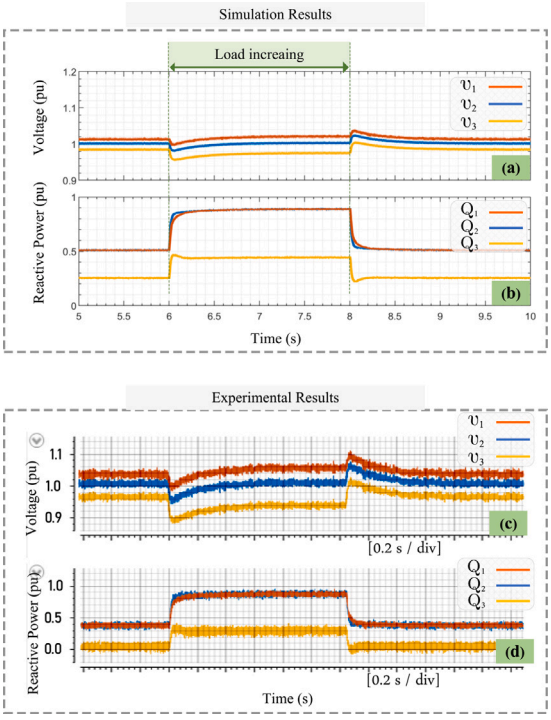

Fig. 8. Voltage regulation and reactive power sharing among DG units (a) Voltage of DG units, and (b) Reactive power sharing among DG units.

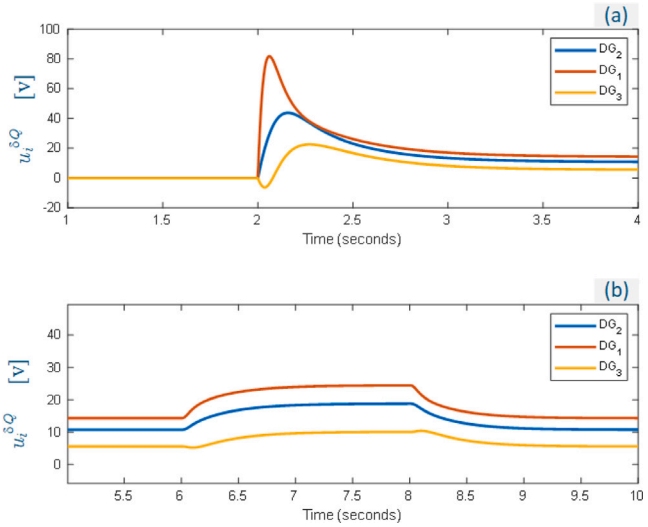

Fig. 9. Control input signals $u_i^{\delta Q}$ of the DG units under (a) secondary control (SC) activation at $t = 2$ s and (b) a load change scenario at $t = 6$ s.

Over the course of the load increase event, the control mechanism ensures that the voltages stabilize back near their nominal value of 1 pu, effectively maintaining system voltage regulation. This adaptive response to the load change demonstrates the learning-based LQG algorithm's capability to adjust control actions in real-time, ensuring system stability even in the face of disturbances. Fig. 8(b), the focus shifts to reactive power sharing dynamics among the three DG units. Before the load increase, the reactive power contributions are relatively low and balanced. However, as the load increases at 6.5 s, the learning-based LQG reacts by adjusting the reactive power outputs of the DGs to meet the new demand. Accurate reactive power contribution reflects the learning-based LQG algorithm's optimization process, which consider factors such as the capacity of each DG when determining their respective contributions. As the load normalizes around 8.5 s, the reactive power levels also stabilize, illustrating the system's ability to effectively share reactive power in a balanced manner post-disturbance. Overall, the results show that the learning-based LQG policy scheme efficiently handles voltage regulation and reactive power sharing, demonstrating its adaptability in real-time to load changes and its effectiveness in dynamically stabilizing the system under varying conditions.

Fig. 9 shows the control inputs $u_i^{\delta Q}$ for the three DG units during (a) SC activation and (b) a load disturbance. In (a), the control inputs rise sharply right after the SC is activated at $t = 2$ s, then gradually settle as the system reaches stability. In (b), a load change at $t = 6$ s causes a smooth and proportional adjustment in the control signals. The responses remain stable and bounded in both cases, confirming the effectiveness and robustness of the proposed Q-learning-based controller in managing both transitions and disturbances without excessive control action.

4.3. Effects of communication delay

The experimental results depicted in Fig. 5 present a comprehensive evaluation of the learning-based LQG policy scheme in the SC layer. These results explore the impact of two different time delays (100 ms and 200 ms) on the SC's ability to maintain voltage regulation and effectively manage the reactive power sharing. The study aims to determine how the learning-based LQG performs in real-time disturbance scenarios like the effect of communication delays among different agents. The experiments assess four key metrics: voltage regulation, reactive power distribution among DG units, reactive power sharing indices, and active power sharing indices. Each of these metrics provides insights into the efficiency and robustness of the control strategy in scenarios where communication delays, a common issue in distributed systems, are present. The first set of results, displayed in the first subplot of Fig. 10(a), presents the voltage profiles of all the DG units during the load change event for the 100 ms delay. In the 100 ms delay scenario, the same load change at approximately 6.5 s results in a short-lived fluctuation in the voltage profiles of all three DG units. These transient deviations are expected during communication delays

In contrast, the 200 ms delay case in Fig. 11 presents more oscillations. The recovery time for mitigating the oscillations due to 200 ms time delayv is significantly large, and the transient fluctuations in the reactive power-sharing are more pronounced. The extended communication delay results in delayed control actions, leading to longer periods of oscillations following the load disturbance. Although the system eventually restores the voltage levels to their nominal values, the slower response time increases the risk of transient instability, especially in more complex or larger MGs. This underscores the importance of minimizing communication data exchanges in MG control systems, as longer delays directly impact the system's ability to quickly stabilize. Subplots (c) and (d) of Figs. 10 and 11 delve into the confirmation of reactive power and active power sharing using proportional indices.

From a comparison point of the view, during the load change event, there is a significant increase in the reactive power demand, which the system must distribute proportionally across the DG units to maintain stable voltage levels and power quality. In the 200 ms delay scenario, the system faces more significant oscillations in reactive power sharing during the initial load change event. This finding highlights the importance of the event-triggered mechanism in mitigating the adverse effects of communication delays. By focusing control actions only when necessary, the learning-based LQG approach prevents excessive reactive power fluctuations, even in scenarios with significant delays, ensuring that the system remains operationally stable and that power-sharing goals are met.

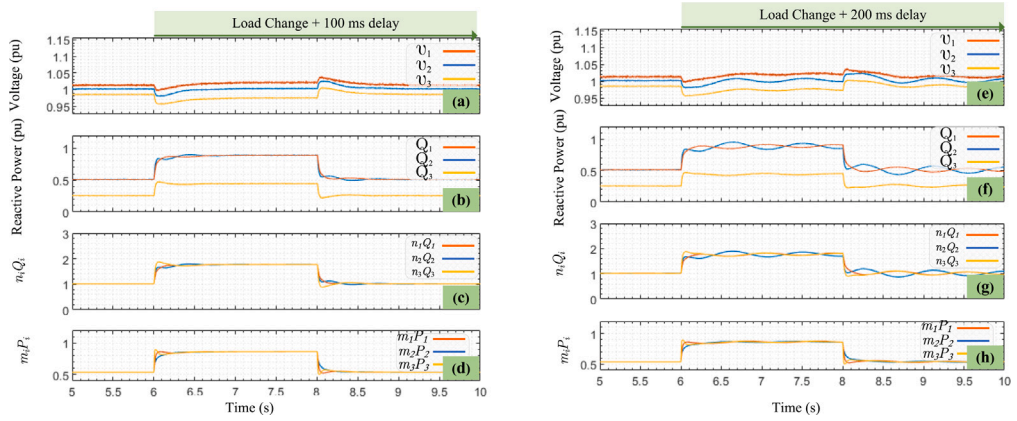

Fig. 10. Simulation results to show the performance of the proposed Q-learning SC enhanced by the ETM scheme, when different time delays are applied with a load change. Left subplots are for 100 ms time delay and right subplots are for 200 ms time delay. From top to bottom of both subplots, waveforms show: (a) voltage of the DG units, (b) reactive power-sharing among DG units (c) reactive power sharing confirmation by depicting the indices $n_iQ_i = n_jQ_j$ among all the DG units, and (d) keeping the active power sharing by depicting the indices $m_iP_i = m_jP_j$ among all the DG units.

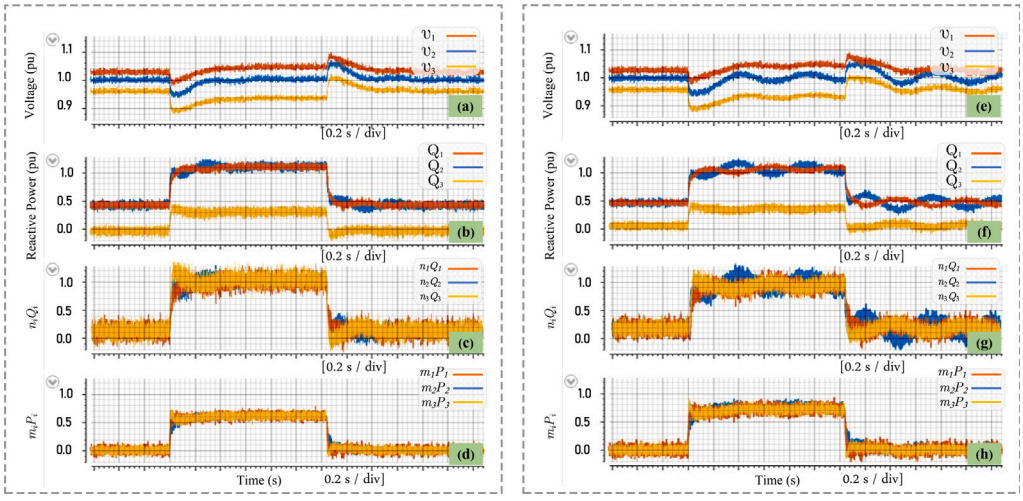

Fig. 11. Experimental results to show the performance of the proposed Q-learning SC enhanced by the ETM scheme, when different time delays are applied with a load change. Left subplots are for 100 ms time delay and right subplots are for 200 ms time delay. From top to bottom of both subplots, waveforms show: (a) voltage of the DG units, (b) reactive power-sharing among DG units (c) reactive power sharing confirmation by depicting the indices $n_iQ_i = n_jQ_j$ among all the DG units, and (d) keeping the active power sharing by depicting the indices $m_iP_i = m_jP_j$ among all the DG units.

4.4. Comparison with other approaches

Fig. 12 provides a quantitative benchmark of the proposed EMT Q-learning LQG against three state-of-the-art schemes:(1) conventional continuous time-triggered LQG (fixed gain $K = -(R + B^T PB)^{-1} B^T PA$ from the algebraic Riccati equation), (2) PSO-tuned ETM LQG (Feng and Wang, 2021), and (3) time-triggered LQG (Mejia-Ruiz et al., 2022) under identical microgrid conditions (same load profiles, initial states, switching instants, and disturbance sequences). The comparison is structured along three defined metrics: maximum overshoot $\max_t |y(t) - y_{\text{ref}}|$ (%), 10% settling time t_s (time after which $|y(t) - y_{\text{ref}}| \leq 0.1|y_{\text{ref}}|$ for all subsequent t), and total communication events (number of transmitted packets).

The overshoot subplot reveals that the proposed method yields the lowest peak deviations (typically $\prec$ 12% for voltage and reactive-power channels) because the learned gain matrix $K^l = -H_{uu}^{-1} H_{ux}$ is continuously refined via policy iteration on real input data, automatically minimizing the quadratic cost $J = \sum \gamma^k (z_k^T Q z_k + u_k^T R u_k)$ at each operating point. In contrast, the fixed gain time-triggered LQG exhibits the highest overshoot (up to 45%) since its K is optimal only for the nominal linearized model and cannot adapt to feeder-impedance variations or load steps. The PSO-based scheme, while optimized offline, remains conservative because the particle-swarm cost function does not directly incorporate the infinite-horizon Bellman residual, leading to intermediate overshoot values (≈28%–35%).

Settling-time performance further underscores the advantages that the proposed controller converges in 120–180 ms in most cases, driven by the Q-value-deviation trigger $\|Q_K^t - Q_K^{t-1}\|_F + \gamma_z > \varepsilon_i$, which fires precisely when the policy sub-optimality gap grows, ensuring rapid re-optimization without unnecessary updates. The continuous LQG requires 600–900 ms because every sample is processed regardless of error magnitude, while the PSO-ETM and time-triggered variants settle in 450–700 ms due to either overly frequent triggering or static thresholds that ignore policy evolution. The Lyapunov analysis (Section 4.5) confirms that the adaptive H keeps the closed-loop eigenvalues inside a tighter stability margin than any fixed-gain benchmark.

Communication efficiency is quantified in the final subplot: the proposed scheme generates only 280–420 events across the 4 s test window (leads to 68%–74% reduction), as the trigger directly monitors policy change rather than the state error. Time-triggered methods transmit at every 10 ms tick (≈1800 events), while PSO-ETM still requires 1100–1400 events because its threshold is state-based and non-adaptive. This reduction directly lowers cyber-layer load while preserving stability, guarantees under 100–200 ms delays.

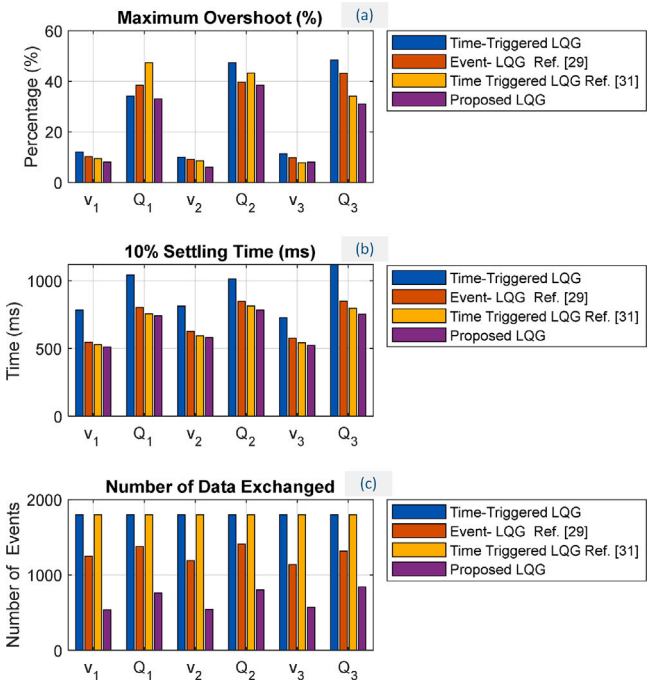

Fig. 12. Comparison results to illustrate the performance of the proposed Q-learning SC enhanced by the ETM scheme, with a time triggered LQG (employs a continuous communication infrastructure with the bandwidth of 300 Hz), and a PSO-based LQG scheme (Feng and Wang, 2021), and the LQG manner in Mejia-Ruiz et al. (2022). Reactive power sharing is not met in Feng and Wang (2021). (a) Maximum overshoots, (b) settling time values, and (c) number of the events for exchanging data in the communication layer for the voltage and reactive power quantities of each DG unit.

All simulations used identical conditions and benchmark tuning approach. The observed superiority therefore based purely on the model-free policy iteration combined with Q-value triggering-features absent in the fixed-gain or PSO-tuned alternatives—making the proposed controller the preferred solution for communication-constrained, uncertainty-rich islanded microgrids.

4.5. Discussions

4.5.1. Sensitivity of the Q-learning policy

Regarding the sensitivity of the learned policy to the modeling parameters across different operating conditions, it is worth to highlight that the learned policy $K = -H_{uu}^{-1} H_{ux}$ is obtained from the converged Q-matrix $\bar{H}$ for the infinite-horizon quadratic cost

$$J = \sum_{k=0}^{\infty} \gamma^k \left(z_k^T Q z_k + u_k^T R u_k \right),$$

where z_k is the augmented state (local voltage error + consensus errors $\bar{e}_{v,i}$, $\bar{e}_{Q,i}$) and $Q = \mathrm{diag}(Q_{vv}, Q_{QQ}, Q_{eQ}, Q_{eV})$, $R = \mathrm{diag}(R_v, R_Q)$.

To quantify sensitivity analytically, consider a perturbation ΔQ in the weighting matrices. The resulting perturbation in the optimal gain satisfies the first-order Lyapunov sensitivity equation derived from the ARE:

$$\Delta K \approx -(R + B^T P B)^{-1} B^T \left(\Delta P A + P \Delta A \right),$$

where ΔP is the solution of the perturbed Lyapunov equation

$$(A_{cl}^T \Delta P + \Delta P A_{cl}) = -(\Delta Q + K^T \Delta R K).$$

Because the closed-loop matrix A_{cl} remains Hurwitz (guaranteed by the initial stabilizing policy and the trigger condition), $\|\Delta P\| \le \kappa \|\Delta Q\| / \alpha$, with $\kappa = \|P\| / \lambda_{\min}(Q + K^T R K)$ and $\alpha > 0$ the stability margin.

In the present design, the dominant term is the consensus weighting ratio $\rho = Q_{eQ}/Q_{vv}$. Differentiating the policy improvement step with respect to ρ yields

$$\frac{\partial K}{\partial \rho} = -H_{uu}^{-1} \left(\frac{\partial H_{ux}}{\partial \rho} \right) \approx -H_{uu}^{-1} \left(\gamma \sum_{j \in \mathcal{N}_i} a_{ij} B_e^T P B_e \right),$$

where B_e is the input matrix column corresponding to the consensus error channels.

4.5.2. Communication graph topology and delays

Regarding the directed graph topology and its effect, it is worth to note that by replacing the symmetric adjacency matrix A_G by a row-stochastic matrix $D^{-1} A_G$ (standard in directed consensus). The consensus errors become

$$\bar{e}_{Q,i} = \sum_{j \in \mathcal{N}_i} d_{ij} (n_{Q,i} Q_i - n_{Q,j} Q_j), \quad d_{ij} \text{ row-stochastic.}$$

The same Q-learning iteration converges because the stacked Laplacian still possesses a simple zero eigenvalue with left eigenvector $\mathbf{1}^T$ (provided the digraph contains a spanning tree). The Lyapunov analysis of Section 4.5 carries over directly with the same σ-bound on the holding error, since only the input matrix B_e changes (its norm remains bounded).

(i) Time-varying topologies: Let $\mathcal{G}(t)$ be piecewise constant with minimum dwell time $\tau_D \ge 50$ ms (larger than the sampling period). The closed-loop system becomes a switched linear system $\dot{z} = A_{cl,\sigma(t)} z + B K e(t)$, where $\sigma(t)$ is the active topology index. Because each $A_{cl,\sigma}$ is Hurwitz with a common quadratic Lyapunov function $V(z) = z^T P z$ (constructed from the worst-case topology via LMI feasibility, which holds under the chosen $\rho = 500$), global uniform asymptotic stability follows from the common Lyapunov function theorem. The Q-value trigger (29) remains valid because $\|\Delta \bar{H}\|$ is evaluated locally and is independent of the instantaneous topology. (ii) Partially disconnected graphs: When the communication graph loses connectivity, each connected component runs its own independent Q-learning instance with fallback to pure droop control ($u_i^{\delta Q} = 0$). The semi-periodic ETM allows reconnection detection. Convergence inside each component is preserved by the same arguments as the connected case; global RPS degrades gracefully to the droop level $n_{Q,i} Q_i$ (already demonstrated in the pre-secondary-control region of Fig. 6). Reconnection instantly restarts policy iteration with the last stored $\bar{H}$, guaranteeing finite-time recovery of exact RPS.

4.5.3. Limitations and future works

As mentioned earlier, islanded MGs face vulnerabilities due to low inertia, weak grid characteristics, unknown line impedances, and dynamic loads, leading to unbalanced power flows and instability. The proposed approach integrates a Q-learning to adaptively determine optimal control policies based on Q-values, enhanced by ETM, it achieves physical layer objectives (RPS and voltage regulation) while reducing cyber layer data exchanges. Main limitations of existing work can be given by the learning transients, need for bounded delays, and dependency for Q-function shaping. The main future directions can be integration of adaptive RL, topology-aware triggering, moving to nonlinear Q-learning (e.g., deep LQG), and investigating it for switching communication graphs.

5. Conclusion

In this paper, a distributed learning-based linear quadratic gaussian (LQG) policy scheme enhanced by an ETM mechanism for employing in the SC level of autonomous MGs was developed to simultaneously fulfill both the operational objectives in the physical layer of the MG and the reduction in data exchange at the cyber level of MG. Unlike LQG schemes, the proposed Q-learning-based LQG control

scheme effectively enhances reactive power-sharing and voltage regulation in microgrids while reducing communication data exchange through event-triggered mechanisms. Our results demonstrate significant improvements in system stability and performance, highlighting the potential of integrating reinforcement learning techniques into hierarchical control structures. Future work should address the limitations related to communication delays and training requirements, paving the way for more resilient and efficient microgrid control strategies. Uncertainty in the grid impedance and load changes in autonomous MGs is always known as a challenging factor for voltage requirements and reactive power sharing; addressing this challenge increases the importance of this feature of the Q-learning based LQG, even with communication time delays. Simulation and experimental results showed the effectiveness of the scheme through a laboratory-scale MG set-up and show that the proposed learning based LQG policy scheme efficiently handles voltage regulation and reactive power sharing, demonstrating its adaptability in real-time applications.

CRediT authorship contribution statement

Salam Nikravesh: Writing – original draft, Software, Methodology, Investigation, Conceptualization. **Amin Karimi:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization. **Karvan Karimizadeh:** Writing – review & editing, Project administration, Methodology. **Yousef Khayat:** Writing – review & editing, Validation, Software, Methodology, Conceptualization. **Saeed Golestan:** Writing – review & editing, Validation, Methodology, Conceptualization.

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.

Data availability

Data will be made available on request.

References

Ahmed, K., Seyedmahmoudian, M., Mekhilef, S., Mubarak, N., Stojcevski, A., 2020. A review on primary and secondary controls of inverter-interfaced microgrid. J. Mod. Power Syst. Clean Energy 9 (5), 969–985.

Bevrani, H., François, B., Ise, T., 2017. Microgrid Dynamics and Control. John Wiley & Sons.

Bidram, A., Davoudi, A., Lewis, F.L., Guerrero, J.M., 2013. Distributed cooperative secondary control of microgrids using feedback linearization. IEEE Trans. Power Syst. 28 (3), 3462–3470.

Chen, M., Xiao, X., 2018. Secondary voltage control in islanded microgrids using event-triggered control. IET Gener. Transm. Distrib. 12 (8), 1872–1878.

Chen, M., Xiao, X., Guerrero, J.M., 2017. Secondary restoration control of islanded microgrids with a decentralized event-triggered strategy. IEEE Trans. Ind. Inform. 14 (9), 3870–3880.

Dehkordi, N.M., Baghaee, H.R., Sadati, N., Guerrero, J.M., 2018. Distributed noise-resilient secondary voltage and frequency control for islanded microgrids. IEEE Trans. Smart Grid 10 (4), 3780–3790.

Dehkordi, N.M., Nekoukar, V., 2025. Adaptive distributed stochastic deep reinforcement learning control for voltage and frequency restoration in islanded AC microgrids with communication noise and delay. Sci. Rep. 15 (1), 27315.

Ding, L., Han, Q.-L., Zhang, X.-M., 2018. Distributed secondary control for active power sharing and frequency regulation in islanded microgrids using an event-triggered communication mechanism. IEEE Trans. Ind. Inform. 15 (7), 3910–3922.

Falope, T., Lao, L., Hanak, D., Huo, D., 2024. Hybrid energy system integration and management for solar energy: A review. Energy Convers. Manag.: X 100527.

Fan, Y., Hu, G., Egerstedt, M., 2016. Distributed reactive power sharing control for microgrids with event-triggered communication. IEEE Trans. Control Syst. Technol. 25 (1), 118–128.

Feng, Y., Wang, X., 2021. Microgrid event trigger optimization and control based on asynchronous-sampling data system approach. In: 2021 33rd Chinese Control and Decision Conference. CCDC, IEEE, pp. 1431–1438.

Ge, X., Han, Q.-L., Ding, L., Wang, Y.-L., Zhang, X.-M., 2020. Dynamic event-triggered distributed coordination control and its applications: A survey of trends and techniques. IEEE Trans. Syst. Man Cybern.: Syst. 50 (9), 3112–3125.

Guo, Q., Cai, H., Wang, Y., Chen, W., 2017. Distributed secondary voltage control of islanded microgrids with event-triggered scheme. J. Power Electron. 17 (6), 1650–1657.

Islam, M.A., Ali, M.N., Mollick, T., Islam, A., Benitez, I.B., Habib, S.S., Al Mansur, A., Lipu, M.S.H., Flah, A., et al., 2024. Assessing the feasibility and quality performance of a renewable energy-based hybrid microgrid for electrification of remote communities. Energy Convers. Manag.: X 23, 100674.

Jafari, M., Sarfi, V., Ghasemkhani, A., Livani, H., Yang, L., Xu, H., 2019. Adaptive intelligent secondary control of microgrids using a biologically-inspired reinforcement learning. In: 2019 IEEE Power & Energy Society General Meeting. PESGM, IEEE, pp. 1–5.

Kang, W., Guan, Y., Wang, H., Vasquez, J.C., Agundis-Tinajero, G.D., Guerrero, J.M., 2024. Distributed control of virtual energy storage systems for voltage regulation in low voltage distribution networks subjects to varying time delays. Appl. Energy 376, 124295.

Khayat, Y., Naderi, M., Shafiee, Q., Batmani, Y., Fathi, M., Guerrero, J.M., Bevrani, H., 2018. Decentralized optimal frequency control in autonomous microgrids. IEEE Trans. Power Syst. 34 (3), 2345–2353.

Khayat, Y., Shafiee, Q., Heydari, R., Naderi, M., Dragičević, T., Simpson-Porco, J.W., Dörfler, F., Fathi, M., Blaabjerg, F., Guerrero, J.M., et al., 2019. On the secondary control architectures of AC microgrids: An overview. IEEE Trans. Power Electron. 35 (6), 6482–6500.

Li, X., Hu, C., Luo, S., Lu, H., Piao, Z., Jing, L., 2025. Distributed hybrid-triggered observer-based secondary control of multi-bus DC microgrids over directed networks. IEEE Trans. Circuits Syst. I. Regul. Pap. 72 (5), 2467–2480.

Lu, J., Zhao, M., Golestan, S., Dragicevic, T., Pan, X., Guerrero, J.M., 2021. Distributed event-triggered control for reactive, unbalanced, and harmonic power sharing in islanded AC microgrids. IEEE Trans. Ind. Electron. 69 (2), 1548–1560.

Mahmood, M., Chowdhury, P., Yeassin, R., Hasan, M., Ahmad, T., et al., 2024. Impacts of digitalization on smart grids, renewable energy, and demand response: An updated review of current applications. Energy Convers. Manag.: X 100790.

Mahmud, M.A., Hossain, M., Pota, H., Oo, A., 2014. Robust nonlinear distributed controller design for active and reactive power sharing in islanded microgrids. IEEE Trans. Energy Convers. 29 (4), 893–903.

Mejia-Ruiz, G.E., Cárdenas-Javier, R., Paternina, M.R.A., Rodríguez-Rodríguez, J.R., Ramirez, J.M., Zamora-Mendez, A., 2021. Coordinated optimal volt/var control for distribution networks via d-pmus and ev chargers by exploiting the eigensystem realization. IEEE Trans. Smart Grid 12 (3), 2425–2438.

Mejia-Ruiz, G.E., Paternina, M.R.A., Sevilla, F.R.S., Korba, P., 2022. Fast hierarchical coordinated controller for distributed battery energy storage systems to mitigate voltage and frequency deviations. Appl. Energy 323, 119622.

Mohammadi, K., Azizi, E., Choi, J., Hamidi-Beheshti, M.-T., Bidram, A., Bolouki, S., 2021. Asynchronous periodic distributed event-triggered voltage and frequency control of microgrids. IEEE Trans. Power Syst. 36 (5), 4524–4538.

Qian, T., Liu, Y., Zhang, W., Tang, W., Shahidehpour, M., 2019. Event-triggered updating method in centralized and distributed secondary controls for islanded microgrid restoration. IEEE Trans. Smart Grid 11 (2), 1387–1395.

Sen, S., Kumar, V., 2020. Decentralized output-feedback-based robust LQR Vf controller for PV-battery microgrid including generation uncertainties. IEEE Syst. J. 14 (3), 4418–4429.

Suresh, A., Bisht, R., Kamalasadan, S., 2020. ADMM based LQR for voltage regulation using distributed energy resources. In: 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems. PEDES, IEEE, pp. 1–6.

Wan, X., Tian, Y., Wu, J., Ding, X., Tu, H., 2021. Distributed event-triggered secondary recovery control for islanded microgrids. Electronics 10 (15), 1749.

Wang, J., Xu, W., Gu, Y., Song, W., Green, T.C., 2021. Multi-agent reinforcement learning for active voltage control on power distribution networks. Adv. Neural Inf. Process. Syst. 34, 3271–3284.

Weng, S., Yue, D., Dou, C., Shi, J., Huang, C., 2018. Distributed event-triggered cooperative control for frequency and voltage stability and power sharing in isolated inverter-based microgrid. IEEE Trans. Cybern. 49 (4), 1427–1439.

Yan, S., Gu, Z., Ding, L., Park, J.H., Xie, X., 2023a. Weighted memory H_∞ stabilization of time-varying delayed Takagi-Sugeno fuzzy systems. IEEE Trans. Fuzzy Syst..

Yan, S., Gu, Z., Park, J.H., Xie, X., 2022. Sampled memory-event-triggered fuzzy load frequency control for wind power systems subject to outliers and transmission delays. IEEE Trans. Cybern..

Yan, S., Gu, Z., Park, J.H., Xie, X., 2023b. A delay-kernel-dependent approach to saturated control of linear systems with mixed delays. Automatica 152, 110984.

Yang, C., Yao, W., Fang, J., Ai, X., Chen, Z., Wen, J., He, H., 2019. Dynamic event-triggered robust secondary frequency control for islanded AC microgrid. Appl. Energy 242, 821–836.