Preprints.org

Article

Not peer-reviewed version

# Design and Evaluation of PPO-Based Coordinated P–Q Control for a Smart-Inverter BESS in PV–EV Distribution Networks

Thunyaput Kittiphotiklang , Praphat Phuangchuen , Pantree Khompittaya , Prakasit Prabpal , Somchat Sonasang , Niwat Angkawisittpan , Thananan Chooseang * , Krischonmee Bhumkittipich

Posted Date: 19 October 2026

Keywords: smart-inverter BESS; P–Q control; voltage unbalance; proximal policy optimization; deep reinforcement learning; electric vehicles; distribution network; battery energy storage system

Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

*Article*

# Design and Evaluation of PPO-Based Coordinated P–Q Control for a Smart-Inverter BESS in PV–EV Distribution Networks

**Thunyaput Kittiphotiklang [1],Praphat Phuangchuen [2], Pantree Khompittaya [3], Prakasit Prabpal [4], Somchat Sonasang [5], Niwat Angkawisittpan [6], Thananan Chooseang [7,*] and Krischonmee Bhumkittipich [8]**

[1] Department of Electrical Technology, Institute of Vocational Education Eastern Region, Rayong 21000, Thailand
[2] Department of Mechanical Technology, Institute of Vocational Education Eastern Region, Rayong 21000, Thailand
[3] Department of Electronics and Robotics, That Phanom College, Nakhon Phanom University, Nakhon Phanom 48110, Thailand
[4] Department of Electrical Technology, Udon Thani Technical College, Institute of Vocational Education Northeastern Region 1, Udon Thani 41000, Thailand
[5] Electrical Elements and Wireless Devices Unit (EEWDU), Faculty of Industrial Technology, Nakhon Phanom University, Nakhon Phanom 48000, Thailand
[6] Research Unit for Electrical and Computer Engineering Technology (RECENT), Faculty of Engineering, Mahasarakham University, Maha Sarakham 44150, Thailand
[7] Department of Electronic Technology, Pathum Thani Technical College, Institute of Vocational Education Central Region 1, Pathum Thani 12000, Thailand
[8] Department of Electrical Engineering, Faculty of Engineering, Rajamangala University of Technology Thanyaburi (RMUTT), Pathum Thani, Thailand
* Correspondence: thananan@pttc.ac.th

## Abstract

Phase-asymmetric electric-vehicle (EV) charging can aggravate voltage unbalance in radial distribution feeders, while a smart-inverter battery energy storage system (BESS) can provide both energy shifting and fast reactive-power support. This study designs and evaluates a proximal policy optimization (PPO)-based controller for coordinated active- and reactive-power dispatch under asymmetric EV demand. The analysis uses a corrected phase-decoupled three-phase extension of the IEEE 33-bus feeder, with voltage unbalance quantified using the negative-to-positive sequence voltage ratio. Active-power-only (P-only) and coordinated P–Q controllers are evaluated using identical BESS hardware, operating limits, and terminal state-of-charge (SOC) requirements, enabling the incremental contribution of inverter reactive-power capability to be assessed under matched conditions. Across three independent training seeds, P–Q control reduces daily feeder loss from 3874.822 to 3809.751 kWh/day, worst voltage-unbalance factor (VUF) from 2.7879% to 2.5940%, and mean hourly maximum VUF from 1.5422% to 1.3628%, corresponding to reductions of 1.679%, 6.954%, and 11.632%, respectively, relative to the no-BESS case. The corresponding P-only reductions are 0.610%, 4.362%, and 8.427%, respectively. All reported learned-policy evaluations return the BESS to 50% terminal SOC, preventing the comparison from being biased by net battery depletion. Without retraining, the P–Q policies retain approximately 5.66–7.86% mean worst-VUF reduction under the tested load, photovoltaic, EV-demand, and combined-profile perturbations. The results demonstrate repeatable voltage-unbalance mitigation and show that coordinated P–Q operation provides an incremental benefit beyond active-power-only control under the investigated conditions. However, residual undervoltage under severe loading and the phase-decoupled feeder representation limit the scope of the conclusions.

**Keywords:** smart-inverter BESS; P–Q control; voltage unbalance; proximal policy optimization; deep reinforcement learning; electric vehicles; distribution network; battery energy storage system

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



## 1. Introduction

Electric-vehicle (EV) charging can become a significant and phase-asymmetric load in distribution feeders, particularly when charging demand is temporally concentrated. The network impact depends not only on the energy required by the vehicles, but also on when and where charging occurs and how the chargers are distributed among phases. Evening clustering can increase feeder losses and deepen voltage drops, while unequal single-phase or phase-selective charging can add a negative-sequence component that is not captured by an average voltage-magnitude index alone. Earlier studies have examined the network impact of EV charging and the value of coordinated scheduling. In particular, Chitgreeyan *et al.* treated multi-period EV demand control in an unbalanced distribution system and considered temporal load shifting together with feeder losses and voltage unbalance [1]. This provides a useful starting point for the present work, where the EV demand is again phase dependent but the controllable resource is a grid-connected BESS rather than the charging schedule itself.

A BESS can influence the feeder in two different ways. Active power changes the net energy exchanged with the upstream network and couples decisions across time through the battery state of charge (SOC). Reactive power, by contrast, can be supplied or absorbed by the inverter without an equivalent change in stored energy, subject to the converter apparent-power limit. This second degree of freedom is relevant when voltage quality is part of the control objective. Kabir *et al.* showed how photovoltaic reactive power and battery storage can be coordinated to improve feeder voltage profiles [2], while Prabpal *et al.* examined BESS sizing and placement together with VAR-control strategies in a distribution system [3]. Work from the same research group has also considered EV–BESS–V2G operation in an IEEE 13-bus microgrid [4]. These studies establish the electrical motivation for combining storage and converter support, but they do not answer the specific question considered here: how much additional benefit is obtained when the same BESS is allowed to coordinate both $P$ and $Q$ in response to phase-asymmetric EV loading.

The control problem is sequential and constrained. A dispatch that improves the feeder at one hour can reduce the energy available later, and an apparent improvement over a 24-h horizon can be misleading if the battery is allowed to finish the day at a lower SOC. Conventional voltage-control research has addressed this type of problem through distributed optimization, robust Volt/VAR control, coordinated inverter control, and multi-timescale scheduling [5–11]. Deep reinforcement learning (DRL) offers a different route by learning a policy from repeated interaction with the network model. PPO is attractive for continuous control because it constrains policy updates through a clipped surrogate objective while retaining a relatively simple implementation [12]. In distribution systems, DRL has already been applied to Volt/VAR optimization, safe inverter control, model-assisted voltage regulation, and EV-aware operation [13–17]. More recent work has emphasized safety layers, robustness, multi-agent coordination, and physics-informed EV charging [18–22].

The literature therefore makes two points clear. First, reactive-power flexibility can materially affect distribution-network voltage behavior. Second, a learning-based controller should be evaluated on engineering quantities and operating constraints, not only on accumulated training reward. What is less often isolated is the incremental value of reactive-power control when the hardware, battery energy, active-power limit, SOC range, and terminal SOC requirement are held fixed. This distinction is important for an unbalanced EV case because a comparison between controllers with different energy budgets or terminal conditions cannot show whether an observed improvement comes from the control strategy or simply from greater use of the battery.

This paper addresses that narrower question using a corrected phase-decoupled three-phase extension of the IEEE 33-bus feeder. Three cases are evaluated under the same network and operating profiles: no BESS, active-power-only PPO (P-only), and coordinated active–reactive-power PPO (P–Q). Voltage unbalance is calculated from symmetrical components as $|V_2|/|V_1|$, and the assessment reports daily feeder loss, worst VUF, mean hourly maximum VUF, minimum voltage, SOC feasibility, variation across three independent training seeds, and performance under perturbed load, PV, and EV profiles. The purpose is not to claim complete voltage restoration or a universally optimal DRL architecture.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



Instead, the study isolates the additional network benefit associated with using the reactive-power capability of the same smart-inverter BESS under the tested EV-loading conditions.

The main contributions are:

1. a phase-dependent EV-charging case in which voltage unbalance is evaluated from symmetrical components rather than from an ad-hoc phase-magnitude deviation;
2. a matched P-only versus P–Q PPO comparison using identical BESS energy, power, inverter, SOC, and terminal-energy constraints;
3. joint reporting of feeder loss, worst VUF, mean hourly maximum VUF, minimum voltage, and terminal SOC over three independently trained seeds; and
4. a no-retraining robustness evaluation under load ±10%, PV ±20%, EV ±20%, and combined favorable and stress profiles.

## 2. Related Work and Design Rationale

The literature relevant to this study can be organized into four closely connected groups: EV-charging impact and coordination, inverter-based voltage control, BESS-assisted distribution-system control, and reinforcement-learning-based Volt/VAR regulation. These research directions address different parts of the same engineering problem. EV charging introduces additional and potentially phase-asymmetric demand; inverter-based resources provide fast active- and reactive-power flexibility; battery storage introduces temporal energy coupling through the state of charge (SOC); and reinforcement learning offers a data-driven mechanism for coordinating these variables over time. Table 1 summarizes representative studies that are most closely related to the formulation considered in this work.

**Table 1.** Representative State of the Art and Position of the Present Study

| Study | Main resource/control | Primary network objective | Learning/solution approach | Relation to the present work |
|---|---|---|---|---|
| Zhang *et al.* [9] | PV inverter + utility devices | Voltage/VAR under PV uncertainty | Robust optimization | Establishes coordinated inverter VAr control; not EV-driven sequence VUF. |
| Jha *et al.* [10] | Smart inverters + legacy devices | CVR and voltage control | Bi-level optimization | Detailed unbalanced feeder optimization; no learned BESS policy. |
| Wei *et al.* [14] | Distribution VVC devices | Voltage and loss | Safe off-policy DRL | Strong safety formulation; not focused on asymmetric EV charging. |
| Gao and Yu [15] | VVC devices | Voltage and loss | Model-augmented safe RL | Addresses sample efficiency and constraint satisfaction. |
| Cui *et al.* [27] | Inverter-based DER | Voltage stability | Decentralized safe RL | Provides stability-oriented inverter learning; no BESS energy coupling. |
| Liu *et al.* [17] | EV charging + Volt/VAR devices | Charging cost and voltage magnitude | DDPG | Direct EV/voltage-control precedent; sequence VUF is not the central metric. |
| Xiang *et al.* [29] | Distributed storage | Voltage regulation | Topology-aware MADRL | Uses storage as autonomous voltage-control agents; different disturbance and objective. |
| Huang *et al.* [18] | PV inverters | Voltage and loss | Collaborative MARL | Demonstrates scalable multi-agent voltage control in larger feeders. |
| Shen *et al.* [25] | Reactive-power resources + EV forecast | Voltage/loss under EV demand | DRL | Closely related EV-driven reactive optimization; emphasis is forecasting and VVO. |
| Wu *et al.* [20] | PV + EVCS | Volt/VAR optimization | Safe RL | Coordinates PV and EVCS flexibility with safety constraints. |
| Orfanoudakis *et al.* [22] | EV/V2G fleet | Voltage-constrained charging | Physics-informed TD3 | Strong grid-aware EV RL benchmark; focuses on charging coordination and voltage constraints. |
| **This work** | **Single smart-inverter BESS** | **Sequence VUF, loss, voltage, terminal SOC** | **PPO; P-only vs. P–Q** | **Same BESS hardware in both learned cases, three seeds, terminal-SOC matching, and no-retraining profile perturbations.** |

EV charging can substantially alter the operating condition of a distribution feeder, particularly when charging demand is concentrated within a limited time window or unevenly distributed among

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



phases. Previous studies have shown that unmanaged or weakly coordinated charging can increase feeder peak demand, deepen local voltage drops, increase network losses, and create additional operating stress [17,23,24]. Consequently, EV coordination has been studied using deterministic optimization, transactive control, prediction-assisted scheduling, and DRL. Liu *et al.*, for example, coupled EV charging scheduling with distribution-network voltage control using DDPG on the IEEE 33-bus system [17]. More recent studies have extended this direction through EV-load forecasting, coordinated Volt/VAR control, ESS–PV scheduling, safe reinforcement learning, and physics-informed charging strategies [20,22,25,26]. These contributions improve charging coordination and voltage management, but their primary performance indicators are commonly voltage magnitude, energy cost, charging objective, or aggregate operational performance. Sequence-based voltage unbalance arising from explicitly asymmetric phase loading is less frequently used as the principal control metric.

A second research direction concerns voltage regulation using smart inverters and coordinated reactive-power support. This body of literature has established the importance of reactive power as a fast and locally effective control variable in active distribution systems. Distributed voltage control, feedback-based reactive-power regulation, robust inverter VAr dispatch, and coordinated operation of smart inverters with storage and conventional voltage-control devices have all been reported [2,5–11]. These studies provide the electrical basis for coordinated $P$–$Q$ control. In particular, active power and reactive power affect the system through different physical mechanisms. Active-power dispatch changes feeder loading and simultaneously modifies battery energy and future SOC availability, whereas reactive-power dispatch can provide voltage support without requiring an equivalent change in stored energy. This distinction is especially important for a smart-inverter BESS, because an active-power-only strategy does not exploit the full apparent-power capability of the converter.

The third group of studies applies RL and DRL to voltage and Volt/VAR control. Safe off-policy DRL, model-augmented reinforcement learning, decentralized safe control, multi-agent coordination, topology-aware learning, and residual model-assisted methods have all been investigated [14,15,18,27–30]. The recent direction of this field increasingly emphasizes constraint handling, robustness to operating uncertainty, physically meaningful action spaces, sample efficiency, and scalable coordination across multiple controllable resources [21,31–33]. These developments demonstrate the usefulness of learning-based control for nonlinear and sequential distribution-system problems. However, many existing studies are formulated around PV-rich feeders, generic Volt/VAR regulation, or voltage-magnitude objectives. The specific combination of phase-dependent EV demand, sequence-based voltage-unbalance assessment, and matched P-only versus coordinated P–Q BESS operation has received comparatively less attention.

A fourth group focuses on BESS and V2G flexibility in distribution networks. These studies show that flexible active-power exchange can reduce peak demand, feeder loss, operating cost, or voltage deviation, while the achievable benefit depends strongly on battery energy capacity, inverter rating, SOC restrictions, charging and discharging limits, and terminal operating conditions [29,34,35]. This observation has an important implication for comparative controller evaluation. If two control strategies are allowed to use different energy budgets, different terminal SOC values, or different converter capabilities, then any apparent performance improvement may result from additional hardware utilization rather than from the control strategy itself. A meaningful comparison between active-power-only and active–reactive-power control therefore requires the underlying BESS hardware and energy constraints to remain unchanged.

This requirement motivates the design philosophy adopted in the present study. The P-only and P–Q PPO controllers are evaluated using the same BESS energy capacity, the same per-phase active-power limit, the same inverter apparent-power rating, the same SOC bounds, and the same initial and terminal SOC conditions. The two learned cases are also exposed to the same feeder model, load profile, PV generation profile, and phase-asymmetric EV demand. The main structural difference is that the P–Q controller can additionally use inverter reactive power while the P-only controller

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



cannot. This matched design allows the incremental contribution of inverter reactive-power capability to be examined without confounding it with changes in battery size or energy availability.

The research question is therefore deliberately narrower than a general claim of optimal or universally superior DRL control. PPO is used here as the policy-learning mechanism rather than proposed as a new reinforcement-learning algorithm. Likewise, the IEEE 33-bus phase-decoupled extension is used as a controlled evaluation platform and is not presented as a replacement for a fully parameterized unbalanced distribution feeder. The study asks whether coordinated use of active and reactive power from the same smart-inverter BESS produces a measurable improvement over active-power-only operation when the feeder is subjected to repeatable phase-asymmetric EV charging.

To answer this question, the evaluation is based on several complementary engineering metrics rather than training reward alone. Sequence VUF is used to quantify phase unbalance, feeder energy loss is used to assess network efficiency, the minimum phase voltage is reported to distinguish unbalance mitigation from voltage-magnitude restoration, and terminal SOC is enforced to preserve comparability across the 24-h operating horizon. In addition, independently trained PPO policies are evaluated across multiple random seeds, and the learned P–Q policies are tested under structured load, PV, and EV-demand perturbations without retraining. This design links the controller evaluation directly to physically interpretable quantities and provides a more transparent assessment of the benefits and limitations of coordinated smart-inverter BESS operation.

## 3. Materials and Methods

This section describes the feeder model, phase-dependent operating conditions, voltage-unbalance metric, smart-inverter BESS model, PPO-based coordinated control, and the training and evaluation protocol. The methodology is designed to provide a controlled comparison between active-power-only and coordinated active–reactive-power operation. In particular, the P-only and P–Q controllers use the same feeder, operating profiles, BESS energy capacity, active-power rating, SOC constraints, initial SOC, and terminal SOC. Consequently, the comparison isolates the additional network benefit provided by inverter reactive-power capability rather than differences in hardware size or available stored energy.

### *3.1. Distribution Feeder and Operating Conditions*

The study is based on the IEEE 33-bus radial distribution feeder, a widely used benchmark for distribution-system loss and voltage studies [36]. The feeder operates at 12.66 kV on a 10-MVA base. The corrected benchmark data are used throughout the study, including a reactive load of 600 kVAr at Bus 30.

Before PV generation, EV charging, and BESS control are introduced, the balanced benchmark is evaluated as a consistency check. The resulting active-power loss is 202.678 kW and the minimum bus-voltage magnitude is 0.913090 p.u. These values are used to verify the corrected feeder implementation before phase-dependent operating conditions are applied.

The conventional IEEE 33-bus data set does not provide complete $3 \times 3$ phase-domain impedance matrices for all branches. Therefore, unsupported mutual-coupling impedances are not introduced. Instead, the validated branch impedance is applied independently to phases A, B, and C. The resulting model is a *phase-decoupled three-phase extension* of the benchmark feeder. This representation permits phase-dependent loading and phase-specific control while avoiding artificially constructed coupling parameters. It should not, however, be interpreted as a complete electromagnetic model of an untransposed practical feeder with explicit phase coupling and neutral-conductor effects.

A deterministic 24-h load profile is applied with phase scaling factors of 0.85, 1.30, and 0.75 for phases A, B, and C, respectively. Thus, the base operating condition itself is phase dependent, with phase B representing the heaviest-loaded phase. A 500-kW peak PV source is connected at Bus 18 and its active-power output is divided equally among the three phases.

Additional EV charging demand is connected at Bus 30 to represent a repeatable evening charging disturbance. Phase B receives 120 kW during hours 17–22, while phase C receives 80 kW during hours

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



18–21. The combination of phase-dependent base loading and asynchronous EV charging produces a controlled unbalanced operating condition against which the controllers can be evaluated.

All controller cases are subjected to the same 24-h load, PV, and EV profiles. This common chronological operating horizon permits direct comparison of phase voltages, voltage unbalance, feeder losses, BESS dispatch, and SOC trajectories.

### *3.2. Voltage-Unbalance Metric*

Voltage unbalance is evaluated using symmetrical components rather than a simple deviation among phase-voltage magnitudes. For the complex phase voltages $V_a$, $V_b$, and $V_c$, the positive- and negative-sequence components are

$$V_1 = \frac{1}{3}\left(V_a + aV_b + a^2V_c\right), \tag{1}$$

$$V_2 = \frac{1}{3}\left(V_a + a^2V_b + aV_c\right), \tag{2}$$

where

$$a = e^{j2\pi/3}. \tag{3}$$

The voltage-unbalance factor is then defined as

$$\text{VUF} = \frac{|V_2|}{|V_1|} \times 100\%. \tag{4}$$

The sequence-based definition in Eq. (4) is used consistently in the nominal and robustness evaluations. At each hourly operating point, VUF is computed at every bus. Two aggregate indicators are subsequently reported: the worst VUF observed during the 24-h horizon and the mean of the hourly maximum VUF values. The first quantity identifies the most severe unbalance condition, whereas the second reflects the persistence of voltage unbalance over the daily operating cycle.

Minimum phase-voltage magnitude is reported separately from VUF. This distinction is important because mitigation of negative-sequence voltage does not necessarily imply restoration of all phase voltages to an acceptable magnitude range.

### *3.3. BESS and Smart-Inverter Model*

The BESS is connected at Bus 18 with an energy capacity of 800 kWh. The active-power command on each phase is limited to

$$-100\ \text{kW} \le P_{b,\phi,t} \le 100\ \text{kW}, \tag{5}$$

where $\phi \in \{A, B, C\}$ and $t$ denotes the hourly operating interval. The inverter is rated at 100 kVA per phase. For coordinated P–Q operation, the feasible operating region is constrained by

$$P_{b,\phi,t}^2 + Q_{b,\phi,t}^2 \le S_{b,\phi}^2, \tag{6}$$

with $S_{b,\phi} = 100$ kVA. Hence, the available reactive-power range is reduced as the magnitude of active power approaches the inverter apparent- power rating.

The BESS SOC is constrained by

$$0.05 \le \text{SOC}_t \le 0.95, \tag{7}$$

with both the initial and terminal SOC set to 50%. Charging and discharging efficiencies are both 0.95. For a one-hour time step, the SOC transition is

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



$$\mathrm{SOC}_{t+1} = \mathrm{SOC}_t - \frac{\Delta t}{E_{\mathrm{BESS}}}\left[\frac{\sum_\phi P^+_{b,\phi,t}}{\eta_d} - \eta_c \sum_\phi P^-_{b,\phi,t}\right], \tag{8}$$

where $P^+_{b,\phi,t} = \max(P_{b,\phi,t}, 0)$ denotes discharge power, $P^-_{b,\phi,t} = \max(-P_{b,\phi,t}, 0)$ denotes charging power, $E_{\mathrm{BESS}} = 800$ kWh, $\eta_c = \eta_d = 0.95$, and $\Delta t = 1$ h.

A terminal-reachability projection is applied before the SOC update. At each hour, the requested active-power action is restricted so that the remaining charging or discharging capability is sufficient to return the BESS to its prescribed 50% terminal SOC by the end of the 24-h horizon. At the final hour, this projection enforces the terminal-energy condition directly. This treatment prevents one controller from obtaining an apparent network benefit by using more net stored energy than another controller.

### *3.4. PPO-Based Coordinated Control*

Three operating cases are considered: (i) no BESS, (ii) P-only PPO, and (iii) P–Q PPO. The no-BESS case provides the uncontrolled reference. The P-only controller schedules three phase-dependent active-power commands, whereas the P–Q controller additionally schedules three reactive-power commands. The two learned controllers use identical BESS and feeder conditions; therefore, the difference between their performance represents the incremental contribution of inverter reactive-power capability.

The PPO state vector contains 138 quantities,

$$\mathbf{s}_t = [\mathbf{v}_t, \boldsymbol{\nu}_t, \mathrm{SOC}_t, \tau_t, \ell_t, p^{\mathrm{PV}}_t, p^{\mathrm{EV,B}}_t, p^{\mathrm{EV,C}}_t]^{\mathrm{T}}, \tag{9}$$

where $\mathbf{v}_t \in \mathbb{R}^{99}$ contains the three phase-voltage magnitudes at the 33 buses and $\boldsymbol{\nu}_t \in \mathbb{R}^{33}$ contains the corresponding bus-wise VUF values. The remaining elements are the shared BESS SOC, normalized hour index, normalized load level, normalized PV generation, and normalized EV demands on phases B and C.

The P-only action vector is

$$\mathbf{a}^P_t = [a_{P_A}, a_{P_B}, a_{P_C}]^{\mathrm{T}}, \tag{10}$$

whereas the P–Q controller uses

$$\mathbf{a}^{PQ}_t = [a_{P_A}, a_{P_B}, a_{P_C}, a_{Q_A}, a_{Q_B}, a_{Q_C}]^{\mathrm{T}}. \tag{11}$$

The Gaussian policy outputs are transformed by $\tanh(\cdot)$, restricting each normalized action to $[-1, 1]$. Active-power actions are scaled by 100 kW per phase. For P–Q operation, requested reactive-power actions are scaled by 100 kVA and then projected onto the inverter capability limit,

$$|Q_{b,\phi,t}| \le \sqrt{S^2_{b,\phi} - P^2_{b,\phi,t}}. \tag{12}$$

The PPO agent is trained with a scalarized engineering objective rather than a formal Pareto multi-objective formulation. The implemented reward is

$$\begin{aligned} r_t = -\Big[&10 J_{\mathrm{VUF},t} + 10 J_{\mathrm{UV},t} + 5 \times 10^{-4} J_{\mathrm{loss},t} \\ &+ 2 J_{\mathrm{SOC},t} + 1.5 J_{\mathrm{act},t} + 0.10 J_{\mathrm{ramp},t} + 0.1 J_{\mathrm{vdev},t}\Big], \end{aligned} \tag{13}$$

where

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



$$J_{\mathrm{VUF},t} = \max_i \mathrm{VUF}_{i,t}, \tag{14}$$

$$J_{\mathrm{UV},t} = \frac{1}{3N_b} \sum_{i=1}^{N_b} \sum_{\phi \in \{A,B,C\}} \max\left(0, 0.95 - |V_{i,\phi,t}|\right), \tag{15}$$

$$J_{\mathrm{loss},t} = P_{\mathrm{loss},t}, \tag{16}$$

$$J_{\mathrm{SOC},t} = |\mathrm{SOC}_t - \mathrm{SOC}_0|, \tag{17}$$

$$J_{\mathrm{act},t} = \frac{\sum_\phi |P_{b,\phi,t}|}{3P_{\max}}, \tag{18}$$

$$J_{\mathrm{ramp},t} = \frac{\sum_\phi (P_{b,\phi,t} - P_{b,\phi,t-1})^2}{3P_{\max}^2}, \tag{19}$$

$$J_{\mathrm{vdev},t} = \frac{1}{3N_b} \sum_{i=1}^{N_b} \sum_{\phi \in \{A,B,C\}} \left(|V_{i,\phi,t}| - 1\right)^2. \tag{20}$$

Here, $N_b = 33$ and $P_{\max} = 100$ kW. At the end of each 24-h episode, an additional terminal-SOC penalty is applied,

$$r_T \leftarrow r_T - 300|\mathrm{SOC}_T - \mathrm{SOC}_0|. \tag{21}$$

The $J_{\mathrm{act}}$ and $J_{\mathrm{ramp}}$ terms are controller-shaping quantities that discourage excessive active-power use and abrupt power changes. They are treated only as operational-stress proxies and are not interpreted as a calibrated electrochemical battery-degradation or state-of-health model. Physical feasibility is enforced separately through active-power limits, the inverter P–Q capability circle, SOC limits, and the terminal-reachability projection.

A shared actor–critic neural network is implemented directly in PyTorch. The 138-dimensional state vector passes through two fully connected hidden layers of 128 neurons each with hyperbolic-tangent activation. The actor produces the mean of a diagonal Gaussian policy with trainable log-standard deviations, whereas the critic produces a scalar state-value estimate.

Table 2 summarizes the verified PPO implementation parameters.

**Table 2.** Verified PPO implementation parameters.

| Parameter | Implemented value |
|---|---|
| Observation dimension | 138 |
| Action dimension, P-only | 3 |
| Action dimension, P–Q | 6 |
| Hidden layers | 128–128 |
| Activation function | Tanh |
| Policy distribution | Diagonal Gaussian |
| Action transformation | Tanh squashing |
| Optimizer | Adam |
| Learning rate | $3 \times 10^{-4}$ |
| Discount factor, $\gamma$ | 0.99 |
| GAE parameter, $\lambda$ | 0.95 |
| PPO clipping ratio | 0.20 |
| Rollout length | 1024 interactions |
| Optimization epochs | 6 |
| Minibatch size | 128 |
| Value-loss coefficient | 0.5 |
| Entropy coefficient | 0.005 |
| Gradient-norm limit | 0.5 |
| Training interactions | 10,000 per seed/controller |
| Independent seeds | 41, 42, 43 |
| Episode horizon | 24 h |
| Time step | 1 h |

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



*3.5. Training, Evaluation, and Reproducibility Protocol*

Separate P-only and P–Q policies are trained using random seeds 41, 42, and 43. Each controller is trained for 10,000 environment interactions using the nominal 24-h operating profile. Scenario randomization is disabled during the verified training runs. Therefore, variations in load, PV generation, and EV demand considered later in the study are out-of-training-profile tests rather than disturbances encountered during training.

During PPO training, actions are sampled from the squashed Gaussian policy. Generalized advantage estimation is used with $\gamma = 0.99$ and $\lambda = 0.95$. PPO updates use a clipping ratio of 0.20, rollout blocks of up to 1024 interactions, six optimization epochs per rollout, and minibatches of 128 samples. The network parameters are optimized using Adam with a learning rate of $3 \times 10^{-4}$. The value-loss coefficient is 0.5, the entropy coefficient is 0.005, and the gradient norm is limited to 0.5.

During feeder-performance evaluation, stochastic policy sampling is disabled. The deterministic action is obtained from

$$\mathbf{a}_t = \tanh(\boldsymbol{\mu}_\theta(\mathbf{s}_t)), \tag{22}$$

where $\boldsymbol{\mu}_\theta$ is the learned actor mean. Consequently, differences among the three reported learned-policy results arise from independently trained policy parameters rather than from random action sampling during evaluation.

No individual seed is selected as the "best" policy. The reported learned controller results are aggregated across all three independent seeds. For a quantity $x$, the mean and sample standard deviation are calculated as

$$\bar{x} = \frac{1}{N_s} \sum_{i=1}^{N_s} x_i, \tag{23}$$

$$\sigma_x = \sqrt{\frac{1}{N_s - 1} \sum_{i=1}^{N_s} (x_i - \bar{x})^2}, \tag{24}$$

where $N_s = 3$.

Nominal performance is assessed using four principal feeder-level metrics: daily feeder energy loss, worst VUF, mean hourly maximum VUF, and worst minimum phase-voltage magnitude. Terminal SOC is also inspected to verify that performance improvements are not obtained through net battery depletion.

The trained P–Q policies are subsequently evaluated without retraining under structured profile perturbations consisting of load variations of $\pm 10\%$, PV variations of $\pm 20\%$, EV-demand variations of $\pm 20\%$, and combined favorable and stress cases. For each case, the feeder operating point is recomputed using the modified profiles. The reported robustness results are therefore obtained from explicit feeder simulations rather than by scaling nominal performance curves.

This robustness analysis should be interpreted as a bounded sensitivity assessment rather than proof of universal robustness. It does not represent arbitrary feeder topologies, communication delays, measurement errors, or field-observed stochastic disturbances.

The numerical evidence follows a traceable workflow,

$$\text{network model} \rightarrow \text{trained policy} \rightarrow \text{24-h evaluation} \rightarrow \text{reported results}. \tag{25}$$

Trained models, training logs, machine-readable evaluation outputs, and figure-generation data are retained in the verified reproducibility package. Only results obtained from the corrected computational model and actual trained policies are used as numerical evidence in this manuscript.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



## 4. Results

### *4.1. Nominal Voltage-Unbalance Performance*

Figure 1 presents the hourly maximum VUF for the no-BESS, P-only PPO, and P–Q PPO cases over the nominal 24-h operating horizon. The no-BESS case consistently exhibits the largest voltage unbalance, with the most pronounced deterioration occurring during the evening period when the phase-dependent EV charging demand is active. This response is consistent with the operating conditions defined in Section 3, where the feeder already contains phase-dependent base loading and receives additional EV demand on phases B and C.

Both learned BESS controllers reduce the VUF relative to the no-BESS case. However, the coordinated P–Q controller produces the lowest VUF over most of the affected operating period. This behavior indicates that active-power redistribution alone provides useful mitigation, while the additional reactive-power degree of freedom further improves the ability of the smart inverter to respond to phase-dependent voltage conditions.

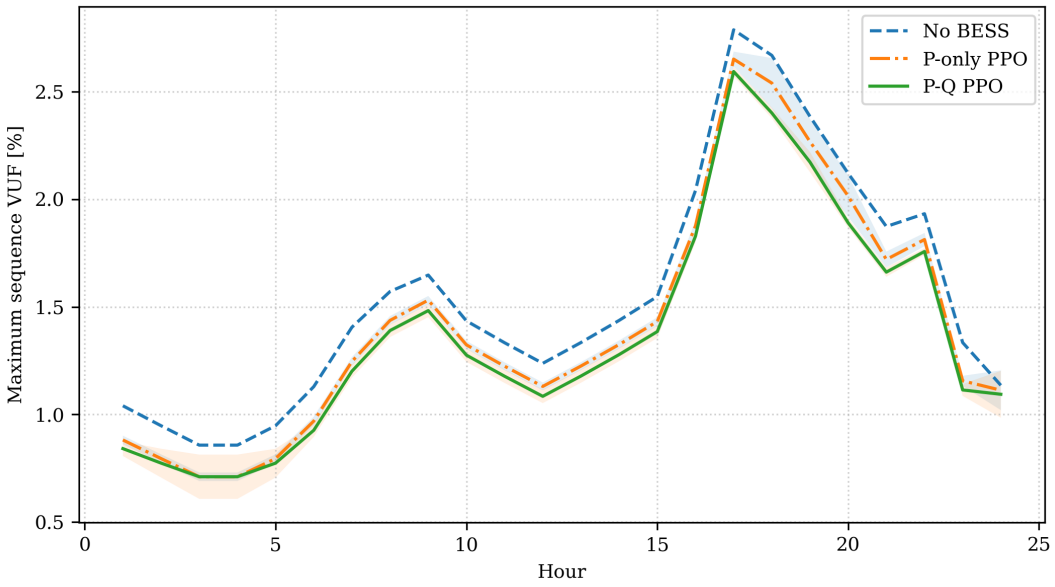



**Figure 1.** Verified hourly voltage-unbalance performance for the no-BESS, P-only PPO, and P–Q PPO cases. Learned-controller curves report the mean and variability across the three independent seeds.

Table 3 summarizes the nominal feeder-level performance. The no-BESS case results in a daily feeder energy loss of 3874.822 kWh/day, a worst VUF of 2.7879%, a mean hourly maximum VUF of 1.5422%, and a worst minimum voltage of 0.8342 p.u.

P-only PPO reduces the daily energy loss to $3851.179 \pm 25.931$ kWh/day, the worst VUF to $2.6663 \pm 0.0264$%, and the mean hourly maximum VUF to $1.4122 \pm 0.0081$%. Relative to the no-BESS condition, these values correspond to reductions of 0.610%, 4.362%, and 8.427%, respectively.

When coordinated reactive-power capability is enabled, P–Q PPO further reduces daily energy loss to $3809.751 \pm 59.152$ kWh/day, worst VUF to $2.5940 \pm 0.0208$%, and mean hourly maximum VUF to $1.3628 \pm 0.0263$%. The corresponding reductions relative to the no-BESS case are 1.679%, 6.954%, and 11.632%, respectively.

**Table 3.** Verified nominal performance.

| Case | Daily loss (kWh/day) | Worst VUF (%) | Mean hourly max. VUF (%) | Worst min. voltage (p.u.) |
|---|---|---|---|---|
| No BESS | 3874.822 | 2.7879 | 1.5422 | 0.8342 |
| P-only PPO | $3851.179 \pm 25.931$ | $2.6663 \pm 0.0264$ | $1.4122 \pm 0.0081$ | $0.8354 \pm 0.0016$ |
| P–Q PPO | $3809.751 \pm 59.152$ | $2.5940 \pm 0.0208$ | $1.3628 \pm 0.0263$ | $0.8367 \pm 0.0013$ |

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



The relative improvement is larger for the VUF-based indicators than for daily feeder energy loss. This result is consistent with the multi-criteria reward described in Section 3.4. The controller is not trained as a pure loss-minimization algorithm; it simultaneously responds to voltage unbalance, undervoltage, loss, SOC deviation, battery activity, active-power ramping, and mean voltage deviation. Therefore, the resulting operating point should be interpreted as a compromise among several feeder and BESS objectives rather than as the minimum-loss solution.

The comparison between P-only and P–Q control is particularly important. Because the two controllers use the same BESS energy capacity, active-power limits, SOC bounds, initial SOC, terminal SOC, feeder model, and operating profiles, the additional improvement observed in the P–Q case cannot be attributed to a larger battery or a larger active-power budget. Within the tested system, it instead reflects the additional flexibility provided by reactive-power control.

### *4.2. Minimum Voltage and Feeder Energy Loss*

Figures 2 and 3 show the corresponding minimum-voltage and feeder-energy-loss results. Although the P–Q controller provides the best nominal minimum-voltage result among the three cases, the absolute improvement in minimum voltage is modest. The worst minimum voltage increases from 0.8342 p.u. in the no-BESS case to $0.8354 \pm 0.0016$ p.u. for P-only PPO and $0.8367 \pm 0.0013$ p.u. for P–Q PPO.

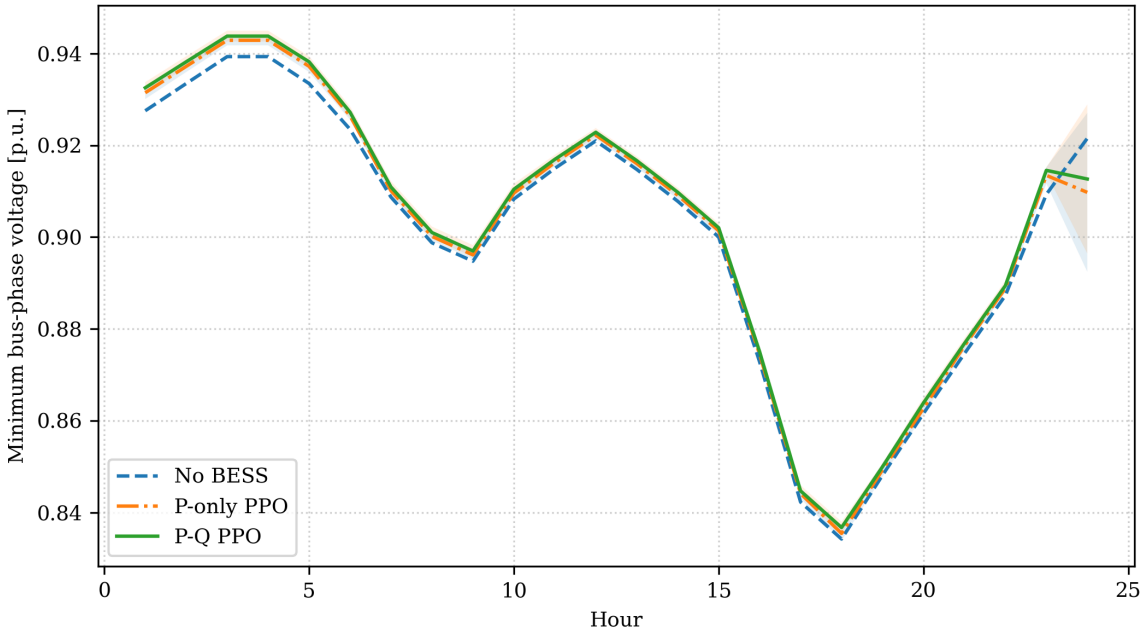



**Figure 2.** Minimum-voltage performance across the nominal 24-h operating horizon.

The limited increase in minimum voltage should be distinguished from the clearer reduction in VUF. The two quantities measure different electrical phenomena. VUF is governed by the negative-sequence component relative to the positive sequence, whereas the minimum-voltage metric identifies the lowest individual phase-voltage magnitude. Consequently, improved phase balance does not necessarily imply that all low-voltage conditions have been removed.

The feeder-energy-loss results in Figure 3 exhibit a similar ordering. P–Q PPO provides the lowest mean daily loss, followed by P-only PPO and the no-BESS condition. Nevertheless, the reduction in energy loss is smaller in relative terms than the improvement in the sequence-based unbalance metrics. This again confirms that the learned controller is balancing multiple objectives rather than optimizing feeder loss alone.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



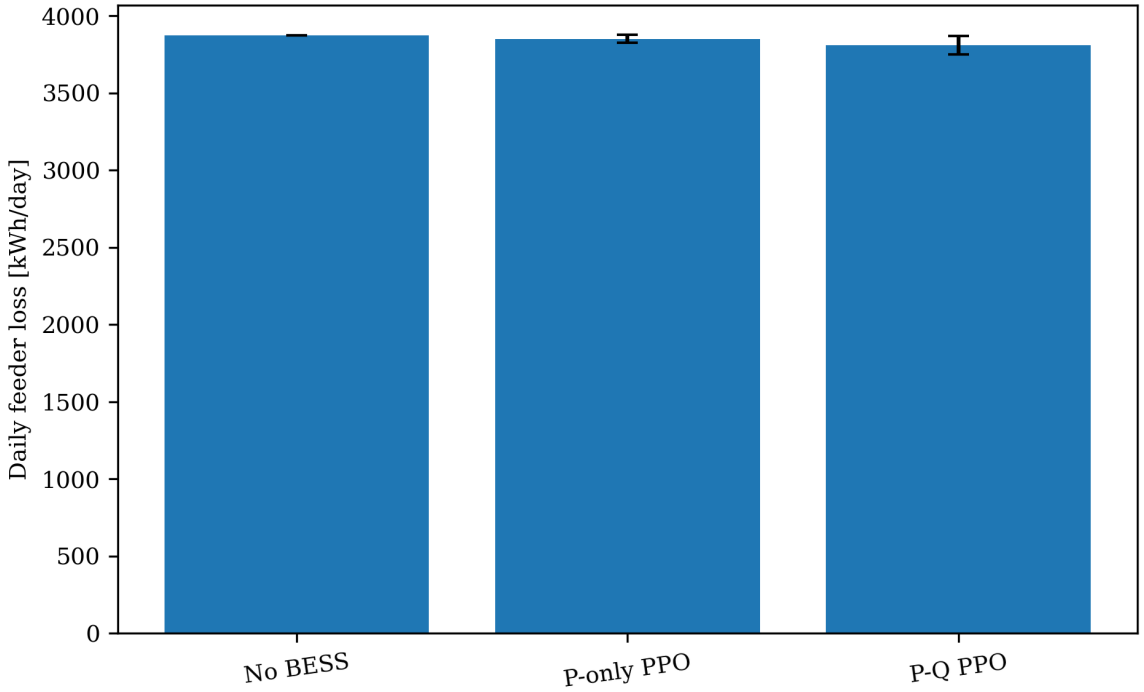



**Figure 3.** Daily feeder energy loss for the verified control cases.

These results support a voltage-unbalance mitigation claim but not a feeder-wide voltage-restoration or voltage-compliance claim. Even in the nominal case, the minimum voltage remains substantially below 1.0 p.u., indicating that the available BESS and inverter capability primarily improves the operating condition rather than fully correcting the feeder voltage profile.

### *4.3. Active–Reactive Dispatch and SOC Feasibility*

Figures 4 and 5 show the active- and reactive-power trajectories of a representative P–Q policy corresponding to seed 41. These figures illustrate how the two control channels are used within the same smart-inverter BESS.

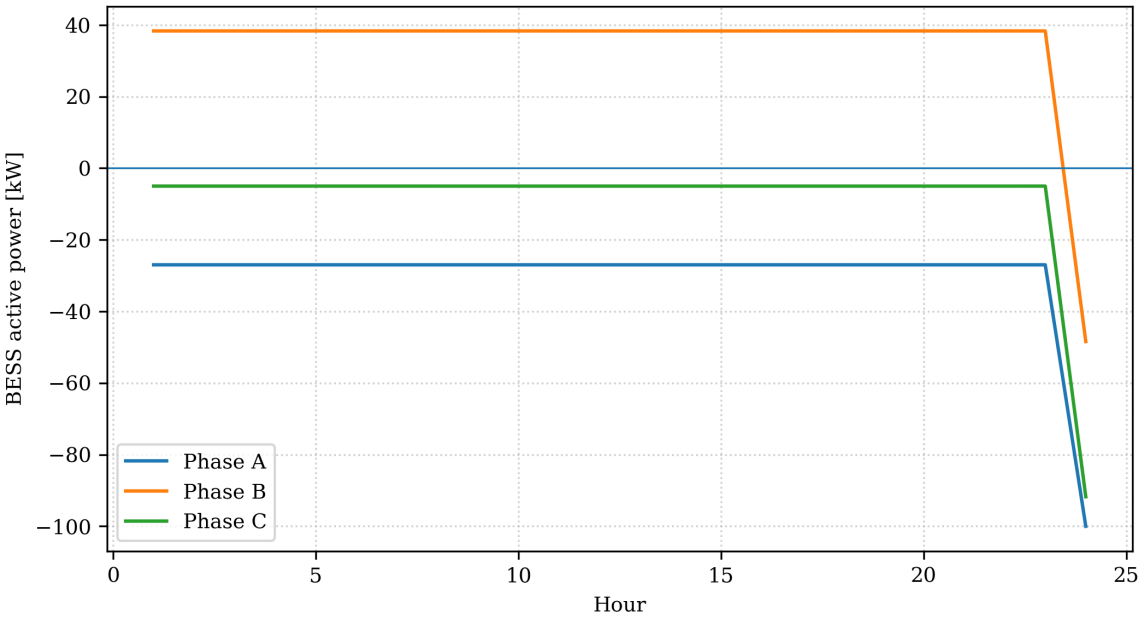



**Figure 4.** Representative active-power dispatch of the P–Q PPO controller for seed 41.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



The active-power dispatch alters the net real-power loading seen by the feeder and simultaneously changes BESS energy content. Consequently, each active-power decision affects not only the present operating point but also the remaining SOC flexibility available later in the 24-h horizon.

Reactive-power dispatch, shown in Figure 5, provides an additional control mechanism that does not require an equivalent transfer of stored battery energy. However, reactive-power availability is not independent of active power. As defined by the inverter capability constraint in Equation (6), the allowable magnitude of $Q$ decreases when the corresponding active-power magnitude approaches the 100-kVA per-phase inverter rating.

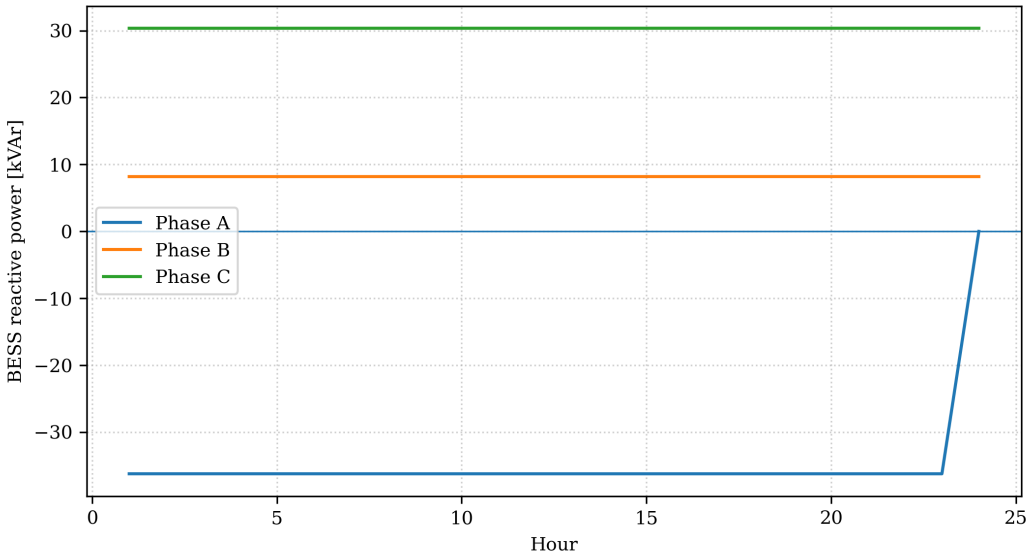



**Figure 5.** Representative reactive-power dispatch of the P–Q PPO controller for seed 41.

The SOC trajectory in Figure 6 provides an important feasibility check. The controller is allowed to charge and discharge during the day, but the terminal-reachability projection ensures that the BESS retains sufficient remaining capability to return to its prescribed terminal state. All reported learned-policy evaluations return the SOC to 50% at hour 24.

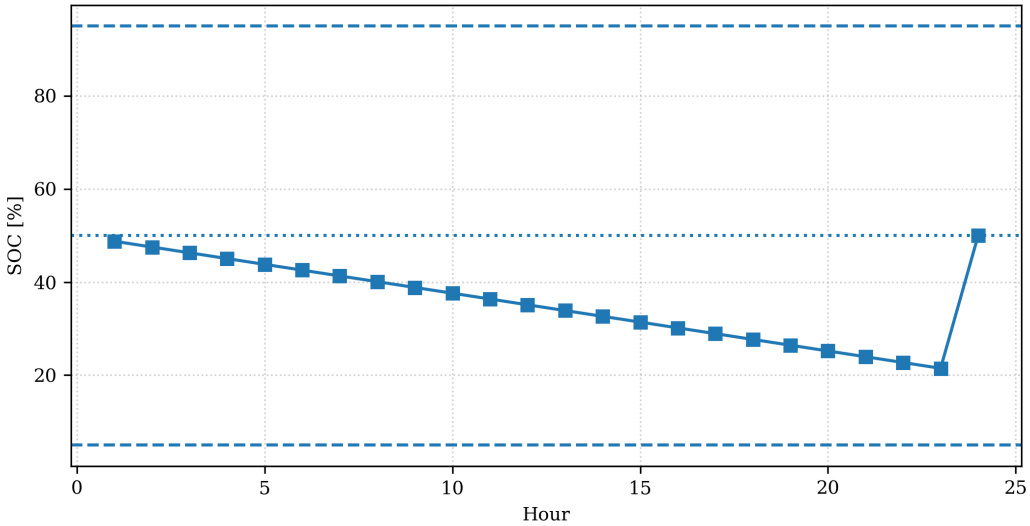



**Figure 6.** Representative SOC trajectory for seed 41. The terminal-reachability projection enforces return to the prescribed 50% terminal SOC.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



This result is essential for interpreting the controller comparison. Without terminal-energy matching, one controller could appear to reduce loss or improve voltage simply by ending the day with less stored energy. Because both learned control cases begin and end at the same SOC, the reported improvements are not produced by net depletion of the BESS.

### *4.4. Phase-Voltage and Spatial VUF Behavior*

Figure 7 shows the three-phase voltage response at Bus 33 for the representative seed 41 evaluation. The individual phase-voltage trajectories illustrate the effect of the phase-dependent loading defined in the feeder model. Differences among the three voltage magnitudes become more visible during heavily loaded intervals and during the asymmetric EV-charging period.

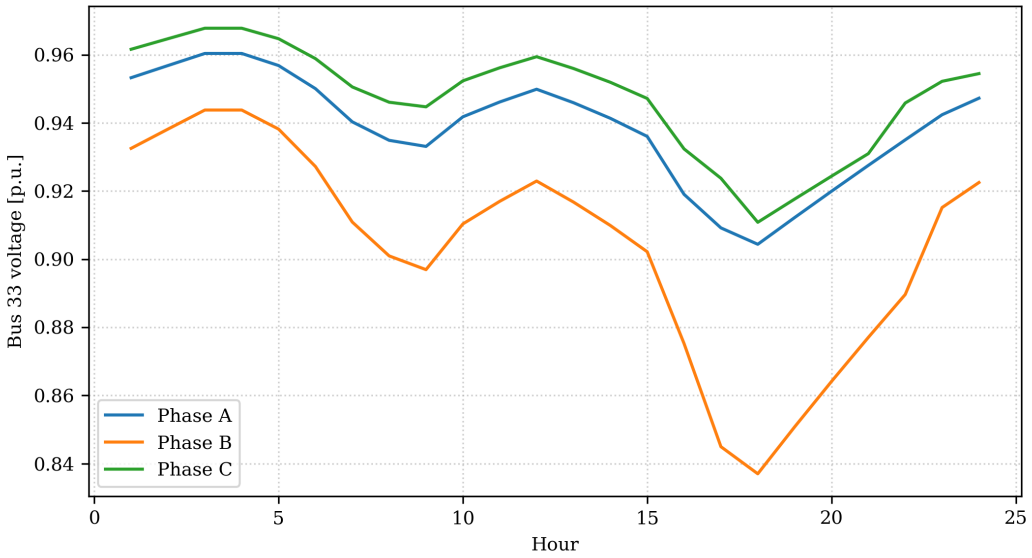



**Figure 7.** Three-phase Bus 33 voltage profile for the representative verified evaluation using seed 41.

Figure 8 extends this interpretation from a single bus to the complete feeder. The bus–time VUF heat map shows that voltage unbalance is spatially and temporally nonuniform. This observation is important because a single minimum-voltage value cannot identify where negative-sequence effects are concentrated or how long they persist.

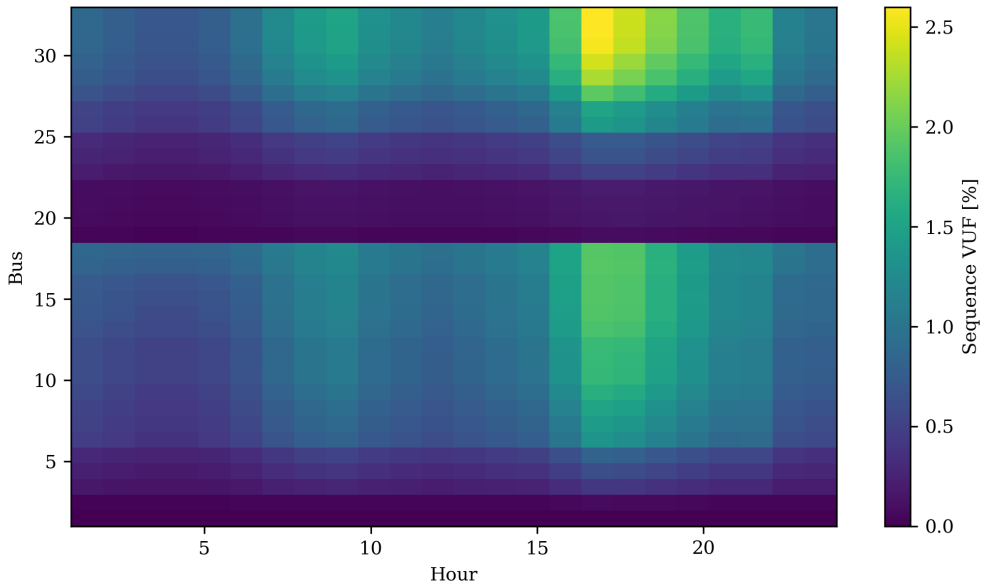



**Figure 8.** Bus–time VUF heat map for the representative P–Q PPO evaluation using seed 41.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



Together, Figures 7 and 8 demonstrate why both voltage magnitude and sequence VUF are retained as evaluation metrics. A controller may reduce the negative-sequence component and improve phase balance while only modestly changing the lowest phase-voltage magnitude. The two indicators should therefore be interpreted as complementary rather than interchangeable measures of feeder voltage performance.

### *4.5. Robustness Without Retraining*

The trained P–Q policies are next evaluated without retraining under the structured operating-profile perturbations defined in Section 3.5. These cases include load variations of ±10%, PV variations of ±20%, EV-demand variations of ±20%, and combined favorable and stress conditions. Because scenario randomization is disabled during the verified training runs, these perturbations represent out-of-training-profile evaluations.

Figure 9 shows the resulting worst-VUF performance. Across the tested perturbations, the mean worst-VUF reduction produced by P–Q PPO remains approximately 5.66–7.86% relative to the corresponding no-BESS case. The continued reduction across multiple modified profiles indicates that the learned policy is not simply reproducing a fixed nominal dispatch trajectory.

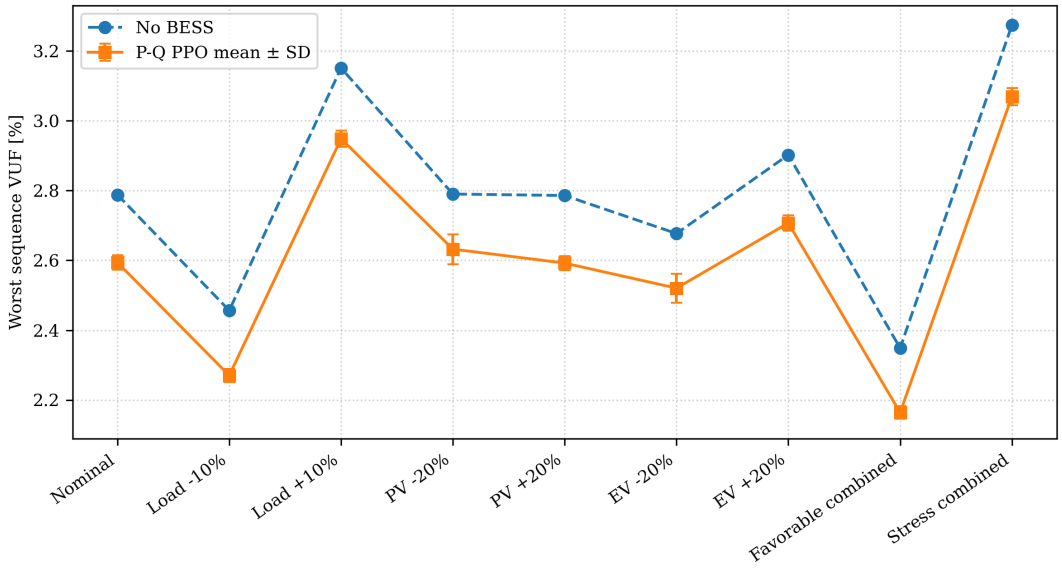



**Figure 9.** Robustness of worst VUF under load, PV, EV-demand, and combined operating-profile perturbations. The trained P–Q policies are evaluated without retraining.

Nevertheless, the robustness result should be interpreted within the tested range. The perturbations alter load, PV generation, and EV demand while retaining the same feeder topology, BESS placement, network parameters, and controller structure. The analysis therefore provides bounded evidence of generalization to operating-profile variation rather than proof of universal robustness.

Figure 10 provides the corresponding minimum-voltage results. The combined stress condition is particularly informative. Under this case, the worst minimum voltage is approximately 0.8121 p.u. without BESS and 0.8142 p.u. for the P–Q PPO mean.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



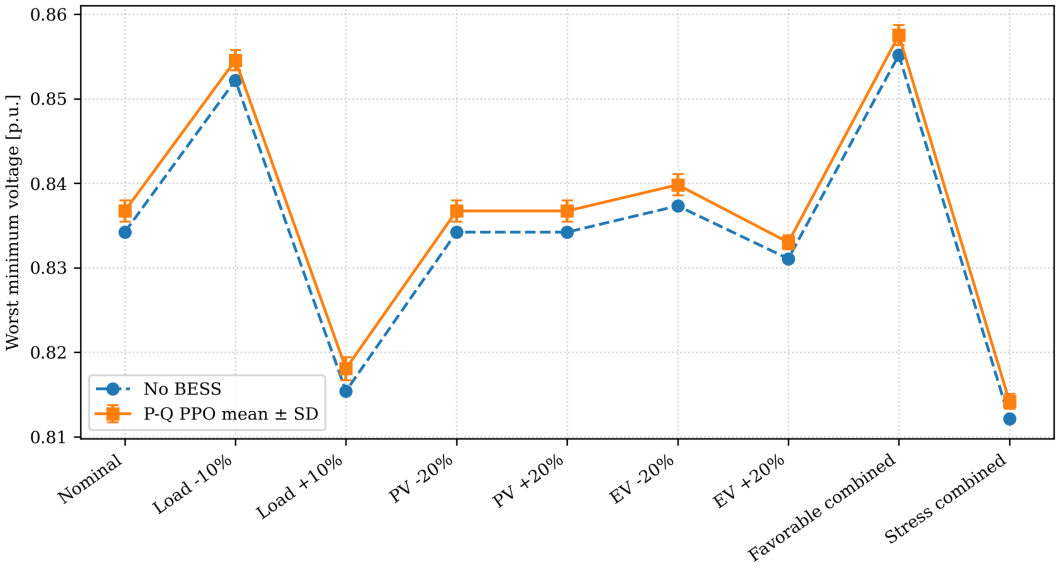



**Figure 10.** Minimum-voltage robustness under the same operating-profile perturbations.

Although the controller continues to provide VUF improvement under the combined stress case, the minimum-voltage recovery remains small. This result exposes a physical limitation of the available BESS and inverter capability. A learned controller can redistribute the available active and reactive power, but it cannot provide voltage-support capacity beyond the rating and location of the installed resource.

### *4.6. Training Repeatability*

Figure 11 presents the actual PPO episode returns obtained from the P-only and P–Q training runs for seeds 41, 42, and 43. These curves are used only as learning diagnostics. They are not treated as direct evidence of electrical superiority because a higher accumulated reward does not by itself specify the resulting VUF, feeder loss, minimum voltage, or terminal SOC.

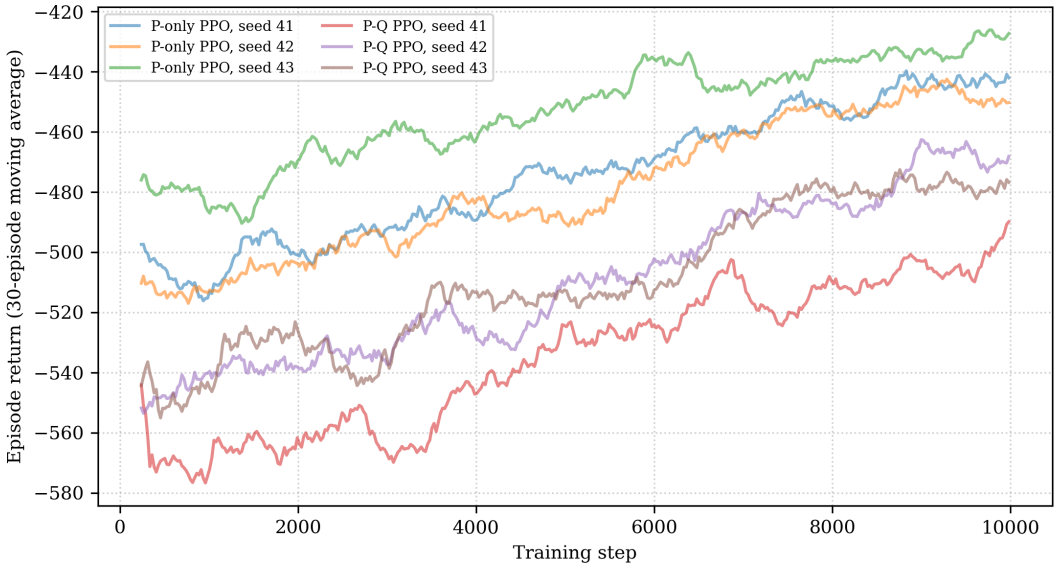



**Figure 11.** Actual PPO training returns for P-only and P–Q controllers across seeds 41, 42, and 43. No synthetic DDPG or rescaled convergence trace is used.

The final comparison is therefore based on the feeder-level metrics obtained after deterministic re-evaluation of the trained policies. The relatively small dispersion in worst VUF suggests that the

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



principal voltage-unbalance mitigation is reproduced across the three independent runs. In contrast, daily feeder loss exhibits greater variation, particularly for P–Q PPO.

This difference is plausible in a scalarized control problem. More than one active–reactive dispatch trajectory can produce similar VUF performance while yielding different feeder-loss values. The observed variability therefore reinforces the importance of reporting multiple independently trained policies rather than selecting only the numerically best seed.

## 5. Discussion

### *5.1. Engineering Interpretation of Coordinated P–Q Control*

The results demonstrate a consistent performance ordering across the verified nominal cases: no BESS, P-only PPO, and P–Q PPO. The principal engineering finding is not that PPO itself is superior to all alternative control methods, but that enabling reactive-power control on the same smart-inverter BESS provides an incremental benefit beyond active-power-only operation under the investigated phase-asymmetric EV-loading condition.

This interpretation follows directly from the matched experimental design. The P-only and P–Q controllers use the same 800-kWh storage capacity, $\pm$100-kW per-phase active-power limit, 100-kVA per-phase inverter rating, SOC limits, initial SOC, terminal SOC, feeder model, and operating profiles. Therefore, the stronger performance of P–Q control cannot be attributed to additional stored energy or a larger active-power actuator.

Relative to the no-BESS case, P-only PPO reduces daily feeder loss by 0.610%, worst VUF by 4.362%, and mean hourly maximum VUF by 8.427%. With reactive-power control included, the corresponding reductions increase to 1.679%, 6.954%, and 11.632%, respectively.

The larger relative improvement in the VUF metrics is physically meaningful. Active-power control affects feeder loading but is coupled directly to the battery energy state. Reactive-power control, by contrast, provides an additional voltage-support variable without requiring an equivalent change in stored energy. The available reactive power remains limited by the per-phase inverter capability circle, so the result should not be interpreted as unlimited independent VAr support. Instead, coordinated P–Q control allows the controller to allocate the same converter capacity between real- and reactive-power functions according to the instantaneous feeder condition.

The minimum-voltage results also clarify the nature of the improvement. The worst minimum voltage changes from 0.8342 p.u. without BESS to $0.8354 \pm 0.0016$ p.u. under P-only PPO and $0.8367 \pm 0.0013$ p.u. under P–Q PPO. These changes are smaller than the relative improvement in VUF. Thus, the principal benefit demonstrated by the present controller is improved phase balance and modest feeder-loss reduction, not complete restoration of low bus voltages.

This distinction is important when comparing the present results with previous EV coordination and inverter-control studies. EV scheduling studies commonly emphasize feeder loss, charging cost, or voltage magnitude [1,17], whereas smart-inverter studies have demonstrated the usefulness of reactive power for voltage regulation [2,3,9]. The present study links these two directions by explicitly evaluating sequence-based VUF under phase-dependent EV demand while maintaining the same BESS energy constraints for the P-only and P–Q cases.

### *5.2. Robustness and Practical Implications*

The no-retraining evaluations provide additional evidence that the learned P–Q behavior is not limited to the nominal operating profile. Across the tested load, PV, EV-demand, and combined perturbations, the mean worst-VUF reduction remains approximately 5.66–7.86% relative to the corresponding no-BESS condition.

This result is relevant because the verified PPO policies are trained on the nominal profile without scenario randomization. Therefore, the robustness cases represent genuinely different operating conditions from those used during training. The retained VUF reduction indicates that the policy has

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



learned a useful mapping between feeder observations and BESS actions rather than only reproducing one fixed hourly action sequence.

However, these results should not be extended beyond the tested disturbance range. The perturbations modify operating profiles but do not change feeder topology, electrical parameters, BESS location, communication assumptions, or measurement quality. The robustness study should therefore be interpreted as a structured sensitivity assessment rather than a demonstration of general robustness to arbitrary field conditions.

The combined stress case provides the clearest example of this limitation. The worst minimum voltage is approximately 0.8121 p.u. without BESS and 0.8142 p.u. for the P–Q PPO mean. Although phase-unbalance mitigation is retained, the absolute voltage remains severely depressed. This indicates that control intelligence cannot substitute for insufficient physical voltage-support capability.

In a practical feeder, more severe undervoltage conditions may require coordination with additional resources such as on-load tap changers, voltage regulators, capacitor banks, distributed PV inverters, or additional strategically placed storage. Therefore, the proposed controller is better viewed as one controllable component within a broader distribution-voltage management framework rather than as a stand-alone solution to all voltage problems.

The terminal-SOC result further strengthens the practical interpretation. All learned-policy evaluations return the BESS to 50% SOC at the end of the 24-h horizon. As a result, the reported network improvements are achieved without net daily depletion of the stored-energy resource. This constraint is particularly important for repeatable day-to-day operation and for a fair comparison between P-only and P–Q control.

### *5.3. Limitations and Required Extensions*

The present results should be interpreted within several clearly defined boundaries.

First, the network representation is a phase-decoupled extension of the IEEE 33-bus feeder. The original benchmark does not provide complete three-phase mutual-impedance matrices, and the present model deliberately avoids introducing unsupported coupling parameters. This choice provides a controlled environment for evaluating phase-dependent loading and sequence-based VUF, but it does not reproduce mutual phase coupling, neutral-conductor effects, line transposition asymmetry, or all properties of a practical unbalanced feeder.

Second, the study establishes the incremental benefit of P–Q control relative to no-BESS and matched P-only PPO operation. It does not compare PPO with a deterministic optimal-dispatch benchmark, model-predictive control, or alternative DRL algorithms. Consequently, the results do not establish global optimality or general algorithmic superiority.

Third, three independent training seeds provide useful repeatability evidence, but they do not constitute a large-sample statistical study. The reported standard deviations should therefore be interpreted as variability among the three verified runs rather than as population-level uncertainty estimates.

Fourth, the robustness analysis uses structured load, PV, and EV-demand perturbations. It does not include measurement noise, communication delay, network reconfiguration, component outages, forecast error distributions, or field-measured stochastic EV arrivals. The present robustness results should therefore not be interpreted as field validation.

Finally, the battery-activity and ramping terms used in the PPO reward are operational-stress proxies only. They do not constitute a physics-based electrochemical degradation or state-of-health model. Future work should therefore combine the present coordinated P–Q formulation with a validated battery-aging representation when lifetime-oriented operation is of interest.

These limitations motivate four natural extensions: evaluation on a fully parameterized unbalanced feeder with explicit phase coupling, comparison against deterministic or model-based optimal dispatch, broader multi-seed and stochastic testing, and validation using measured feeder and EV-demand data.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



## 6. Conclusions

This study developed and evaluated a reproducible PPO-based framework for coordinated active–reactive-power control of a smart-inverter BESS under phase-asymmetric EV charging in a distribution feeder. The comparison was designed so that the P-only and P–Q controllers used the same BESS energy capacity, active-power limits, inverter rating, SOC bounds, initial and terminal SOC conditions, feeder model, and operating profiles. This matched design allowed the additional value of inverter reactive-power capability to be assessed without confounding the comparison with differences in storage hardware or net battery-energy use.

Under the nominal operating profile, P–Q PPO reduced daily feeder loss by 1.679%, worst VUF by 6.954%, and mean hourly maximum VUF by 11.632% relative to the no-BESS case. The corresponding reductions achieved by P-only PPO were 0.610%, 4.362%, and 8.427%, respectively. These results show that active-power scheduling alone provides measurable network benefits, while coordinated reactive-power control provides an additional improvement, particularly in the sequence-based voltage-unbalance metrics.

All reported learned-policy evaluations returned the BESS SOC to 50% at the end of the 24-h horizon. Therefore, the observed performance improvements were not obtained through net depletion of stored battery energy. The robustness assessment further showed that, without retraining, the P–Q policies retained a mean worst-VUF reduction of approximately 5.66–7.86% across the tested load, PV, EV-demand, and combined-profile perturbations. This provides bounded evidence that the learned control behavior extends beyond the nominal profile within the investigated operating range.

The results should nevertheless be interpreted within the scope of the present model and evaluation. The controller improves voltage unbalance and feeder loss but does not achieve feeder-wide voltage restoration, particularly under the combined stress condition. In addition, the phase-decoupled IEEE 33-bus representation does not include explicit mutual phase coupling or all characteristics of a practical unbalanced feeder. The present study also does not establish general superiority of PPO over deterministic optimization, model-based control, or alternative reinforcement-learning algorithms.

Future work should therefore focus on four directions: evaluation using fully parameterized unbalanced feeders with explicit phase coupling; comparison with deterministic and alternative learning-based control baselines; systematic assessment of reward-weight sensitivity and broader stochastic training and testing; and validation using field-derived EV-demand profiles together with physics-based battery-aging models. These extensions would allow the coordinated P–Q control framework to be assessed under progressively more realistic distribution-system conditions.

**Author Contributions:** Conceptualization, T.K., S.S., T.C. and K.B.; methodology, T.K., P.Pb., S.S., N.A. and T.C.; software, T.K., P.P. and P.K.; validation, P.P., P.K., P.Pb. and N.A.; formal analysis, T.K., P.Pb. and S.S.; investigation, T.K., P.P. and P.K.; data curation, T.K. and P.P.; writing—original draft preparation, T.K., P.P. and S.S.; writing—review and editing, T.K., P.P., P.K., P.Pb., S.S., N.A., T.C. and K.B.; visualization, T.K. and P.K.; supervision, S.S., T.C. and K.B.; project administration, T.C. and K.B. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable. This study is a computational power-system simulation and does not involve human participants or animals.

**Informed Consent Statement:** Not applicable.

**Acknowledgments:** The authors acknowledge the support and research facilities provided by Pathum Thani Technical College, Nakhon Phanom University, Udon Thani Technical College, Mahasarakham University, and the Institute of Vocational Education Eastern Region, Thailand. The authors also acknowledge the academic and technical support provided by their respective institutions.

**Conflicts of Interest:** The authors declare no conflicts of interest.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



## References


1. N. Chitgreeyan, P. Pilalum, S. Marsong, S. Sonasang, P. Prabpal, D. N. Vo, K. Buayai, K. Kerdchuen, and Y. Kongjeen, "Multi-period optimization of energy demand control for electric vehicles in unbalanced electrical power systems considering the center load distance of charging station areas," *Engineering Access*, vol. 10, no. 2, pp. 90–102, 2024, doi: 10.14456/mijet.2024.12.
2. M. N. Kabir, Y. Mishra, G. Ledwich, Z. Y. Dong, and K. P. Wong, "Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 967–977, May 2014, doi: 10.1109/TII.2014.2299336.
3. P. Prabpal, Y. Kongjeen, and K. Bhumkittipich, "Optimal battery energy storage system based on VAR control strategies using particle swarm optimization for power distribution system," *Symmetry*, vol. 13, no. 9, Art. no. 1692, 2021, doi: 10.3390/sym13091692.
4. P. Ronyut, P. Prabpal, S. Sonasang, and M. Noosawad, "Study of battery energy management for EV under condition hybrid BESS and V2G in microgrid," *Journal of Industrial Education*, vol. 17, no. 2, pp. 97–108, 2023.
5. B. Zhang, A. Y. S. Lam, A. D. Domínguez-García, and D. Tse, "An optimal and distributed method for voltage regulation in power distribution systems," *IEEE Trans. Power Syst.*, vol. 30, no. 4, pp. 1714–1726, Jul. 2015, doi: 10.1109/TPWRS.2014.2347281.
6. S. Bolognani, R. Carli, G. Cavraro, and S. Zampieri, "Distributed reactive power feedback control for voltage regulation and loss minimization," *IEEE Trans. Autom. Control*, vol. 60, no. 4, pp. 966–981, Apr. 2015, doi: 10.1109/TAC.2014.2363931.
7. H. Zhu and H. J. Liu, "Fast local voltage control under limited reactive power: Optimality and stability analysis," *IEEE Trans. Power Syst.*, vol. 31, no. 5, pp. 3794–3803, Sep. 2016, doi: 10.1109/TPWRS.2015.2504419.
8. Y. Xu, Z. Y. Dong, R. Zhang, and D. J. Hill, "Multi-timescale coordinated voltage/VAR control of high renewable-penetrated distribution systems," *IEEE Trans. Power Syst.*, vol. 32, no. 6, pp. 4398–4408, Nov. 2017, doi: 10.1109/TPWRS.2017.2669343.
9. C. Zhang, Y. Xu, Z. Y. Dong, and J. Ravishankar, "Three-stage robust inverter-based voltage/VAR control for distribution networks with high-level PV," *IEEE Trans. Smart Grid*, vol. 10, no. 1, pp. 782–793, Jan. 2019, doi: 10.1109/TSG.2017.2752234.
10. R. R. Jha, A. Dubey, C.-C. Liu, and K. P. Schneider, "Bi-level Volt-VAR optimization to coordinate smart inverters with voltage control devices," *IEEE Trans. Power Syst.*, vol. 34, no. 3, pp. 1801–1813, May 2019, doi: 10.1109/TPWRS.2018.2890613.
11. J. Li, Z. Xu, J. Zhao, and C. Zhang, "Distributed online voltage control in active distribution networks considering PV curtailment," *IEEE Trans. Ind. Informat.*, vol. 15, no. 10, pp. 5519–5530, Oct. 2019, doi: 10.1109/TII.2019.2903888.
12. J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
13. Y. Zhang, X. Wang, J. Wang, and Y. Zhang, "Deep reinforcement learning based Volt-VAR optimization in smart distribution systems," *IEEE Trans. Smart Grid*, vol. 12, no. 1, pp. 361–371, Jan. 2021, doi: 10.1109/TSG.2020.3010130.
14. W. Wang, N. Yu, Y. Gao, and J. Shi, "Safe off-policy deep reinforcement learning algorithm for Volt-VAR control in power distribution systems," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3008–3018, Jul. 2020, doi: 10.1109/TSG.2019.2962625.
15. Y. Gao and N. Yu, "Model-augmented safe reinforcement learning for Volt-VAR control in power distribution networks," *Applied Energy*, vol. 313, Art. no. 118762, 2022, doi: 10.1016/j.apenergy.2022.118762.
16. D. Cao, J. Zhao, W. Hu, F. Ding, N. Yu, Q. Huang, and Z. Chen, "Model-free voltage control of active distribution system with PVs using surrogate model-based deep reinforcement learning," *Applied Energy*, vol. 306, Art. no. 117982, Jan. 2022, doi: 10.1016/j.apenergy.2021.117982.
17. D. Liu, P. Zeng, S. Cui, and C. Song, "Deep reinforcement learning for charging scheduling of electric vehicles considering distribution network voltage stability," *Sensors*, vol. 23, no. 3, Art. no. 1618, 2023, doi: 10.3390/s23031618.
18. J. Huang, H. Zhang, D. Tian, Z. Zhang, C. Yu, and G. P. Hancke, "Multi-agent deep reinforcement learning with enhanced collaboration for distribution network voltage control," *Engineering Applications of Artificial Intelligence*, vol. 134, Art. no. 108677, 2024, doi: 10.1016/j.engappai.2024.108677.


© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1




19. L. Xue, J. Wang, Y. Qin, Y. Zhang, Q. Yang, and Z. Li, "Two-timescale online coordinated schedule of active distribution network considering dynamic network reconfiguration via bi-level safe deep reinforcement learning," *Electric Power Systems Research*, vol. 234, Art. no. 110549, 2024, doi: 10.1016/j.epsr.2024.110549.
20. M. Wu, L. Hong, C. Xu, Y. Gao, and Y. Wang, "Coordinated Volt/Var optimization of PV and EVCS interfaced power distribution networks using safe reinforcement learning approach," *Sustainable Energy, Grids and Networks*, vol. 43, Art. no. 101896, 2025, doi: 10.1016/j.segan.2025.101896.
21. Z. Wu, Y. Li, X. Zhang, S. Zheng, and J. Zhao, "Distributed voltage control for multi-feeder distribution networks considering transmission network voltage fluctuation based on robust deep reinforcement learning," *Applied Energy*, vol. 379, Art. no. 124984, Feb. 2025, doi: 10.1016/j.apenergy.2024.124984.
22. S. Orfanoudakis, F. A. Oliehoek, P. Palensky, and P. P. Vergara, "Physics-informed reinforcement learning for large-scale EV smart charging considering distribution network voltage constraints," *Applied Energy*, vol. 421, Art. no. 128224, 2026, doi: 10.1016/j.apenergy.2026.128224.
23. F. Ahmed, S. Ahmad, M. T. Rahman, M. R. Hazari, R. Faiz, T. Ahmed, and M. Karimi, "A holistic review of electric vehicle charging impacts on power distribution networks: Technical challenges, smart mitigation strategies and future directions," *Applied Energy*, vol. 402, Part B, Art. no. 126961, Jan. 2026, doi: 10.1016/j.apenergy.2025.126961.
24. J. Hu, G. Yang, and H. W. Bindner, "Network constrained transactive control for electric vehicles integration," in *Proc. IEEE Power & Energy Society General Meeting (PESGM)*, Denver, CO, USA, Jul. 2015, pp. 1–5, doi: 10.1109/PESGM.2015.7286174.
25. X. Shen, J. Shen, Y. Wang, and H. Wu, "A deep reinforcement learning-based reactive power optimization method for distribution networks considering electric vehicle charging load prediction," *Sustainable Energy, Grids and Networks*, vol. 43, Art. no. 101777, Sep. 2025, doi: 10.1016/j.segan.2025.101777.
26. J.-W. Park and H.-S. Yang, "A reinforcement learning-based scheduling for electric vehicle charging stations with distributed energy resources," *Sustainable Energy, Grids and Networks*, vol. 46, Art. no. 102244, Jun. 2026, doi: 10.1016/j.segan.2026.102244.
27. W. Cui, J. Li, and B. Zhang, "Decentralized safe reinforcement learning for inverter-based voltage control," *Electric Power Systems Research*, vol. 211, Art. no. 108609, 2022, doi: 10.1016/j.epsr.2022.108609.
28. H. Li, Z. Wang, and H. He, "Distributed Volt-VAR optimization based on multi-agent deep reinforcement learning," in *Proc. Int. Joint Conf. Neural Networks (IJCNN)*, Shenzhen, China, Jul. 2021, doi: 10.1109/IJCNN52387.2021.9534348.
29. Y. Xiang, Y. Lu, and J. Liu, "Deep reinforcement learning based topology-aware voltage regulation of distribution networks with distributed energy storage," *Applied Energy*, vol. 332, Art. no. 120510, 2023, doi: 10.1016/j.apenergy.2022.120510.
30. Q. Liu, Y. Guo, L. Deng, H. Liu, D. Li, and H. Sun, "Residual deep reinforcement learning with model-based optimization for inverter-based Volt-Var control," *IEEE Trans. Sustain. Energy*, vol. 16, no. 1, pp. 269–283, Jan. 2025, doi: 10.1109/TSTE.2024.3454080.
31. Q. Liu, Y. Guo, and T. Xu, "Robust deep reinforcement learning for inverter-based Volt-Var control in partially observable distribution networks," *Applied Energy*, vol. 399, Art. no. 126445, Dec. 2025, doi: 10.1016/j.apenergy.2025.126445.
32. A. Deanseekeaw, W. Pinthurat, and B. Marungsri, "Multi-objective-based multi-heterogeneous-agent deep reinforcement learning for minimization of voltage deviation and operation cost in active distribution system," *IEEE Access*, vol. 13, pp. 75648–75665, 2025, doi: 10.1109/ACCESS.2025.3565123.
33. S. Bera and N. M. Pindoriya, "A deep reinforcement learning-based grid management framework for PV-rich smart distribution systems," *IEEE Trans. Ind. Informat.*, vol. 22, no. 7, pp. 6082–6093, Jul. 2026, doi: 10.1109/TII.2026.3674845.
34. C. S. Lai, D. Chen, J. Zhang, X. Zhang, X. Xu, G. A. Taylor, and L. L. Lai, "Profit maximization for large-scale energy storage systems to enable fast EV charging infrastructure in distribution networks," *Energy*, vol. 259, Art. no. 124852, Nov. 2022, doi: 10.1016/j.energy.2022.124852.


© 2026 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints.org | NOT PEER-REVIEWED | Posted: 19 October 2026 https://doi.org/10.20944/preprints202610.1652.v1



35. Y. Shiferaw, M. Kiros, K. Yeneneh, and G. Sufe, "Degradation-constrained multi-agent reinforcement learning with centralized training and decentralized execution for vehicle-to-grid optimization in renewable-dominated distribution networks," *Scientific Reports*, vol. 16, Art. no. 25119, 2026, doi: 10.1038/s41598-026-55471-3.
36. M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," *IEEE Trans. Power Del.*, vol. 4, no. 2, pp. 1401–1407, Apr. 1989, doi: 10.1109/61.25627.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

© 2026 by the author(s). Distributed under a Creative Commons CC BY license.