International Journal of Engineering Trends and Technology
ISSN: 2231 – 5381 / https://doi.org/10.14445/22315381/IJETT-V71I2P222

Volume 71 Issue 2, 186-196, February 2023

Original Article

Voltage Control on Distributed Generation Systems based on Multi-Agent Reinforcement learning approach

Tlotlollo S Hlalele[1,2], Yanxia Sun[2], Zenghui Wang[3]

[1,3]*Department of Electrical Engineering, University of South Africa, Johannesburg, South Africa.*
[2]*Department of Electrical and Electronic Engineering Science, University of Johannesburg, South Africa.*

[1,2]*Corresponding Author : hlalets@unisa.ac.za*

Received: 01 September 2022 Revised: 18 December 2023 Accepted: 02 February 2023 Published: 25 February 2023

Abstract *- The voltage control problem due to bidirectional power flows is more apparent when heterogeneous distributed generation systems (DGS) are integrated into the grid. In this paper, a novel method of voltage control in distributed generation systems based on a reinforcement learning technique is proposed. DGS incorporating renewable energy resources are highly complicated nonlinear dynamic systems. There are several challenges in employing the existing control methods. The novel method presented in this paper entrenches the Q learning algorithm into the voltage control problem of DGS. The Q-learning algorithm teaches agents responsible for decision taking in controlling the voltage and award the reward if the aim is achieved. The IEEE 9 bus test system with DG's integrated is used with various controlling agents connected. The results show significant improvement in the reliability of agent communication and the efficiency of the proposed method.*

Keywords *- Distributed generation, Reinforcement learning, Voltage control.*

1. Introduction

The significant task of the control system of a microgrid in the island mode is to control frequency and voltage as well as share the load between DGS. A microgrid has a hierarchical architecture of three levels, namely primary, secondary and tertiary control. Each microgrid comprises a minor generation with dissimilar capabilities and features. These are very difficult to control in nature. Considering the nonlinear dynamic behaviour of the system, the control hitch turns out to be multi-objective inhibited nonlinear in nature, which becomes very difficult to solve with the existing control methods. Multi-agent reinforcement learning (MARL) delivers a method for agents to develop active coordination policies without constructing a comprehensive decision model. MARL permits agents to discover the environment through trial and error and adjust their behaviors to the dynamics of the changeable and embryonic environment[1]. In some studies, RL is considered for optimal protection coordination. Relays are viewed as autonomous agents that can manipulate their time dial settings to respond optimally to signals from their environment, i.e., the power system. [2]This study does not consider bidirectional power flows or DGS. State-coordinated voltage control in an active distribution network is studied in [3]. In coordinated reinforcement learning, agents organize both their action choice of activities and their parameter updates. Within the limits of our parametric illustrations, the agents will determine a jointly ideal action without considering every possible action in their exponentially large joint action space.[4]. In [5], a Multi-agent based voltage control is proposed. It uses a distribution feeder split into a series of overlapping line segments, each of which is allocated an agent. In order to adjust the voltage of that segment, the agent senses the voltage variables in that segment and develops the reactive power compensation required. Paramenter determination of voltage regulator is implement as a voltage control strategy in [6]. A study in fault classification using machine learning techniques and quarter-cycle fault signatures was carried out in [7]. Current-based feature vectors and separate voltage- were described using multi-resolution analysis and input to a two-stage classifier. In this study, the author only concentrates on the classification of faults. The multi-agent modeling and simulation are employed in [8] with distributed reinforcement learning to voltage control. In this application, DG's are not taken into consideration. Reinforcement learning has been considered for cyber-physical security assessment in power systems. [27].The proposed methodology considers transitions of the attackers in the network based on critical contingency pairs of N-2. An online Q-learning reinforcement scheme is designed to solve a Markov decision-making process that models adversarial behaviour, not for voltage control. [10], investigates various techniques of attack and cascading failures from the perspectives of the attackers while the protection strategies of the defenders or operators are ignored. Game theoretical methods are applied to attacker-defender games in the smart grid security area. A new approach for a

multistage game (also called a dynamic game) between the attacker and the defender based on reinforcement learning to establish the optimum sequences of the attack given certain goals (e.g., transmission line outages or loss of generation) is used.

The reinforcement learning technique is a goal-directed computational approach where a computer learns to perform a task by interacting with an unknown dynamic environment. It is further a machine learning approach based on the interaction of agents with the environment and is learning what action to take to maximize the gain. Unlike supervised machine learning, where the agent is not told what action or decision to take but should be able to decide what actions to take to achieve the best reward (gain and losses).

The RL follow the Markov decision process (MDP), where the agent monitors the current state s_t of the surroundings and decide to take an action a_t. The environment then enters the new state (s_{t+1}) with the probability of $T(s_t; a_t; s_{t+1})$. Then an agent gets the reward of r_{t+1} with the probability of $R(s_t; a_t; s_{t+1}; r_{t+1})$. The RL algorithm depends on how well the agent can best take action based on the current state. The rate of captivating an action a in a state s under a policy π is anticipated to return when captivating the action a in the state s following the policy π.

There are two different actions to be followed which are exploration and exploitation, which are chosen based on the policy π such as $\varepsilon - greedy$. The action taken by the agent is expected to affect the future environment state.

The Q value is defined as the reward function based on the history and the observation of the state of the environment.

The microgrid has emerged as an alternative mode of operation, especially with the connection of DEG's to the grid. Figure 1 shows the universal design of a microgrid control system. A control signal can be written as [11]

$$\{x_i = f_i(x_i) + k_i(x_i)D_i + g_i(x_i)u_i\} \quad (1)$$

$$y_i = h_i(x_i) \quad (2)$$

Where,
x_i and $u_i = [\delta_i\ P_i\ Q_i\ \varphi_{di}\ \varphi_{qi}\ \gamma_{di}\ i_{ldi}\ i_{lqi}\ v_{odi}\ v_{oqi}\ i_{odi}\ i_{oqi}]^T$ are control signals. $D_i = [\omega_{com} v_{bdi} v_{bqi}]$ and $f_i\ k_i\ g_i$ form internal dynamics of the DG's. For the i^{th} DG, $v_{odi} = v_{oi} = v_{oqi} = 0$, which yields to $y'_i = A_{yi} + B_{ui} + d_i$. Where $y_i = [v_{odi}\ v'_{odi}]^T$. d_i represents the disturbance in y_i. Where $e'_i = A_{ei} + B_{ui} + d_i$ Becomes the voltage error of the DG integrated into the microgrid. Figure 1 depicts a typical microgrid voltage control.

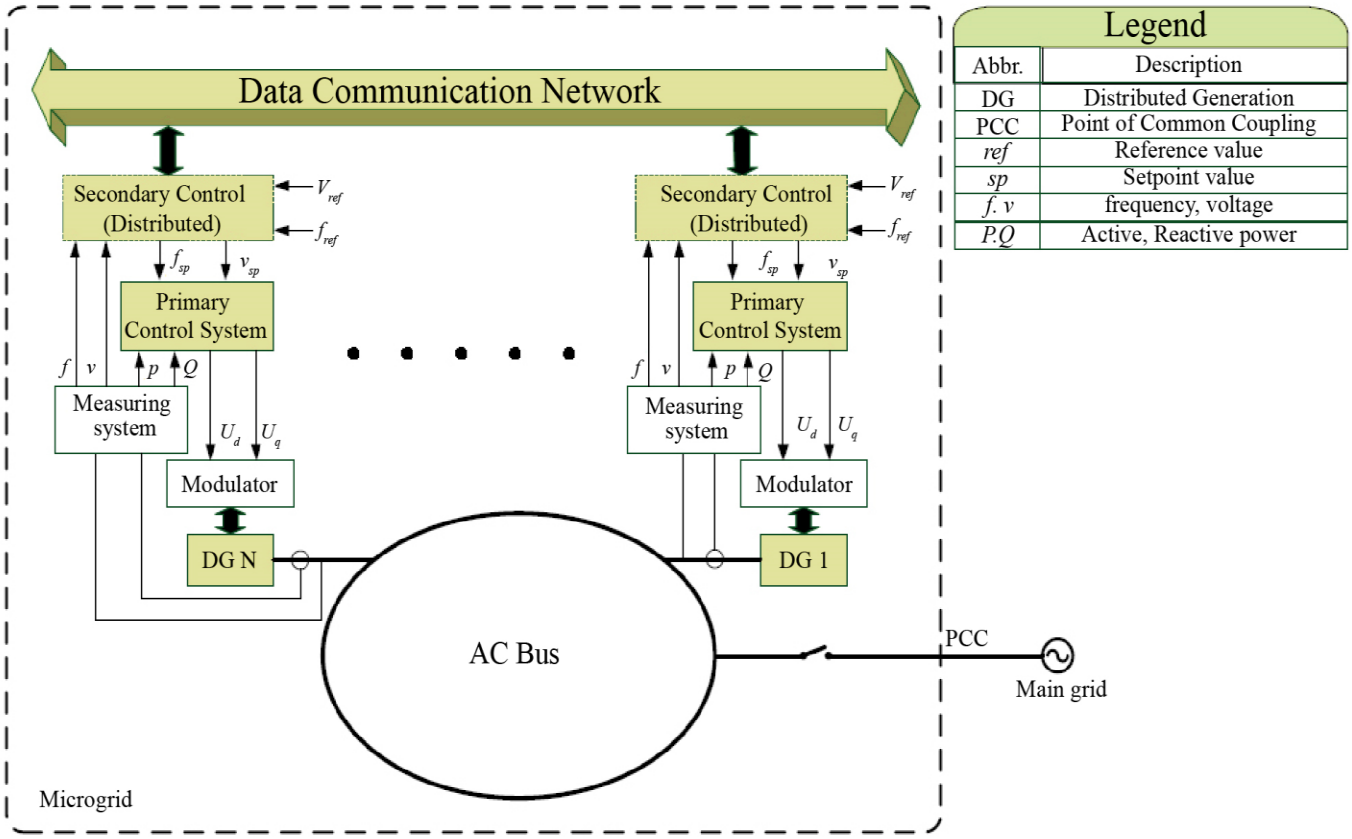

Fig. 1 A typical voltage control for a microgrid [11]

The non-relationship between nodal power injections and nodal voltages complicates the tension control problem solution. Also, models that follow linearizations that prove incapable of handling the rapid variations in volt in large networks effectively, as the synchronization of multiple components would require comprehensive communication. Distributed control schemes suggest subproblems, the decomposition of the voltage regulation problem to ease the global contact requirements[12][28]. Micro-grid voltage control based on the distributed cooperative control of multi-agent systems is proposed in [14][15]. The proposed secondary control is totally disseminated; each distributed generator needs only its in-formation and some neighbours ' information. The disseminated configuration removes the requirements for a central controller and network communication, which increases the systems' efficiency. Linearization of input–output feedback is used to transform secondary voltage power. The voltage control on islanded microgrid is studied in [16]. The microgrid is normally controlled by two loops with dissimilar bandwidths: the inner voltage control loop and the outer control loop. Here a linear quadratic regulator is suggested for voltage control. A new distributed controller for secondary frequency and voltage control in microgrids is presented in [17]. The proposed controller uses localized information and nearest neighbour information exchange to do secondary control actions.

In [18], the control method for the islanded microgrid is based on distributed cooperative control. A predictive control method applied to the secondary level of microgrids is suggested in [19]. It is based on droop and power transfer equations. These consider frequency and voltage regulation control objectives and consensus over the microgrid's real and reactive power contributions from each power unit. An innovative voltage control algorithm founded on peer-to-peer control and gossiping communication is suggested to function in a disseminated mode with no central coordinator[20]. The algorithm can be implemented asynchronously with partial data exchange between the agents. The control strategy employed in [29] uses fuzzy logic combined with a heuristic algorithm for coordinated Volt/Var control for the real-time operation of a smart distribution grid which implements simplicity and suitable performance for real-time application. Voltage regulation in an active power distribution system integrated with natural gas grids using distributed electric, and gas energy resources is presented in [22]. The new algorithm here is suggested for optimal real-time scheduling of power to gas and gas to the power unit to control over/under voltage issues in the active distribution system with the integration of renewables. Active-reactive coordinated optimization method is considered in [23]. A particle swarm optimization algorithm is presented to solve the problem of voltage deviation. Advanced voltage control for smart microgrids using distributed energy resources is suggested in [24]. The suggested method includes solving an optimization problem. Integrating distributed energy resources (DER's), which are necessary to improve the network capacity and meet the power demand, will change the whole architecture of the network into a multi-source network rendering the traditional control method inefficient.

Some of the contributions include the following:

1. The introduction of a Q-learning algorithm for voltage control.
2. The adoption of multi-agent reinforcement learning for a dynamic changing environment.

The paper is organised as follows; Section 2 presents a multi-agent reinforcement learning approach. In section 3 a Q-learning algorithm is introduced, section 4 subsequently discusses IEEE 9 bus with distributed energy resources, section 5 presents some results, and section 6 presents the conclusion.

2. Multi-Agent Reinforcement Learning

The multi-agent reinforcement learning technique is the approach in which agents are able to interact under certain environments through trial and error without making complete decisions. They can further adapt their behaviour to uncertain and evolving environments to better their performance. Figure 1 shows the agent environment action. The action taken by the agent is expected to affect the future environment state.

The goal of Q-learning is to learn a policy which tells an agent what action to take under what circumstances. It does not require a model of the environment and can handle problems with stochastic transitions and rewards without requiring adaptations. There are different algorithms which use MAS variants tabular Q-learning. They are as follows;

1. Markov Decision Process (MDP)
2. Independent learner
3. Coordinated reinforcement learning
4. Distributed value function

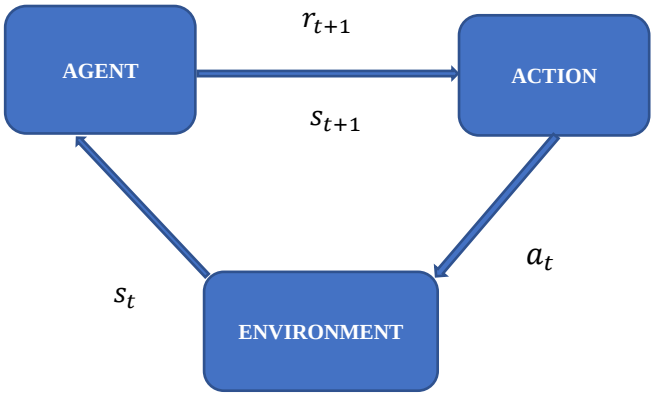

Fig. 2 Agent environment interaction

2.1. MDP Learners

The agent is not obliged to interact, but they must be able to monitor the effected joint actions and the received separate gain. According to [30], MDP is characterized by

Number of states $S = \{s_1, s_2, \ldots, s_n\}$ where s_t is a state in S; Number of actions $A = \{a_1, a_2, \ldots, a_M\}$ accessible to the agent per state s;

Alteration dissemination $T(s'|s,a)$ records a set comprised of a state s and an action a to prospect dissemination of state s';

A reward function $R: S \times A \times S \to R$ provides the probable reward when the agent builds the alteration from state s to state s' via action a. r_t represents the instant scalar reward gained at time t, where

$$r_t = R(S_{t+1} = s'|s_t = s, a_t = a) = E\{r_t|s_{t+1} = s', s_t = s, a_t = a \quad (1)$$

The dissemination of resulting states and rewards is autonomous of the history through the present state and action, such that

$$T(s_{t+1}|s_t a_t) = T(s_{t+1}|s_t, a_t, \ldots, s_1 a_1) \quad (2)$$

The action selection mechanism in MDP is $policy\ \pi: S \times A \to [0,1]$ that stipulates a prospect of choosing a in an exact s. The probable return in a state s.

$$V^{\pi}(s) = E_{\pi}[R_t|s_t = s] = E_{\pi}[\sum_{t=0}^{\infty} \gamma^t . r_t|s_t = s] \quad (3)$$

where γ is the discount factor and R_t signifies the gain. V^{π} is the gain of an agent resulting in the policy π. The action value function for policy π, $Q^{\pi}(s,a)$, is the anticipated gain when acquiring action a in state s under the policy π.

Therefore $Q^{\pi}(s,a) = E_{\pi}[R_t|s_t = s, a_t = a]$. The MDP's goal is to discover the preeminent policy π^* That exploits the probable gain. The optimal s_{value} for any state $is\ V^*(s) = max_{\pi} V^{\pi}(s)$.

Strategies used to find optimal policies, i.e. classification criterion for reinforcement learning approach, are as follows; $Value\ iteration$ updates each iteration according to the policy-given value function such that the current value function updates to intuitive

$$V_{t+1}^{\pi}(s) = \max_{a \in A} \sum_{s' \in S} T(s'|s,a)\,(R(s'|s,a) + \gamma . V_t^{\pi}(s')) \quad (4)$$

Policy reiteration progresses the feature of the policy π over after assessing the value function $V^{\pi}\ of\ the\ fixed\ policy\ \pi$.

Direct policy search. Here there, it is not necessary to realize the value function.

For independent learning

$$Q_i(s,a_i) = Q_i(s,a_i) + \alpha[R_i(s,a) + \gamma max\ a'_i\ Q_i(s',a'_i) - Q_i(s,a_i) \quad (5)$$

In this case, the agent overlooks the actions and gains of other agents.

For coordinated reinforcement learning, the agent must harmonize its action with a few agents and acts self-sufficiently within the environment.

$$Q_i(s_i,a_i) = Q_i(s_i,a_i) + \alpha[R(s,a) + \gamma \max a'\ Q(s',a') - Q(s,a)] \quad (6)$$

This method is disseminated and generates enormous storage and computational savings in the action space.

Distributed value functions

$$Q_i(s_i,a_i) = (1-\alpha)Q_i(s_i,a_i) + \alpha[R_i(s,a)\,\gamma \sum_{j \in (f(i,j) \neq 0} f(i,j) \max a'_i \quad (7)$$

3. Q-Learning Algorithm

$$\begin{cases} Initialize\ the\ Q-function\ and\ V\ values \\ Observe\ the\ current\ state\ s_t \\ Select\ action\ a_t\ and\ take\ it \\ Observe\ the\ reward\ r(s_t,a_t) \\ Perform\ the\ following\ updates\ and\ do\ not\ update\ any\ other\ Q-values \\ Q_{t+1}(s_t,a_t) \leftarrow (1-\alpha)Q_t(s_t,a_t) + \alpha_t\big(r(s_t,a_t) + \beta V_t(s_{t+1})\big) \\ V_{t+1}(s) \leftarrow max_a Q_t(s,a) \\ repeat \end{cases}$$

4. The IEEE 9 bus with Distributed Energy Resources

Consider the IEEE 9 bus with distributed energy resources where several agents have to be coordinated in order to be able to achieve one goal, which is to solve the problem of voltage deviation in the power system.

The voltage at a node in a network can be presented as a function of $P\ and\ Q$. That is $V = f(P, Q)$ where P and Q are the active and reactive power flow on the network.

$$P_i = \sum_{k=1}^{N} V_i V_k Y_{ik} \cos(\delta_i - \delta_k - \delta_{ik}) \quad (8)$$

$$Q_i = \sum_{k=1}^{N} V_i - V_k - Y_{ik} \cos(\delta_i - \delta_k - \delta_{ik}) \quad (9)$$

$Vi\ and\ \delta i$ is the amount of voltage and angle at node i; $Vk\ and\ \delta k$ is the amount of voltage and angle at node k; $Yik\ and\ \delta ik$ is the magnitude and argument of the element (i, k) in the network's admittance matrix[5]

Let the optimal power flow $P(i, u)$ for a pair of (i, u) with $i \in S\ and\ u \in A(i)$ be presented as

$$P(i,u) = \sum_{j \in S} q(i,u,j)\,[g(i,u,j + R^*(j)], \quad (10)$$

$$R^*(i) = min_{(u \in A(i))} \sum_{j \in S} q(i,u,j)[g(i,u,j) + R^*(j)] \forall i \quad (11)$$

Where $g(i, u, j)$ is the reward gained in the alteration from state i to state j under action u. Equating (10) and (11)

$$R^*(i) = min_{u \in A(i)} P(i,u) \forall i. \quad (12)$$

Equating (10) and (12) gives

$$P(i,u) = \sum_{j \in S} q(i,u,j)[g(i,u,j) + min_{v \in A(j)}\, P(j,v)] \forall (i,u). \quad (13)$$

$$P(i,u) \leftarrow (1-\gamma)P(i,u) + \gamma[g(i,u,j) + min_{v \epsilon A_j} P(j,v) \quad (14)$$

The average reward is approximated on the one-time scale and the active power on the other.

If P^x is the vector of P-values at the xth iteration. Let e^x be the sequence of states visited in the simulation till xth.

$$P^{x+1}(i,u) = P^x(i,u) + \alpha\big(m(x,i,u)\big)[g(i,u,e^x_{iu} + min_v Q^k(e^x_{iu}, v) - q^x - P^x(i,u)]I((i,u) = \emptyset^x \quad (15)$$

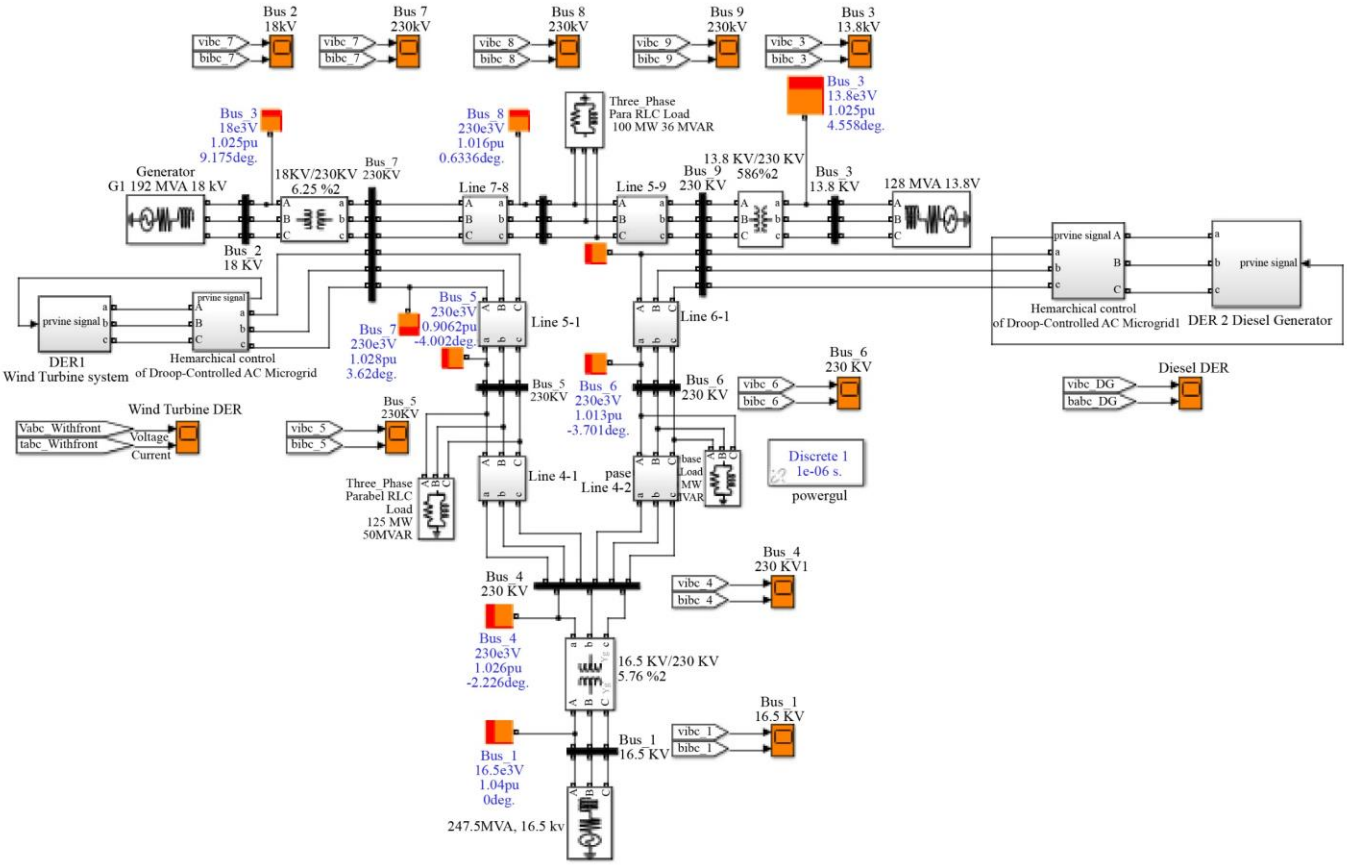

Fig. 3 A modified IEEE 9 bus

$$q^{x+1} = (1 - \beta(k)q^{x} + \beta(k)\frac{[(J(k)q^{x} + g(i,u,e_{iu}^{x})}{J(x+1)} \quad (16)$$

$J(x)$ is the number of state transition up to xth iteration, $\emptyset = (\emptyset^1, \emptyset^2, \ldots.)$ is the process of state action pairs tried in the learning process.

4.1. Primary Control

Adjusts the frequency and amplitude of the voltage reference provided to the inner current and voltage control loops, Simulates the behavior of a synchronous generator, and reduces the frequency when the active power increases using the P/Q droop method.

$$\omega = \omega * - Gp(s) \cdot (P - P*) \quad (17)$$

$$E = E * - G_Q(s) \cdot (Q - Q*) \quad (18)$$

Where ω is the frequency and E amplitude of the output voltage, $\omega *$ and $E *$ the respective references, P active and Q reactive power, $P *$ and $Q *$ the power references, and $GP(s)$ and $GQ(s)$ the respective transfer functions.

$$Gp\,(s) = m \quad (19)$$

$$m = \Delta\omega/2Pmax \quad (20)$$

$$GQ(s) = n \quad (21)$$

$$n = \Delta V/2Qmax \quad (22)$$

$\Delta\omega, \Delta V$, Pmax and Qmax are the maximum values for frequency, voltage, and active and reactive power delivered by the inverter, respectively.

Using power electronics, the output impedance depends on the controller, and the control droops (17), (18) can be modified according to Park's transformation determined by the impedance angle θ.

$$\omega = \omega * - GP\,(s)\,[(P - P*)\,sin\,\theta - (Q - Q*)\,cos\,\theta] \quad (23)$$

$$E = E * - GQ(s)\,[(P - P*)\,cos\,\theta + (Q - Q*)\,sin\,\theta] \quad (24)$$

The output voltage depends on the virtual output-impedance transfer function $Z_D(s)$

$$v_o^* = v_{ref} - Z_D(s) \cdot i_o \quad (25)$$

$$v_{ref} = E\,sin(\omega t) \quad (26)$$

$$Z_D(s) = L_D \frac{2k_1 s^2}{s^2 + 2\varepsilon\omega_1 s + \omega_1^2} + R_i \frac{2k_i s}{s^2 + 2\varepsilon\omega_o s + \omega_o^2} \quad (27)$$

Where ki is the filter's coefficient for every harmonic i term, LD inductive and Ri resistive impedance values, respectively.

4.2. Coordinated Secondary Voltage Control

Voltage control in power systems has three levels with different response times. In the secondary voltage control level, the objectives are to attain better voltage regulation and improvement of power system voltage stability for various system conditions. When a disturbance occurs in power systems, the area and intensity of the disturbance must keep as small as possible

Secondary control is proposed to solve the voltage deviation problem. The voltage level in the MG vMG is sensed and compared with the voltage reference v∗MG, and the error processed through a compensator is sent to all the units δvo to restore the output voltage.

The multiagent system with 11 agents (buses) and Two static compensators STATCOMs are assigned to nodes 7 and 9 in figure 3.

The secondary control ensures that the frequency and voltage deviations are compensated. The frequency and amplitude levels in the MG, ωMG and EMG are sensed and compared with the references ω∗ MG and E∗ MG; the errors processed through the compensators $\delta\omega$ and δE are sent to all the units to restore the output-voltage frequency and amplitude. The limit for the frequency deviation is defined as +-0.1 or 0.2 Hz.

$$dP = -\beta \cdot G - \frac{1}{T_r}\int Gdt \quad (28)$$

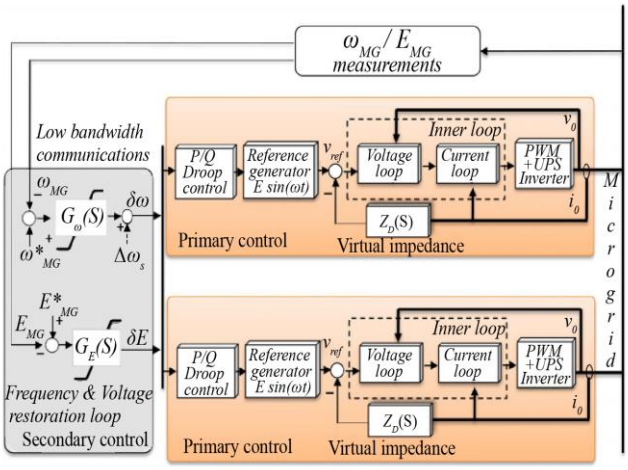

Fig. 4 Microgrid voltage control [26]

δP is the output set point, β is the proportional gain, Tr is the time constant, and G is the area control error (ACE), calculated in 5- to 10-s intervals.

$$G = P_{meas} - P_{sched} + K_{ri}(f_{meas} - f_0) \quad (29)$$

P_{meas} is the measured active power transferred at the PCC, P_{sched} is the resulting exchange program, K_{ri} is the proportional factor, f_{meas} is the instantaneous measured system frequency, and $f_{0\,is}$ the set-point desired frequency.

$$\delta\omega = kp\omega\,(\omega * MG - \omega MG) + ki\omega \int (\omega * MG - \omega MG)\,dt + \Delta\omega S \quad (30)$$

$$\delta E = kpE\,(E * MG - EMG) + kiE \int (E * MG - EMG)\,dt \quad (31)$$

$kp\omega, ki\omega, kpE$, and kiE are the control parameters of the secondary-control compensator, and $\Delta\omega S$ is a synchronization term which remains equal to zero when the grid is not present.

The phase between the grid and the MG will be synchronized with a conventional phase-locked loop PLL, in which the output signal $\Delta\omega S$ will be added to the secondary control and sent to all the modules to synchronize the MG phase, connected to the main grid through a static bypass switch.

4.3. Tertiary Control

It changes the phase in the steady state by adjusting the frequency and amplitude to obtain the desired waveform; by measuring the P/Q through the static bypass switch, PG and QG can be compared with the required P∗ G and Q∗ G. The control laws P_{IP} and P_{IQ} are given as;

$$\omega^*_{MG} = kpP\,(P * G - PG) + kiP \int (P * G - PG)\,dt \quad (32)$$

$$E^*_{MG} = kpQ\,(Q * G - QG) + kiQ \int (Q * G - QG)\,dt \quad (33)$$

5. Results and Analysis

5.1. Time Domain Evaluation

Three different types of Power System Stabilizers (PSS), Multi-Band PSS (MBPSS), Generic Delta Omega PSS, and Generic Delta Pa PSS, are installed in each generator PSS was once utilised to address issues with vibration stability. The improved performance and functionality of PSS have been proven over the past few decades in numerous references and real-world applications. Therefore, PSS models are an essential component of any instrument used for power system modelling.

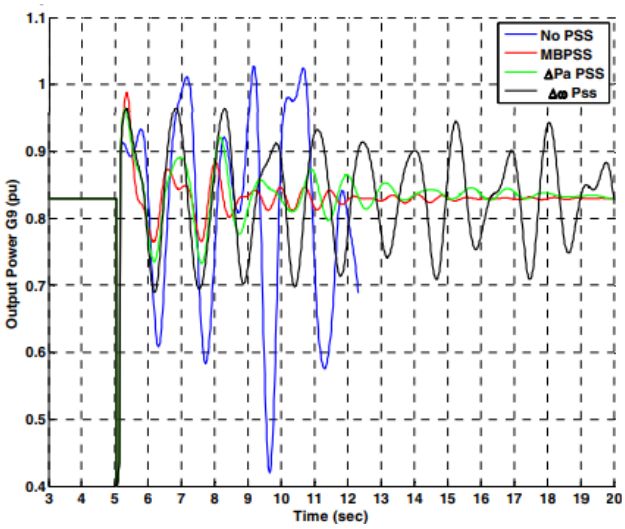

Fig. 5 The G9's active power at the NE9bus with or without PSSs.

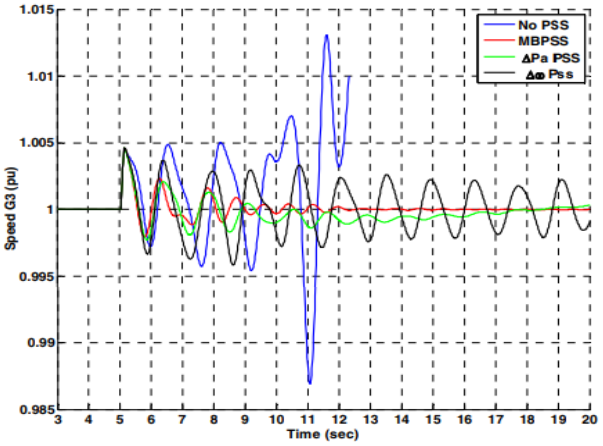

Fig. 6 G3 at NE9bus speed with or without PSSs.

It is simulated that a three-phase fault would resolve at t = 5 s after 6 cycles without damaging equipment. After t=12 seconds, the test system loses synchrony in the absence of PSS (No PSS).

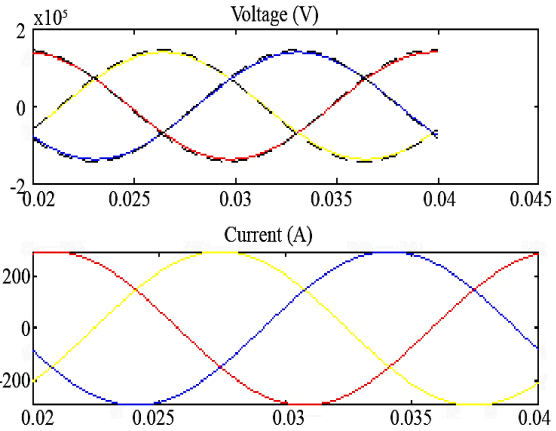

Fig. 7 Wind turbine voltage and current after control at agent bus 7 (230kV)

Table 1. System data for RL model.

	rlDDPGAgent	rlDDPGAgent_1	rlDDPGAgent_2
Status	Training finished	Training finished	Training finished
Episode number	2	2	2
Episode reward	-10	-10	10
Episode steps	10	10	10
Total agent steps	20	20	20
Average reward	-10	-10	-10
Average steps	10	10	10
Episode Q0	-47163824	-6759870	-0.162031292915
Averaging window length	5	5	5
Training stopped by	MaxEpisodes	MaxEpisodes	MaxEpisodes
Training stopped at	Episode 2	Episode 2	Episode 2

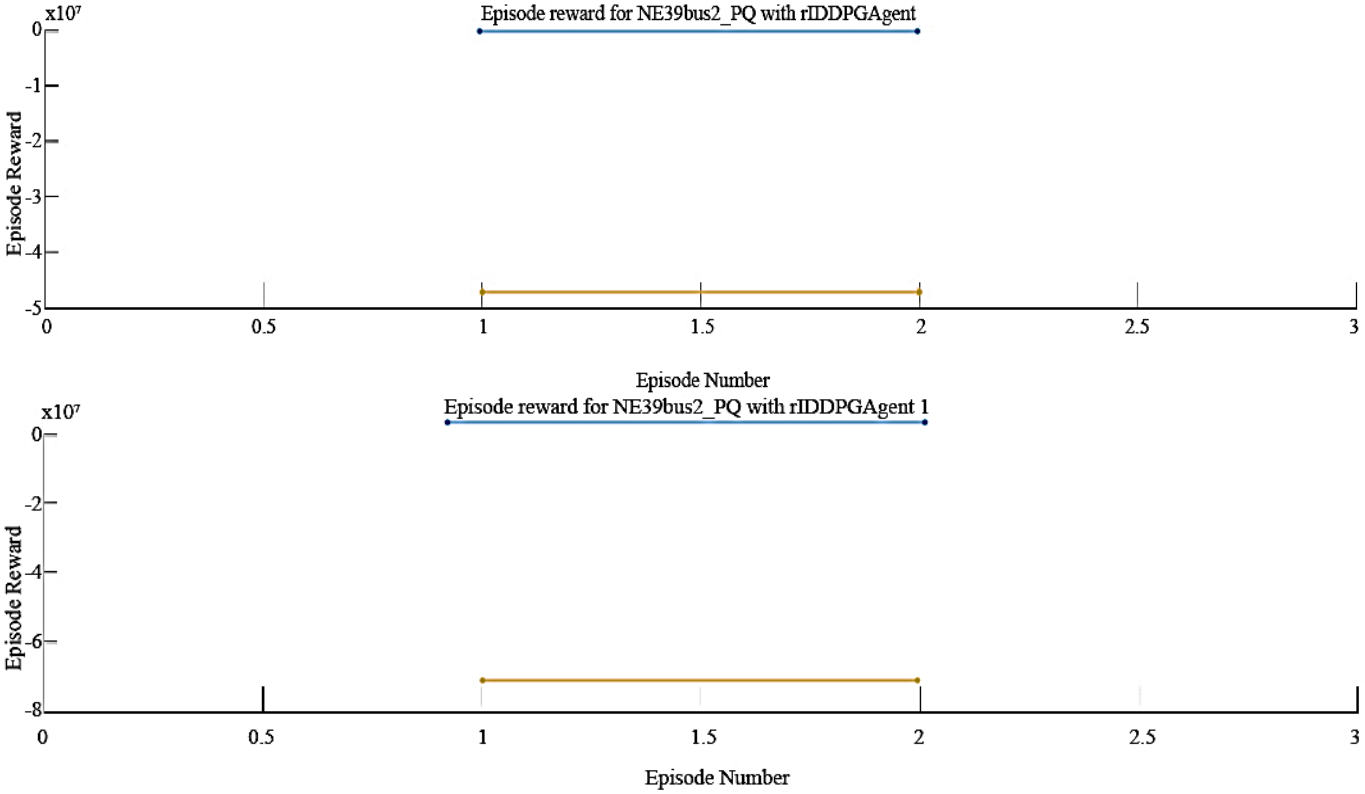

Fig. 8 Episode reward for NE9bus2_PQ with rIDDPGAgent_1

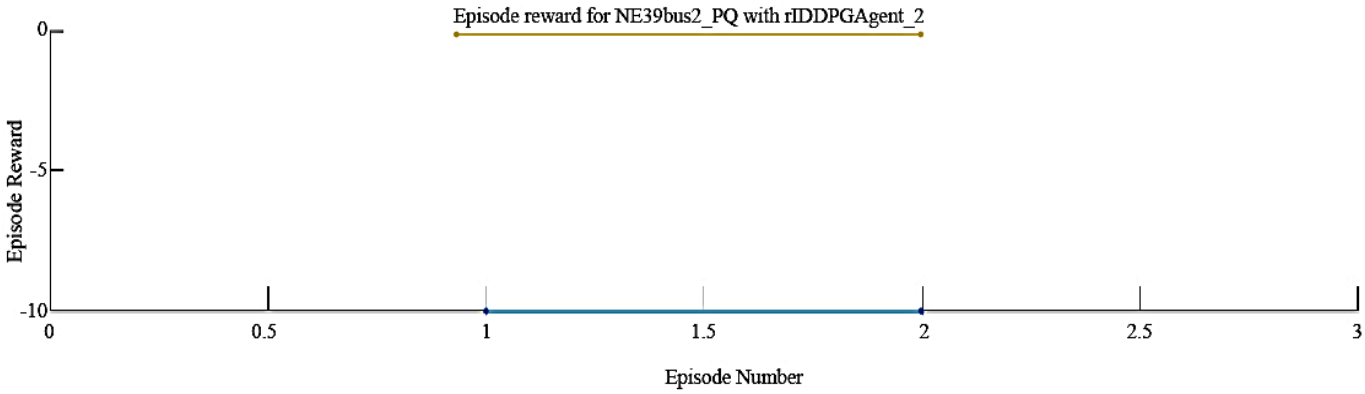

Fig. 9 Episode reward for NE9bus2_PQ with rIDDPGAgent_2

5.2. Output from reinforcement learning for last N episodes 2

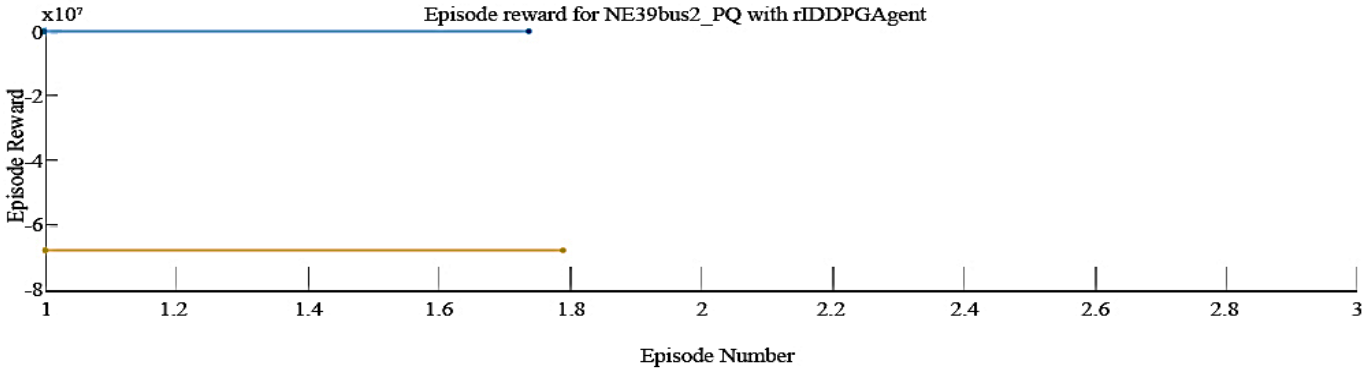

Fig. 10 Episode reward for NE9bus2_PQ with rIDDPGAgent

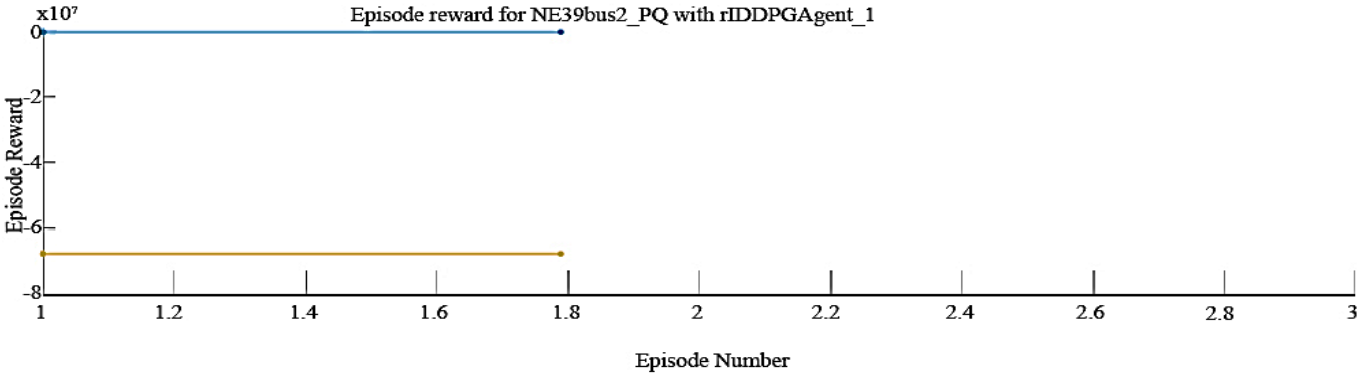

Fig. 11 Episode reward for NE9bus2_PQ with rIDDPGAgent_1

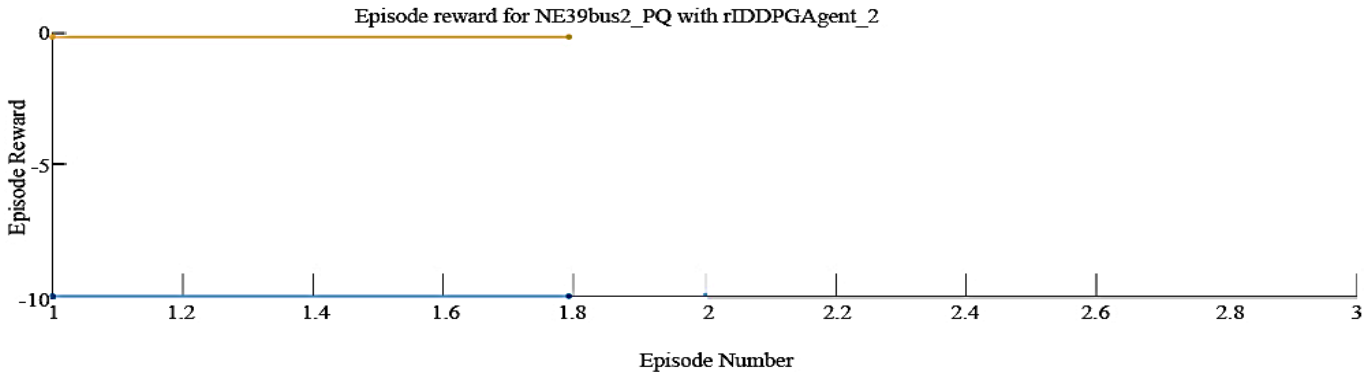

Fig. 12 Episode reward for NE9bus2_PQ with rIDDPGAgent

The output of the reinforcement learning from the episodes suggests that the agent took the incorrect decision at some point as the reward was negative. The positive reward indicates that the correct decision was taken by the agent in the environment, thereby indicating that adequate control is achieved.

6. Conclusion

The voltage control problem is exploited with multi-agent reinforcement learning (MARL). MARL offers an appealing method for agents to evolve applicable synchronisation of policies without explicitly building a complete decision model. MARL allows agents to discover the environment through trial and error and adapt their behaviours to the dynamics of the changeable environment. Q-learning algorithm is shown and applied to a voltage control problem. The main benefit of the proposed solution is that the agents can solve the voltage control problem without any central controller's interference and only by contact between neighbouring agents. The algorithm eliminates the potential for conflict between the agent's control behaviour and reduces the effect of communication failure to boost the control method's robustness. The reward is awarded once the aim is achieved.

Funding Statement

This research is supported partially by South African National Research Foundation Grants (No.120106, 141951 and 132797), South African National Research Foundation Incentive Grant (No. 132159, and the GES fund of the University of Johannesburg.

References

[1] Chongjie Zhang, and Victor Lesser, "Coordinating Multi-Agent Reinforcement Learning with Limited Communication," *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems*, pp. 1101–1108, 2013.

[2] Hasan Can Kiliçkiran, Bedri Kekezoglu, and Nikolaos G. Paterakis, "Reinforcement Learning for Optimal Protection Coordination," *2018 International Conference on Smart Energy Systems and Technologies*, pp. 1-6, 2018. *Crossref*, http://doi.org/10.1109/SEST.2018.8495830

[3] Praveen Prakash Singh, and Ivo Palu, "State Coordinated Voltage Control in an Active Distribution Network with On-Load Tap Changers and Photovoltaic Systems," *Global Energy Interconnection*, vol. 4, no. 2, pp. 117–125, 2021. *Crossref*, http://doi.org/10.1016/j.gloei.2021.05.005

[4] Carlos Guestrin, Michail Lagoudakis, and Ronald Parr, "Coordinated Reinforcement Learning".

[5] Patrick T. Manditereza, and Ramesh C. Bansal, "Multi-Agent Based Distributed Voltage Control Algorithm for Smart Grid Applications," *Electric Power Components and Systems*, vol. 44, no. 20, pp. 2352–2363, 2016. *Crossref*, http://doi.org/10.1080/15325008.2016.1219889

[6] Won Nam Koong et al., "Voltage Control of Distribution Networks to Increase their Hosting Capacity in South Korea," *Journal of Electrical Engineering & Technology*, vol. 16, no. 3, pp. 1305–1312, 2021. *Crossref*, http://doi.org/10.1007/s42835-021-00682-z.

[7] Nicholas S. Coleman, Christian Schegan, and Karen N. Miu, "A Study of Power Distribution System Fault Classification with Machine Learning Techniques," *2015 North American Power Symposium*, pp. 1–6, 2015. *Crossref*, http://doi.org/10.1109/NAPS.2015.7335264

[8] M.Reza Tousi et al., "A Multi-Agent-Based Voltage Control in Power Systems Using Distributed Reinforcement Learning," *SIMULATION*, vol. 87, no. 7, pp. 581–599, 2011. *Crossref*, http://doi.org/10.1177/0037549710367904

[9] S. Anudevi, and Vinayak N. Shet, "Distribution Network with Optimal DG Placement and Protection Impacts: Review Analysis," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 4, no. 2, pp. 12-16, 2017. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V4I2P103

[10] Zhen Ni, and Shuva Paul, "A Multistage Game in Smart Grid Security: A Reinforcement Learning Solution," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 9, pp. 2684–2695, 2019. *Crossref*, https://doi.org/10.1109/TNNLS.2018.2885530

[11] Ali Moradi Amani et al., "Voltage Control in Distributed Generation Systems Based on Complex Network Approach," *Energy Procedia*, vol. 110, pp. 334–339, 2017. *Crossref*, https://doi.org/10.1016/j.egypro.2017.03.149

[12] Kyriaki E. Antoniadou-Plytaria et al., "Distributed and Decentralized Voltage Control of Smart Distribution Networks : Models, Methods, and Future Research," *IEEE Transactions on Smart Grid*, vol. 8, no. 6, pp. 2999–3008, 2017. *Crossref*, https://doi.org/10.1109/TSG.2017.2679238

[13] Mulumudi Rajesh, and A. Lakshmi Devi, "Wind, PV Solar, Hydro and Hybrid Energy Storage System-Based Intelligent Adaptive Control for Standalone Distributed Generation System," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 9, no. 11, pp. 67-94, 2022. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V9I11P108

[14] Ali Bidram et al., "Distributed Cooperative Secondary Control of Microgrids Using Feedback Linearization," *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3462–3470, 2013. *Crossref*, https://doi.org/10.1109/TPWRS.2013.2247071

[15] Zhibin Liu et al., "Simulation Analysis of Distributed Micro-Energy Clusters Participating in Voltage Regulation and Peak Regulation," *Energy Reports*, vol. 7, pp. 1529–1543, 2021. *Crossref*, https://doi.org/10.1016/j.egyr.2021.09.089

[16] Tine L. Vandoorn et al., "Voltage Control in Islanded Microgrids by means of a Linear-Quadratic Regulator," *IEEE Benelux Young Researchers Symposium*, 2010.

[17] John W. Simpson-Porco et al., "Secondary Frequency and Voltage Control of Islanded Microgrids via Distributed Averaging," *IEEE Transactions on Industrial Electronics*, vol. 62, no. 11, pp. 7025–7038, 2015. *Crossref*, https://doi.org/10.1109/TIE.2015.2436879

[18] Xiangyu Wu, Chen Shen, and Reza Iravani, "A Distributed, Cooperative Frequency and Voltage Control for Microgrids," *IEEE Transactions on Smart Grid*, vol. 9, no. 4, pp. 2764–2776, 2018. *Crossref*, https://doi.org/10.1109/TSG.2016.2619486

[19] Juan S. Gómez et al., "Distributed Predictive Control for Frequency and Voltage Regulation in Microgrids," *IEEE Transactions on Smart Grid*, vol. 11, no. 2, pp. 1319–1329, 2020. *Crossref*, https://doi.org/10.1109/TSG.2019.2935977

[20] Jonas Engels, Hamada Almasalma, and Geert Deconinck, "A Distributed Gossip-based Voltage Control Algorithm for Peer-to-Peer Microgrids," *2016 IEEE International Conference on Smart Grid Communications*, pp. 370–375, 2016. *Crossref*, https://doi.org/10.1109/SmartGridComm.2016.7778789

[21] Modu Abba Gana, Ishaku Abdul Dalyop, and Musa Mustapha, "Assessment of Reliability of Distribution Network with Embedded Generation," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 7, no. 2, pp. 34-37, 2020. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V7I2P107

[22] Nader A. El-Taweel, Hadi Khani, and Hany E.Z. Farag, "Electrical Power and Energy Systems Voltage Regulation in Active Power Distribution Systems Integrated with Natural Gas Grids Using Distributed Electric and Gas Energy Resources," *International Journal of Electrical Power & Energy Systems*, vol. 106, pp. 561–571, 2019. *Crossref*, https://doi.org/10.1016/j.ijepes.2018.10.037

[23] Ge Shaoyun et al., "Coordinated Voltage Control for Active Distribution Network on District Distribution Heating and Cooling Coordinated Voltage Control for Active Network Considering the Impact of Energy Storage," *Energy Procedia*, vol. 158, pp. 1122–1127, 2019. *Crossref*, https://doi.org/10.1016/j.egypro.2019.01.277

[24] P. C. Olival, A. G. Madureira, and M. Matos, "Advanced Voltage Control for Smart Microgrids Using Distributed Energy Resources," *Electric Power Systems Research*, vol. 146, pp. 132–140, 2017. *Crossref*, https://doi.org/10.1016/j.epsr.2017.01.027

[25] Um-e-Batool et al., "Controller for Voltage Profile Improvement of Double Fed Induction Generator based Wind Generator," *SSRG International Journal of Electrical and Electronics Engineering*, vol. 7, no. 12, pp. 21-26, 2020. *Crossref*, https://doi.org/10.14445/23488379/IJEEE-V7I12P104

[26] Josep M. Guerrero et al., "Hierarchical Control of Droop-Controlled AC and DC Microgrids - A General Approach Toward Standardization," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 1, pp. 158–172, 2011. *Crossref*, https://doi.org/10.1109/TIE.2010.2066534

[27] Xiaorui Liu, and Charalambos Konstantinou, "Reinforcement Learning for Cyber-Physical Security Assessment of Power Systems," *2019 IEEE Milan PowerTech, PowerTech*, pp. 1–6, 2019. *Crossref*, https://doi.org/10.1109/PTC.2019.8810568

[28] Aleksandar Boričić, Jose Luis Rueda Torres, and Marjan Popov, "Fundamental Study on the Influence of Dynamic Load and Distributed Energy Resources on Power System Short-Term Voltage Stability," *International Journal of Electrical Power & Energy Systems*, vol. 131, 2021. *Crossref*, https://doi.org/10.1016/j.ijepes.2021.107141

[29] Ana Paula Carboni de Mello, Luciano Lopes Pfitsche, and Daniel Pinheiro Bernardon, "Coordinated Volt / VAr Control for Real-Time Operation of Smart Distribution Grids," *Electric Power Systems Research*, vol. 151, pp. 233–242, 2017. *Crossref*, https://doi.org/10.1016/j.epsr.2017.05.040

[30] Madalina M. Drugan, "Reinforcement Learning Versus Evolutionary Computation : A Survey on Hybrid Algorithms," *Swarm and Evolutionary Computation*, vol. 44, pp. 228–246, 2019. *Crossref*, https://doi.org/10.1016/j.swevo.2018.03.011