Development approach of a volt-var control for inverter-coupled renewable energy plants using deep reinforcement learning

Jurek Türk[1], Merten Schuster[1]*, Hartmudt Köppe[1], Bernd Engel[1]

[1]elenia Institute for High Voltage Technology and Power Systems, Brunswick, TU Braunschweig, Germany
**E-mail: m.schuster@tu-braunschweig.de*

Keywords: VOLT-VAR CONTROL, DEEP REINFORCEMENT LEARNING, SOFT ACTOR CRITIC, REACTIVE POWER, RENEWABLE ENERGY SOURCES

Abstract

The connection of RES and the disconnection of controllable synchronous generators pose new challenges for grid management and steady-state voltage control. Removed reactive power sources that have contributed to voltage stability need to be replaced. One approach is the use of inverter-coupled RES as reactive power sources and the application of Artificial Neural Networks (ANN) for optimised reactive power provision. Several studies have already started to explore the possibilities of Deep Reinforcement Learning (DRL) to develop new reactive power control methods for RES. While other studies have mostly looked at the control of a single unit in a low-voltage grid, this study develops a volt-var control for multiple RES investigated in different operating scenarios for a whole year in a CIGRE benchmark medium-voltage distribution grid. The results show that the DRL-trained ANN can control the reactive power supply with a lower voltage deviation compared to conventional methods. However, this is at the cost of more frequent equipment overloads due to an increase in reactive power flows. The application of this developed approach will be further explored in order to validate the methodology in real grid models in the future.

1 Introduction

The strong growth of renewable energy sources (RES) such as solar and wind power is changing the dynamics of electric power systems worldwide. This is particularly evident in Germany, where the implementation of the German energy transition, known as the Energiewende, is leading to significant structural changes within the power system. In particular, the integration of RES into the distribution grid and the shutdown of controllable synchronous generators in the transmission grid pose new challenges for grid control and the associated dispatch of ancillary services. [1]

One of the important tasks of grid control is voltage control, an non-frequency ancillary service that helps maintain voltage within a specified range. Voltage control involves maintenance of voltage through various methods, such as the provision of reactive power from connected reactive power sources. The aforementioned changes due to the Energiewende also have an impact on steady-state voltage control. The deactivated reactive power sources that contributed to voltage stability in the past must be replaced by equivalent sources. In addition, supplying reactive power from new reactive power sources needs to be adaptive, target-oriented and efficient given the ongoing changes in the power system. This leads to the central motivation for this paper.

1.1 Background and Motivation

Traditionally, voltage control in power systems has been performed using rule-based methods, such as Q(P) or Q(V). These methods have proven to be effective, especially Q(V) see [2]. For this reason, the methods are also established in the context of different Technical Connection Rules (TCR), e.g. in Germany. The problem with these methods is the lack of adaptation of the characteristics to the continuously changing operating scenarios in the grid. This can lead to a lack of interoperability with other methods already implemented, resulting in inefficiencies in voltage control.

Given the challenges, the question is whether there are methods that can implement adaptive and efficient voltage control and whether these methods perform better than the aforementioned conventional methods. To address this question, this paper introduces a novel approach to voltage control that is using deep reinforcement learning (DRL) for inverter-coupled RES. The primary objective of this research is to develop and evaluate a method for voltage control by injection or absorbing reactive power, commonly known as volt-var control, which adapts to the dynamic and uncertain nature of RES. [3] The hypothesis underlying this research is that a DRL-based control method can provide a more adaptive, targeted, and efficient solution compared to conventional rule-based approaches. [4] In order to place this hypothesis in the state of the art, existing literature and related work in this field is presented below.

1.2 Literature and Related Work

In the context of steady state voltage control and the challenges associated with it, a variety of different control methods are proposed in the literature [5] and, as mentioned previously, are also effectively used in practice [6]. In addition to the Q(P) or Q(V) control methods mentioned, which are distributed,

autonomous and measurement-based control methods, coordinated Q set-point approaches via remote parameterisation are also state of the art in grid control. In addition, the ongoing development of communication and computing technologies has led to the development of algorithms for distributed voltage control. [7]

In recent years, there has been a significant increase in research and definition in the field application of machine learning in grid operations. The ability of machine learning to automatically adapt to changing scenarios by analysing and learning from output data offers promising opportunities for improving grid operations and integrating RES. In particular, Deep Learning, a branch of machine learning that uses deep neural networks, has shown great promise. These networks are able to autonomously analyse large amounts of data, extract key features and make accurate predictions, which is particularly useful for tasks such as forecasting, detecting anomalies and optimising power flows in networks. [8]

DRL is another branch of machine learning that uses deep neural networks. In reinforcement learning, an agent learns to make decisions by performing actions in an environment to maximise a cumulative reward. Unlike supervised learning, where models are trained using labelled data, and unsupervised learning, where models recognise patterns in unlabelled data, DRL trains models based on feedback from interactions with the environment. [9] The application of DRL in grid operation will have a significant role due to its potential for control and optimisation tasks. [8]

The use of DRL is already part of several research papers related to the optimisation of reactive power control and power flow. One of the earlier works on this subject is [10], where Vlachogiannis et al. use the Q-learning algorithm, which is one of the most popular model-free reinforcement learning algorithms, to successfully simulate the optimal control of reactive power within the physical and operational constraints in an IEEE 14-bus system and an IEEE 136-bus system. It is shown that the method gives better results compared to conventional methods. [10] In the context of voltage var control and reactive power provision of inverter-coupled RES, publications on the topic have increased, especially in recent years. For example, Li et al. 2019 paper explores the use of DRL to coordinate multiple smart inverters of PV systems to improve grid voltage, feed-in curtailment while reducing system losses. [3] Also Beyer et al. [4] or Utama et al. [11] show that DRL is suitable for the training of artificial neural networks for PV systems to improve the provision of reactive power in the scope of voltage control.

This paper builds on the results of the above-mentioned papers and is organised as follows. Section 2 presents the approach to implement and evaluate DRL-based voltage var control within a distribution grid. Section 3 deals with the simulation and evaluation of the results of the implemented models in different scenarios that represent the real conditions of the RES plants within the voltage control. The results of the models are compared with conventional methods, among others. Finally, Section 4 concludes the development approach presented and summarises possible directions for future research and development.

1.3 Key Contributions

The key contributions of this paper include:

1. Open source development approach of a DRL-based model for centralised and decentralised volt-var control.
2. A comprehensive evaluation of the DRL model compared to conventional methods.
3. Insights into the potential for DRL in grid control applications.

2 Methodology

This section outlines the approach to implementing and evaluating a DRL-based volt-var control within a distribution grid. The methodology is essential to establish a realistic environment for testing the effectiveness of volt-var control methods. Figure 1 below presents a simplified schematic structure of the implemented framework.

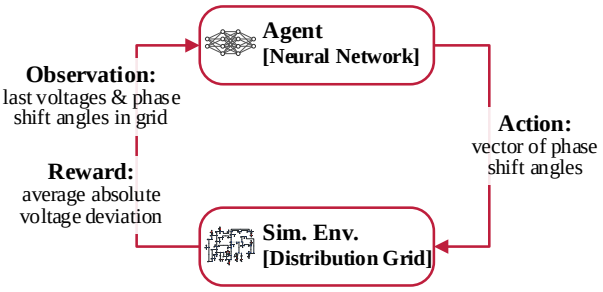

Fig. 1 Simplified schematic structure of the volt-var control framework

The figure shows an agent that uses a policy with artificial neural networks (ANN) to determine the phase shift angles for the RES in the environment. The environment contains a distribution grid and runs power flow simulations. Depending on the outcomes of the power flow simulations, a numeric reward is calculated and passed back to the agent together with an observation vector of the distribution grid. The reward is a numeric indicator how close the network voltages were to the nominal voltages in the last power flow simulation. The agent subsequently uses the reward and observation to adapt the policy in order to maximise the reward. The following subsections provide more detailed descriptions of the agent, the environment, training and evaluation process.

2.1 Deep Reinforcement Learning Volt-Var Control Training

To achieve optimal volt-var control of inverter-coupled RES plants, a robust training procedure is essential. This subsection describes the implemented approach for training, which is visually represented in Figure 2.

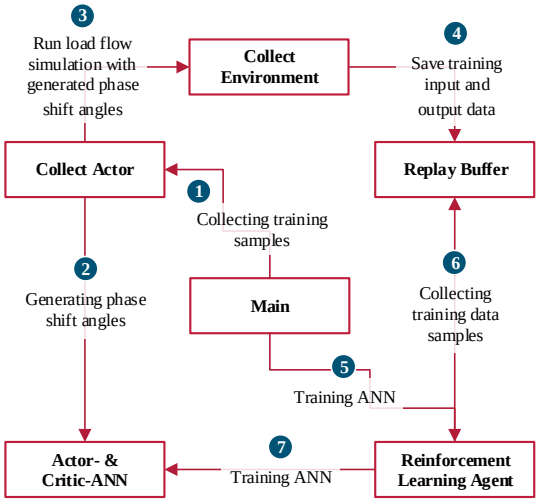

Fig. 2. Training procedure

During each step of a training episode 1) a new training data sample is created by 2) first generating a new set of phase shift angles using the actor net and the last voltages and phase shift angles of the distribution grid. In 3) a distribution grid simulation is performed using the new phase shift angles together with the generation and load data for the next time step. A training data sample is created by combining the input phase shift angles, the output voltages of the distribution grid and the calculated reward. This sample is then stored in a replay buffer in step 4). In the second part of a training step, started by calling 5), random samples are taken from the replay buffer in 6) and passed to the agent for ANN training in 7).

The replay buffer is initialised with data comprising 50,000 random input phase shift angles. Subsequently, the ANN is trained for 1,500,000 steps, which is equivalent to approximately 504 episodes. The training process involves several steps, followed by the evaluation of the ANN over a separate data-set to measure progress and improvement. Details about the evaluation and training data-set, which depend on the load and generation time series used, can be found in subsection 2.3. Once the training is complete, the subsequent step is to evaluate the performance of the ANN. Figure 3 shows the evaluation process.

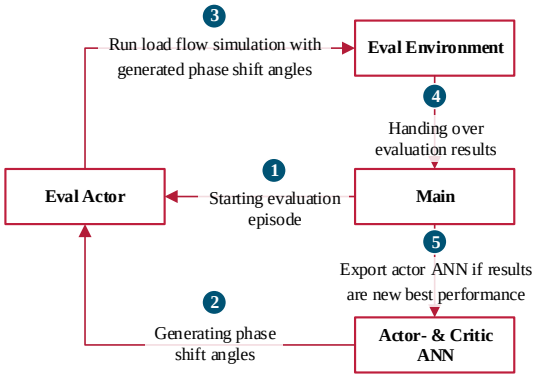

Fig. 3. Evaluation procedure

The evaluation process closely resembles the training process, with the exception that data is not stored in a replay buffer, and the ANN is not subjected to further training. To facilitate the assessment of the learning process, data generated during the evaluation is recorded in an SQL database. An episode consists of a training procedure and a subsequent evaluation procedure.

2.2 Deep Reinforcement Learning Actor

The selection of the DRL actor is crucial for the implementation of the volt-var control of the RES. In this study, a Soft Actor-Critic (SAC) agent is used, as described in [12], which uses multiple ANN to calculate and adapt its policy. The ANN of the agent are trained with the Tensorflow library [13] using its SAC-Agent. SAC is an off-policy form of reinforcement learning that aims to maximise the total expected rewards. This is done by implementing 'double Q-learning' [14] to initially estimate these rewards. A mathematical principle, the Kullback-Leibler divergence, is then implemented to adjust the decision policy. This approach involves two key components: a 'critic' ANN that estimates the expected rewards and an 'actor' ANN that selects the most suitable action based on the current state of the environment.[12] In this study, both the actor and critic networks are designed identically, featuring two fully-connected layers, each with 100 neurons.

2.3 Simulation Environment

The previously mentioned agent is trained in a simulation environment specifically developed for this study. The environment uses a CIGRE benchmark medium voltage grid containing multiple distributed RES as described in [15] and shown in Figure 4. The pandapower [16] Python framework is employed to model the grid and perform power flow calculations.

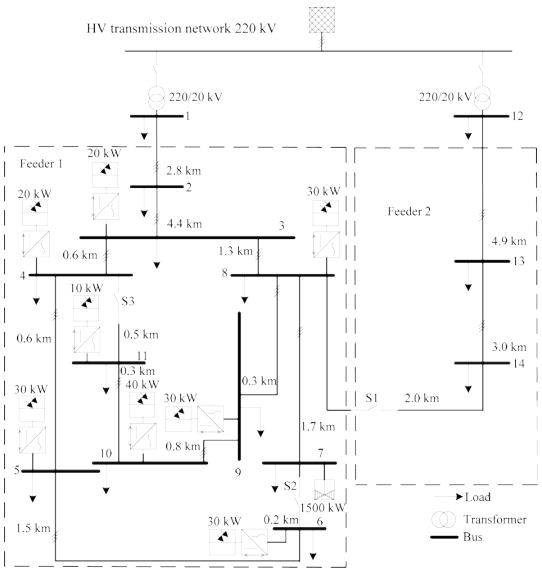

Fig. 4 CIGRE benchmark medium voltage grid with RES [15]

The distribution grid operates at 20 kV, comprising 18 loads and 9 inverter-coupled RES. Of these plants, eight are photovoltaic (PV) systems, while one is a wind power plant. The loads and generators are modeled with time series data taken from the SimBench data-set [17]. The data-set provides load and generation time series for the whole year of 2016, with a resolution of 15-minute intervals. The selected time series and additional details about the loads are listed in Table 1, while the RES attributes and time series are listed in Table 2.

Table 1 Used load time series from the SimBench data-set [17]

ID	Name	P [MW]	Q [Mvar]	Load profile
0	R1	14.99	3.04	-
1	R3	0.28	0.07	mv_semiurb
2	R4	0.43	0.11	mv_urban
3	R5	0.73	0.18	mv_comm
4	R6	0.55	0.14	lv_rural3
5	R8	0.59	0.15	lv_urban6
6	R10	0.48	0.12	mv_semiurb
7	R11	0.33	0.08	mv_urban
8	R12	14.99	3.04	-
9	R14	0.21	0.05	mv_comm
10	CI1	4.85	1.59	-
11	CI3	0.23	0.14	lv_rural1
12	CI7	0.08	0.05	lv_rural2
13	CI9	0.57	0.36	lv_semiurb4
14	CI10	0.07	0.04	lv_semiurb5
15	CI12	5.02	1.65	-
16	CI13	0.03	0.02	mv_rural
17	CI14	0.33	0.21	lv_semiurb4

In addition to the designation of the loads, the table shows the active (P) and reactive (Q) load available according to the pandapower implementation of the CIGRE benchmark grid.

In the listed loads, names with "R" represent residential loads, while "C" indicates industrial loads. Load profiles are chosen based on each load's apparent power, with R3, the mv_comm load profile aligns most closely with the standard load when compared to other profiles. The grid's four largest loads, R1, R12, C1, and C12, represent sub-grid. For simplicity in the simulation, their cross-interface power flows are omitted, and no power profiles are assigned.

Table 2 Used RES time series from the SimBench data-set [17]

ID	RES name	RES profile
0	PV 3	PV8
1	PV 4	PV2
2	PV 5	PV5
3	PV 6	PV1
4	PV 8	PV6
5	PV 9	PV3
6	PV 10	PV4
7	PV 11	PV7
8	WKA 7	WP8

Table 2 shows the name of the RES in the CIGRE benchmark grid in the second column and the corresponding RES time series used from the SimBench data-set in the third column. The PV plants are assigned to the time series randomly.

The RES facilities are expanded by a factor of 1.5 to enable the RES to supply approximately 48% of the necessary energy in the grid throughout the simulation year. This percentage corresponds roughly to the present share of feed-in from RES in Germany in 2022 [18].

In order to obtain a summary of the diverse grid load scenarios for every time step from the specified load and generation parameters, the total load and generation capacities can be calculated for each time step. This results in a total of 35,136 grid load scenarios, as shown in Figure 5 below.

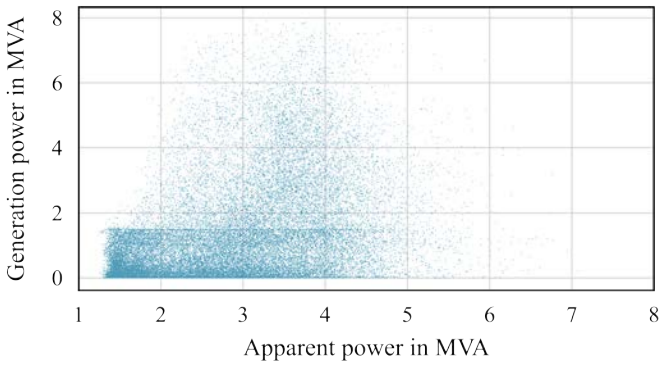

Fig. 5 Representation of the occurring grid operation scenarios

A point in the figure represents a grid operating scenario, consisting of the sum of all load outputs and the sum of all generation outputs. As the scatter of the scenarios shows, there are different operating scenarios. It can be seen that there are both peak and off-peak load cases with increased RES penetration. Within the scatter-plot, one line in particular can be seen at the 1.5 MW generation capacity across different load cases. This is the generation capacity of the wind turbine that reaches its maximum contribution at several times. The PV generation capacity does not show a clear maximum due to the larger distribution and smaller scaled generation capacity.

The time series data from the data-set are divided into a training data-set and an evaluation data-set. In order to have representative data for the whole year in both data-sets, the time series from the odd months are used for training and the even months are used for evaluation. During the training procedure the net is always evaluated over all six evaluation months after training with two randomly selected training months. This means that the training procedure in 2.1 is run consecutively for every time step in the two training months first. Afterwards the trained Actor-ANN is evaluated over all time steps in all six evaluation months.

2.4 Training and Model Evaluation

The objective of the SAC-Agent is to reduce the voltage deviation of all grid buses, the distribution grid comprises a total

of B = 15 buses. To attain this objective, the average absolute deviation from the nominal voltage $\overline{|\Delta U_t|}$ is calculated for each bus b at every time step t. The first bus, which has a constant voltage of 1.03 pu and is connected to the external grid, is excluded from the calculation of $\overline{|\Delta U_t|}$ because its voltage is constant in the grid simulation. The absolute value of the voltage difference, $|\Delta U_t|$, is calculated according to Equation 1.

$$|\Delta U_t| = \frac{\sum_{b=1}^{B-1} |1.0 - U_b|}{B} \tag{1}$$

The reward function is based on $\overline{|\Delta U_t|}$ with a slight modification. As the reward function aims to be maximised while minimising the voltage deviation, a negative coefficient is used to multiply $\overline{|\Delta U_t|}$. Since $\overline{|\Delta U_t|}$ is a very small deviation, the reward function given in Equation 2 is multiplied by a factor of -100.

$$R_t = -100 * |\Delta U_t| = -100 * \frac{\sum_{b=1}^{B-1} |1.0 - U_b|}{B} \tag{2}$$

To evaluate the training performance, the average of $\overline{|\Delta U_t|}$ over all evaluation steps $\overline{|\Delta U_t|}_{total}$ is calculated. If this average is less than the average of all previous evaluation episodes, a new best actor network has been trained and the actor net is exported. The exported net can then be used for further evaluations.

$$\overline{|\Delta U_t|}_{total} = \frac{\sum_{n=1}^{N} \overline{|\Delta U_t|}}{N} \tag{3}$$

2.5 Volt-Var Control Approaches

With this training method, two control approaches are developed. First a centralised system with one actor network which receives all bus voltages and last phase shift angles as input. The centralised approach runs a single instance of the training and evaluation procedure, with its objects as described in 2.1. In the second approach, each RES plant is controlled by an independent actor net. The actor nets receives only the local voltage and the last local shift angle as input and calculate only a shift angle output for the local plant. As this approach does not require any global information exchange, it is decentralised. The decentralised approach uses a separate instance of the training and evaluation process for each RES. In order to run the distribution grid power flow simulation only once per time step, the actors work with a simulation interface that only retrieves data from a simulation environment running the distribution grid simulation. The data flow during training is shown in Figure 6.

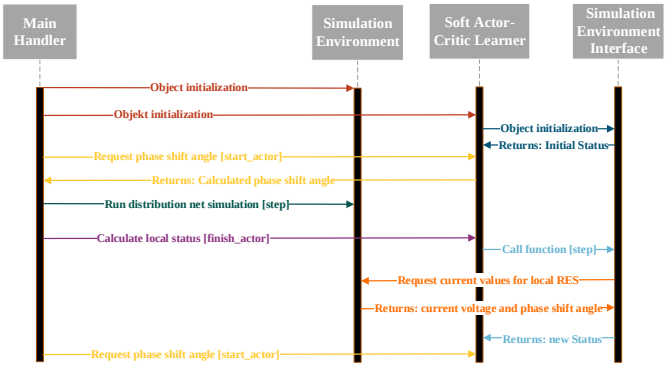

Fig. 6. Distributed actor training

3 Training and Evaluation Results

This section provides an in-depth analysis of the training and evaluation of the actor ANN for volt-var control. It starts by analysing the learning curve of the ANN throughout the training phase and then proceeds to a comparative analysis with existing volt-var control methods.

3.1 Training Evaluation

First, the training is evaluated below by analysing the learning process of the actor ANN. This analysis provides insight into whether the actor net is improving and effectively learning the intended function. This is done by plotting the total average absolute voltage deviation $\overline{|\Delta U_t|}_{total}$ for all evaluation episodes in Figure 7.

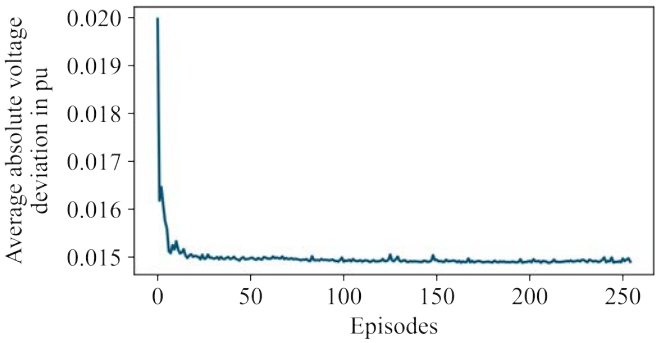

Fig. 7. Development of $\overline{|\Delta U_t|}_{total}$ during the training

The plot shows that the actor ANN converges very fast towards an optimum and afterwards only improves very slowly. This implies a intended and fast learning process that maximises the given reward function over the evaluation episode.

3.2 Post Training Evaluation

As stated earlier, this study compares the results obtained from the centralised and decentralised ANN approaches with the current volt-var control methods used in Germany, which are specified in [19]. Since there are various control methods and

parameterisations available, the following four control methods were chosen for analysis:

1. Static phase shift angle of $0°$
2. Static phase shift angle of $-18°$
3. Q(V) characteristic
4. Extended Q(V) characteristic with limited step size

The first two methods are based on a basic active power supply and a static capacitive reactive power supply. The Q(V) characteristic, derived from the definition outlined in [19], follows the characteristic presented in Figure8. The Q(V) characteristic is derived from the specification defined in [19]. In the context of grid planning, the Q(V) characteristic is parameterised according to the local requirements; in this work, the curve from the TCR corresponding to Figure 8 is used for simplification. The TCR [19] also requires a PT-1 controller with the Q(V) characteristic, which cannot be applied in a simulation with a 15 minute step width. For this reason we compare the developed ANN with a directly applied Q(V) characteristic and an extended Q(V) characteristic with a limited action step. For the second characteristic, the output shift angle is further limited to a maximum change of $0.3°$, resulting in a desirable average $\overline{|\Delta U_t|}total$. Additionally the maximum phase shift angle of the extended Q(V) characteristic is raised from the standard Q(V) value of 18.26° to 25.84°, which is the same as that used by the volt-var control techniques of ANN.

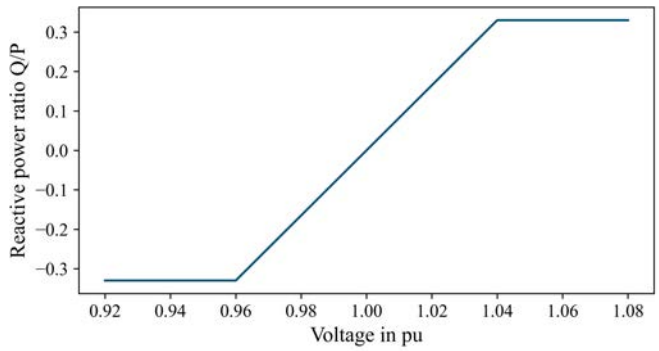

Fig. 8. Used Q(V) characteristic from the TCR 4410 [19]

The comparison of the different vol-var controls is evaluated over the entire SimBench data-set with 35,136 scenarios from the training and evaluation data-set. Four parameters, outlined in Table 3, are analysed for evaluation. The first parameter evaluates the average voltage deviation from the nominal voltage. The second and third parameter show the over- and under-voltage violations using an applied voltage band with a lower voltage of 0.95 pu and an upper voltage of 1.05 pu. The final parameter quantifies the cases when line currents exceed the rated current of the cable.

Table 3 Evaluation parameters and their definition

Evaluation Parameter	Definition
Average overall voltage deviation	$\overline{\lvert\Delta U_t\rvert}_{total}$ defined in (3)
Number of over-voltages $\#U_{ov}$	Sum of line voltages $U_t > 1.05$
Number of under-voltages $\#U_{uv}$	Sum of line voltages $U_t < 0.95$
Number of line overloads $\#l_{ol}$	Sum of line loads $l_t > 100\%$

3.3 Volt-var control comparison

With the two ANN approaches and the four control methods, following six different volt-var control methods are compared to each other.

1. Centralised ANN
2. Decentralised ANN
3. Static $\varphi = 0°$
4. Static $\varphi = -18°$
5. Plain Q(V) characteristic
6. Q(V) characteristic with limited step size

Based on the specified evaluation parameters, the following Table 4 shows the results of the simulation.

Table 4 Comparison of different volt-var controls

Control Approach	$\overline{\lvert\Delta U_t\rvert}_{total}$	$\#U_{ov}$	$\#U_{uv}$	$\#l_{ol}$
1	1,489	0	335	656
2	1,497	0	354	538
3	2,001	3348	457	126
4	1,636	0	790	605
5	1,779	172	386	357
6	1,650	0	644	637

The results in Table 4 demonstrate that the developed ANN approaches for volt-var control have the potential to decrease the total average absolute voltage deviation in the distribution gird. At the same time they minimise occurrences of voltage band violations. The minimal difference between the centralised and decentralised approaches indicates that global information about the distribution grid offers limited benefits for volt-var control. However the better voltage behaviour of the ANN approaches comes at the cost of an increased reactive power flow which leads to more line overloads. These findings so far only apply to the simulated distribution grid and with the applied simulation scenarios.

In order to analyse the behaviour of Volt-var control, the voltage in three operating scenarios is studied on selected buses. Figure 9 shows the bus voltages on a balanced RES generation and grid load day (base scenario). Figure 10 shows a low RES generation day (low scenario) and Figure 11 shows a high RES generation day (high scenario).

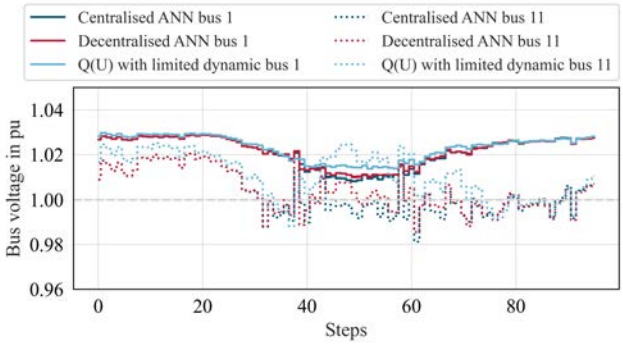

Fig. 9. Bus voltages base scenario

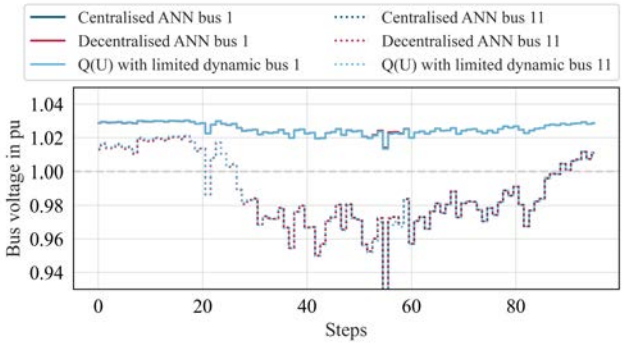

Fig. 10. Bus voltages low scenario

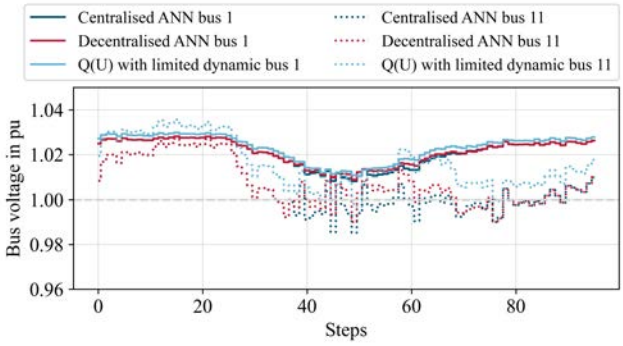

Fig. 11. Bus voltages high scenario

The base and high scenarios show that the voltages of the ANN volt-var controls are closer to the nominal voltage than those of the other methods. The low scenario shows that the voltage differences between the different voltage control methods become minimal. This shows, that the DRL volt-var control approaches have the potential to improve the voltage stability if the RES generate a substantial amount of the energy within in a distribution grid. To further evaluate the behaviour of the volt-var control methods, the generated output values of the phase shift angles can be accumulated over the entire simulation year. These distributions, shown in Figure 12 and Figure 13 give an overview of the differences in the volt-var control strategy.

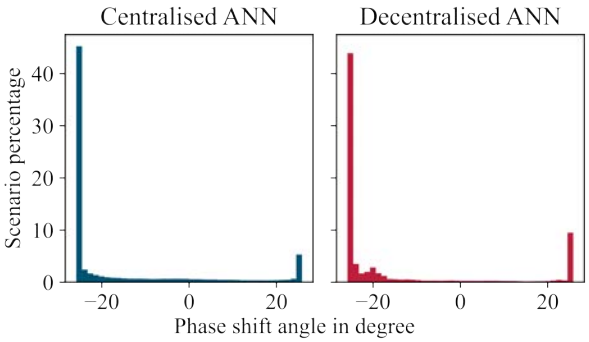

Fig. 12 Distribution of phase shift angles used by centralised and decentralised volt-var control

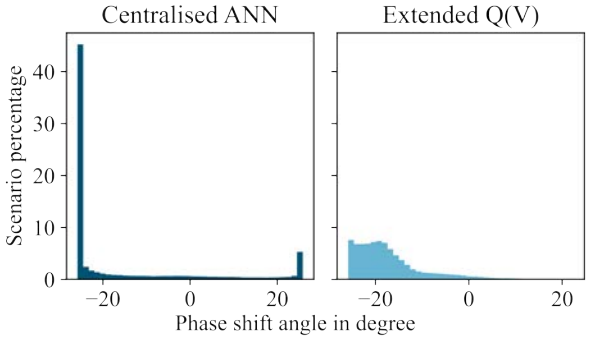

Fig. 13 Distribution of phase shift angles used by centralised volt-var control and extended Q(V) characteristic with limited step size

The distribution shows that the ANN volt-var control methods have a tendency to choose larger absolute phase shift angles more often than the Q(V) control. The strategy differences between the centralised and the decentralised approach are very minimal as seen in 12 which lead to the similar results already seen in table 4.

4 Conclusion

This paper investigates the potential application of Deep Reinforcement Learning (DRL) in the field of volt-var control in distribution networks. Conventional methods are effective but have limitations, e.g. in continuously adapting to changing operating scenarios. The introduction of a DRL-based approach aims to fill this gap and explore a more adaptive and efficient solution.

To this end, this paper presents an open-source-based framework that demonstrates the application of a centralised and decentralised approach to volt-var control within a CIGRE benchmark medium-voltage grid with a DRL agent. Furthermore, the results are presented within a comprehensive evaluation with conventional volt-var control methods.

The results show that ANN trained within a DRL to maintain voltage have the potential to improve voltage stability within the distribution grid. This has already been shown in smaller grids in [4], [3] and [20]. The present study extends the scope of investigation to a medium-voltage grid with multiple RES and places the results in the context of conventional voltage maintenance methods. In addition, the simulation period is extended to a full year at 15-minute intervals.

A more detailed analysis of the phase shift angles used shows that a more frequent use of larger phase shift angles, i.e. a larger reactive power injection, improves the voltage stability, but at the same time causes a higher load on the grid. This is mainly due to the fact that the ANNs are only trained on voltage stability. Furthermore, when comparing the centralised and decentralised ANN volt-var control approaches, only marginal differences can be identified between the two approaches, mainly due to the local effect of reactive power.

Future work will build on the framework developed and take into account other grid variables, such as asset load within the DRL. In addition, the development approach will be investigated in further and real grid models with real measurement data in order to simulatively map the technical and economic advantages and disadvantages in the context of existing volt-var control methods.

5 Acknowledgements

We acknowledge the support of our work by the German Ministry for Economic Affairs and Climate Action and the Project Management Jülich within the project Q-REAL - Reactive power management in real application: Network planning, network operation and reactive power sources for transmission and distribution grids (FKZ 03EI4063B). Only the authors are responsible for the content of this publication. Supplementary source code and data related to this article can be found online at `https://github.com/jtscs/DRL_volt-var_control`.

6 References

[1] Merten Schuster et al. “Design of a Data Driven Reactive Power Forecasting for an Active Cross-Voltage Level Reactive Power Management”. In: *ETG Congress 2023*. 2023, pp. 1–6.

[2] Ole W. Marggraf, Bernd Engel, and Rolf Witzmann. *Auslegung und Bewertung autonomer Spannungsregelkonzepte für Verteilungsnetze*. ger. 1. Auflage. München: Verlag Dr. Hut, 2020.

[3] Changfu Li, Chenrui Jin, and Ratnesh Sharma. “Coordination of PV Smart Inverters Using Deep Reinforcement Learning for Grid Voltage Regulation”. In: *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*. Boca Raton, FL, USA: IEEE, 2019, pp. 1930–1937.

[4] Kirstin Beyer et al. “Adaptive Online-Learning Volt-Var Control for Smart Inverters Using Deep Reinforcement Learning”. *Energies* 14.7 (2021), p. 1991.

[5] Hongbin Sun et al. “Review of Challenges and Research Opportunities for Voltage Control in Smart Grids”. *IEEE Transactions on Power Systems* 34.4 (2019), pp. 2790–2801.

[6] Priyank Srivastava et al. “Voltage regulation in distribution grids: A survey”. *Annual Reviews in Control* 55 (2023), pp. 165–181.

[7] Fco. Javier Zarco-Soto, José L. Martínez-Ramos, and Pedro J. Zarco-Periñán. “Voltage control in active distribution networks”. en. In: *Power Quality in Modern Power Systems*. Elsevier, 2021, pp. 193–217.

[8] Lefeng Cheng and Tao Yu. “A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems”. *International Journal of Energy Research* 43.6 (2019), pp. 1928–1973.

[9] M. Lapan. *Deep Reinforcement Learning Hands-On: Apply modern RL methods to practical problems of chatbots, robotics, discrete optimization, web automation, and more*. Packt Publishing, 2020.

[10] J.G. Vlachogiannis and N.D. Hatziargyriou. “Reinforcement Learning for Reactive Power Control”. *IEEE Transactions on Power Systems* 19.3 (2004), pp. 1317–1325.

[11] Christian Utama et al. “Reactive power control in photovoltaic systems through (explainable) artificial intelligence”. *Applied Energy* 328 (2022), p. 120004.

[12] Tuomas Haarnoja et al. *Soft Actor-Critic Algorithms and Applications*. arXiv:1812.05905 [cs, stat]. 2019.

[13] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems*. arXiv:1603.04467 [cs]. 2016.

[14] H. V. Hasselt. “Double Q-learning”. *In Advances in Neural Information Processing Systems (NIPS),pp. 2613–2621* (2010).

[15] Conseil international des grands réseaux électriques, ed. *Benchmark systems for network integration of renewable and distributed energy resources*. eng. Paris: CIGRÉ, 2014.

[16] Leon Thurner et al. “Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems”. *IEEE Transactions on Power Systems* 33.6 (2018), pp. 6510–6521.

[17] Steffen Meinecke et al. “SimBench—A Benchmark Dataset of Electric Power Systems to Compare Innovative Solutions Based on Power Flow Analysis”. *Energies* 13.12 (2020), p. 3290.

[18] Bundesnetzagentur. *Bundesnetzagentur veröffentlicht Daten zum Strommarkt 2022*. 28, 2023.

[19] *VDE-AR-N 4110 Anwendungsregel: 2018-11: Technische Regeln für den Anschluss von Kundenanlagen an das Mittelspannungsnetz und deren Betrieb (TAR Mittelspannung)*. Tech. rep. Volume: 29.240. Berlin: VDE-Verlag, 2018.

[20] Changfu Li et al. “Online PV Smart Inverter Coordination using Deep Deterministic Policy Gradient”. *Electric Power Systems Research* 209 (2022), p. 107988.