electronics

MDPI

Article

Volt-VAR Control in Active Distribution Networks Using Multi-Agent Reinforcement Learning

Shi Su [1], Haozhe Zhan [2,*], Luxi Zhang [2,3], Qingyang Xie [1], Ruiqi Si [2], Yuxin Dai [2], Tianlu Gao [2], Linhan Wu [2], Jun Zhang [2] and Lei Shang [2]

[1] Electric Power Research Institute of Yunnan Power Grid Co., Ltd., Kunming 650217, China
[2] The School of Electrical Engineering and Automation, Wuhan University, Wuhan 430072, China; luxizhang@brandeis.edu (L.Z.); ruiqi.si@whu.edu.cn (R.S.); yuxindai@whu.edu.cn (Y.D.); tianlu.gao@whu.edu.cn (T.G.)
[3] Physics Department, Brandeis University, Waltham, MA 02453, USA
* Correspondence: hz_zhan@whu.edu.cn

Abstract: With the advancement of power systems, the integration of a substantial portion of renewable energy often leads to frequent voltage surges and increased fluctuations in distribution networks (DNs), significantly affecting the safety of DNs. Active distribution networks (ADNs) can address voltage issues arising from a high proportion of renewable energy by regulating distributed controllable resources. However, the conventional mathematical optimization-based approach to voltage reactive power control has certain limitations. It heavily depends on precise DN parameters, and its online implementation requires iterative solutions, resulting in prolonged computation time. In this study, we propose a Volt-VAR control (VVC) framework in ADNs based on multi-agent reinforcement learning (MARL). To simplify the control of photovoltaic (PV) inverters, the ADNs are initially divided into several distributed autonomous sub-networks based on the electrical distance of reactive voltage sensitivity. Subsequently, the Multi-Agent Soft Actor-Critic (MASAC) algorithm is employed to address the partitioned cooperative voltage control problem. During online deployment, the agents execute distributed cooperative control based on local observations. Comparative tests involving various methods are conducted on IEEE 33-bus and IEEE 141-bus medium-voltage DNs. The results demonstrate the effectiveness and versatility of this method in managing voltage fluctuations and mitigating reactive power loss.

Keywords: active distribution network; Volt-VAR control; network partitioning; soft actor-critic; multi-agent reinforcement learning

Citation: Su, S.; Zhan, H.; Zhang, L.; Xie, Q.; Si, R.; Dai, Y.; Gao, T.; Wu, L.; Zhang, J.; Shang, L. Volt-VAR Control in Active Distribution Networks Using Multi-Agent Reinforcement Learning. *Electronics* **2024**, *13*, 1971. https://doi.org/10.3390/electronics13101971

Academic Editor: Andreas Mauthe

Received: 5 April 2024
Revised: 7 May 2024
Accepted: 13 May 2024
Published: 17 May 2024

1. Introduction

With the widespread integration of a high proportion of new energy sources into active distribution networks (ADNs), issues such as voltage violation, voltage fluctuation, and power loss caused by distributed new energy sources such, as photovoltaic (PV), are becoming increasingly prominent [1]. In order to alleviate these problems, Volt-VAR Control (VVC) is widely applied to improve the voltage quality of distribution networks (DNs) and reduce network losses [2].

The objective of VVC is to ensure the robust operation of DNs through the control of voltage and reactive power [3]. In conventional DNs, methods such as distribution network reconfiguration [4], shunt capacitor banks (SCBs) [5], and on-load tap changers (OLTCs) [6] are commonly employed for voltage regulation. However, these methods, which rely on mechanical devices, encounter challenges such as slow response speeds and limited accuracy. These limitations hinder the execution of rapid and flexible control strategies, falling short of the requirements for quick and precise regulation in ADNs. An emerging trend involves the gradual integration of PV inverters into the voltage regulation of DNs.

This approach achieves voltage regulation by adjusting the reactive power injection of the inverter and is progressively gaining traction and widespread application.

Based on the limitations posed by DNs in terms of computing and communication capabilities, the control of ADNs can be categorized into the following three types: local control, centralized control, and distributed control. The local control method operates relatively independently and does not rely on complex communication equipment. Each distributed generation device can make decisions based on local information, ensuring a fast response and autonomy [7]. Centralized control methods consolidate control functions within a central controller, enabling unified management and regulation of grid voltage [8]. Consequently, centralized control strategies require dependable and rapid communication devices as a fundamental prerequisite to fulfill real-time data acquisition and transmission necessities. The distributed control approach is specifically designed for regions characterized by unreliable communication links and limited computational resources at the central station. It accomplishes distributed management and regulation of voltage by exchanging boundary information between autonomous units. The alternating direction method of multipliers (ADMM) is the most widely used distributed algorithm in distribution networks. ADMM is used to solve the stochastic and distributed optimal energy management problem for ADNs within office buildings [9]. This is achieved through coordination and optimization in a hierarchical and zoned manner, accounting for control speed and global coordination capability [10,11]. This method only requires a limited number of communication links to obtain local observation information for problem-solving, allowing it to adapt to the complexity and dynamics of large-scale DNs.

The VVC problem within DNs can be formulated as a mixed-integer nonlinear programming problem. Heuristic algorithms, valued for their ease of programming implementation, find widespread use in solving optimization problems in DNs. In the VVC problem of DNs, commonly used intelligent optimization algorithms include Genetic Algorithm [12], Particle Swarm Optimization [13], etc. However, heuristic algorithms are prone to converging towards local optima and have exponential time complexity in solving problems, making them inadequate for handling large-scale VVC problems. Consequently, there are practical limitations associated with their widespread application.

Transforming the non-convex problem into a convex one allows for the utilization of convex optimization techniques, thereby achieving a globally optimal solution efficiently. In [14], a method based on mixed-integer quadratic programming is proposed for VVC in DNs. This innovative approach transforms the initially non-convex problem into a convex optimization problem by exact relaxation of the non-convex constraints. While traditional mathematical methods offer high solution accuracy, they struggle to adapt to VVC scenarios with stringent real-time control requirements. Hence, there is a need to investigate methods for developing millisecond-level control strategies.

With the advancement of Artificial Intelligence (AI), AI algorithms play a pivotal role in optimization strategies for DN control. AI algorithms facilitate real-time online optimization by transferring computational load from online to offline processes. Leveraging the reinforcement learning (RL) approach, real-time responsiveness and adaptability to the system are facilitated through the interactive learning process between the agent and the environment. For instance, the Deep Deterministic Policy Gradient (DDPG) algorithm, introduced in [15], was devised to mitigate voltage violations arising from uncertainty in power systems. However, most RL methods typically employ a single agent within a virtual environment, rendering them unsuitable for direct application to larger-scale systems. Consequently, RL methods based on multi-agent systems have been proposed. In [16], a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) algorithm was introduced for two-stage VVC of ADNs. Nonetheless, its evaluation was confined to an IEEE 33-bus DN, limiting the demonstration of its scalability.

Based on the above analysis, this paper proposes a Volt-VAR control strategy based on MARL in ADNs. The contributions of this paper are outlined as follows:

1. A method based on electrical distance is proposed for partitioning DNs. The partitioned DN exhibits highly aggregated characteristics within regions and low coupling between regions, laying the foundation for achieving distributed VVC of PV inverters.
2. This paper proposes a framework for VVC in DNs based on the Multi-Agent Soft Actor-Critic (MASAC) algorithm. The framework employs a strategy of centralized training followed by distributed execution, aiming to reduce communication and computation demands on the DNs during execution. This approach alleviates the resource-intensive nature of centralized control strategies in terms of real-time computation and storage requirements. The established framework enables the coordinated control of PV systems to minimize voltage deviations while simultaneously minimizing reactive power losses in the DN. Importantly, this coordination occurs with agents interacting only with local information from sub-regions of the DN.
3. In order to validate the effectiveness and versatility of the proposed framework, experiments were conducted using five MARL algorithms, including MASAC, on IEEE 33-bus and IEEE 141-bus network. The results demonstrate that the proposed method can effectively achieve VVC in DNs, relying solely on local observation information after training.

2. Proposed VVC Strategy

2.1. The Model of Volt-VAR Control

Consider an ADN consisting of $N+1$ individual nodes with a high penetration of PV integration. The DN is modeled as graph $\varsigma = (V, E)$, where $V = \{V_0, V_1, \ldots, V_N\}$ and $E = \{E_1, E_2, \ldots, E_N\}$ is the set of nodes and edges. For each bus $i \in V$ in the network, let $s_i = p_i + jq_i$ be the complex power injection and $\widetilde{v}_i = v_i \angle \theta_i$ be the complex voltage. Power flow constraints at bus i are as follows:

$$p_{i,PV} - p_{i,L} = |v_i| \sum_{j=1}^{N} |v_j| \left(G_{ij} \cos \theta_{ij} + B_{ij} \sin \theta_{ij} \right) \tag{1}$$

$$q_{i,PV} - q_{i,L} = |v_i| \sum_{j=1}^{N} |v_j| \left(G_{ij} \sin \theta_{ij} - B_{ij} \cos \theta_{ij} \right) \tag{2}$$

where $p_{i,PV}$ and $q_{i,PV}$ are the active and reactive power injection of PV, $p_{i,L}$ and $q_{i,L}$ are the complex power of the load connected to node i, and G_{ij} and B_{ij} denote the real and imaginary parts of the inter-node conductance for node i and node j, respectively.

In ensuring the safe and optimal operation of the DN, it is imperative to consider constraints on voltage at individual nodes to alleviate voltage deviations.

$$v_{i,\min} \leq v_i \leq v_{i,\max} \tag{3}$$

where $v_{i,min}$ and $v_{i,max}$ represent the lower and upper bounds. The upper voltage bound is $v_{i,max} = 1.05$ p.u., and the lower voltage bound is $v_{i,min} = 0.95$ p.u.

In addressing the voltage instability resulting from the integration of renewable energy sources, PV inverters serve as controllable instruments. In the presence of a PV inverter at a node, limitations are imposed on the reactive power output of the PV inverter as follows:

$$\overline{s}_{i,PV} \geq \sqrt{(p_{i,PV})^2 + (q_{i,PV})^2} \tag{4}$$

where $\overline{s}_{i,PV}$ denotes the rated capacity of the PV inverter. To guarantee that the rated capacity of the PV generating units utilizing inverters is met, the inverters retain the capacity to provide reactive power support. Generally, rated capacity is established at 1.0 to 1.1 times the PV generating units' active power capacity.

2.2. Network Partition Based on Electrical Distance

An ADN can be subdivided into multiple sub-regions, enabling the resolution of the VVC problem in a distributed fashion. This approach transforms a large-scale DN optimization problem into multiple smaller optimization problems within various sub-regions. This not only reduces the demands of the control strategy on communication equipment and computational resource but also enhances the synergy and stability of the control.

This study employs electrical distance based on voltage sensitivity to cluster nodes sharing similar characteristics. Electrical distance can be derived from the Newton–Raphson method-based power flow calculation equation. The modified expression for the power flow calculation can be articulated as:

$$\begin{bmatrix} \Delta\delta \\ \Delta U \end{bmatrix} = \begin{bmatrix} S_{P\delta} & S_{Q\delta} \\ S_{PU} & S_{QU} \end{bmatrix} \begin{bmatrix} \Delta P \\ \Delta Q \end{bmatrix} \tag{5}$$

where $S_{P\delta}$ and $S_{Q\delta}$ are the impact of unit active power and reactive power, respectively, on the phase angle alteration of the nodal voltage. Similarly, S_{PU} and S_{QU} denote the effect of unit active and reactive power, respectively, on the magnitude of the node voltage. Then, the voltage magnitude satisfies the following equation:

$$\Delta U = S_{PU}\Delta P + S_{QU}\Delta Q \tag{6}$$

The objective of this paper is to regulate voltage by controlling the reactive power of PV inverters. Consequently, only the influence of reactive power on voltage magnitude is considered. Reactive voltage sensitivity is chosen to characterize the electrical distance D_{ij} [17]:

$$D_{ij} = S_{QU}^{ii} + S_{QU}^{jj} - S_{QU}^{ij} - S_{QU}^{ji} \tag{7}$$

where D_{ij} denotes the degree of electrical coupling between the nodes. S_{QU}^{ij} represents the sensitivity of the voltage magnitude at node i to the injected reactive power at node j. The smaller the numerical value of the electrical distance, the tighter the electrical connection between two nodes. For nodes i and j, a closer electrical connection implies that their mutual reactive power-voltage sensitivity is more similar, resulting in a smaller electrical distance D_{ij}.

The network topology of the community network bears resemblance to the ADN structure. The modularity index not only serves to gauge the quality of connections within the network topology but also guides the direction of network partitioning through intelligent search algorithms, determining the optimal partitioning scheme. In this paper, the modularity index is chosen as the optimization objective for ADN partitioning.

$$Q = \frac{1}{2m}\sum_{i,j}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j) \tag{8}$$

$$A_{ij} = 1 - \frac{D_{ij}}{\max\limits_{m,n} D_{mn}} \tag{9}$$

$$\delta(c_i, c_j) = \sum\nolimits_g \delta(c_i, g)\delta(c_j, g) \tag{10}$$

where Q is the modularity index, A_{ij} elucidates the weights of connections between nodes, $m = \frac{1}{2}\sum_{i,j} A_{ij}$ is the total weight of node in the whole network, and $k_i = \sum_{j=1}^{n} A_{ij}$ is the sum of weights connected to node i. g denotes the cluster number, c_i denotes the cluster number where node i is located. When two nodes i, j belong to the same cluster g, $\delta(c_i, c_j) = 1$; otherwise, $\delta(c_i, c_j) = 0$.

The degree of electrical coupling is proportional to the value of the module degree. In ADNs, a higher modularity index indicates tighter electrical connections within the same

region and sparser electrical connections between different regions. When the modularity index is greater than 0.3, it indicates that the degree of electrical coupling is obvious.

In this study, we employ the unsupervised heuristic Louvain algorithm to explore the optimal clustering. The Louvain algorithm aims to optimize modularity, wherein the fundamental concept involves partitioning nodes into clusters that maximize the network's modularity. Within this algorithm, each node is considered an independent cluster initially, and subsequent merging of adjacent clusters continues until modularity ceases to increase.

2.3. Formulate VVC as a Markov Game

After partitioning the DN into multiple sub-regions during network partitioning, agents have access to decentralized observation data within these sub-regions. Aiming to manage voltage within a specified range by adjusting PV inverters, the VVC problem is commonly expressed as a Markov game (MG) [18]. The essential components of the MG settings are outlined as follows:

State space: S_t is the state set of the agent, and $s_t^i \in S_t$ is the local observation of agent i at moment t. In this study, s_t^i consists of five parts $(p_{i,L}, q_{i,L}, p_{i,PV}, q_{i,PV}, u_{i,v})$, where $q_{i,L}$ is the reactive power generated by PV inverters at the previous step and $u_{i,v}$ is the nodal voltage vector of the node to which the inverter is connected.

Action space: each agent has a set of continuous actions $a_{i,t} = \{-h \le a_{i,t} \le h, h > 0\}$. The reactive power produced by the first i PV inverter is $q_{i,PV} = a_{i,t}\sqrt{(\bar{s}_{i,PV})^2 - (p_{i,PV})^2}$. When $a_{i,t} > 0$, it signifies that the PV inverter penetrates reactive power to the DN; when $a_{i,t} < 0$, it indicates that the PV inverter absorbs reactive power from DN. h represents a hyperparameter that constrains the range of actions available to an agent, which is often chosen according to the load capacity of lines and transformers in the DN to ensure the safe operation of the PV inverter.

State transition: In this paper, the state space includes the PV agent's previous action, and the load demand depends on user behaviors. Therefore, let $s_{t+1}^i = P(s_t^i, a_t^i, w_t)$ describe the state transition function, depicting how the state changes after agents' actions, where w_t represents random environment noise used to simulate the randomness of PV and load demand fluctuations $(p_{i,L}, q_{i,L}, p_{i,PV})$.

Reward function: $r_t^i \in R_t$ symbolizes the reward granted to an agent upon executing an action. All agents subscribe to a unified form of the reward function, aligning with the optimization goal of sustaining voltage within a secure threshold around V_{ref} while also minimizing reactive power losses. Hence, the reward function can be articulated as:

$$r_{i,t} = -\frac{1}{|v|}\sum f_v(v_i) - \alpha \cdot f_q(\boldsymbol{q}_{PV}) \tag{11}$$

where $f_q(\boldsymbol{q}_{PV}) = \frac{1}{|A|}\|\boldsymbol{q}_{PV}\|_1$ is the average reactive power losses, and the control requires as little reactive power generation as possible, $f_q(\boldsymbol{q}^{PV}) < \varepsilon, \varepsilon > 0$.

One of the objectives of VVC is to maintain the voltage within the range of 0.95 p.u. to 1.05 p.u. as much as possible. When the voltage exceeds the safety range, it is desired that the MARL intelligent agents receive greater penalties, thereby guiding the agents to learn strategies that maximize the reward function. Hence, the voltage barrier function $f_v(v_k)$ is defined as:

$$f_v(v_k) = \begin{cases} a \cdot |v_k - v_{\text{ref}}| - b & if |v_k - v_{\text{ref}}| > 0.05 \\ -c \cdot \mathcal{N}(v_k - v_{\text{ref}}, 0.1) + d & otherwise \end{cases} \tag{12}$$

where $\mathcal{N}(v_k - v_{\text{ref}}, 0.1)$ represents a density function that follows a normal distribution with a mean of v_{ref} and a standard deviation of 0.1.

2.4. MASAC-Based VVC Framework

The MASAC algorithm is effective in tackling sequential decision-making challenges within Markov games. To achieve multi-cluster distributed cooperative control without relying on inter-cluster communication, the proposed network partition method divides

regions based on electrical distances, thus creating a partition pattern characterized by a high cohesion of nodes within regions and low coupling between regions. Within each subregion, every PV inverter is represented as an individual SAC agent, and ultimately, the MASAC algorithm is utilized to address the VVC problem. The MASAC-based Centralized Training and Decentralized Execution (CTDE) architecture is introduced, as illustrated in Figure 1. The framework employs centralized training to refine the policy of each agent, subsequently facilitating decentralized execution.

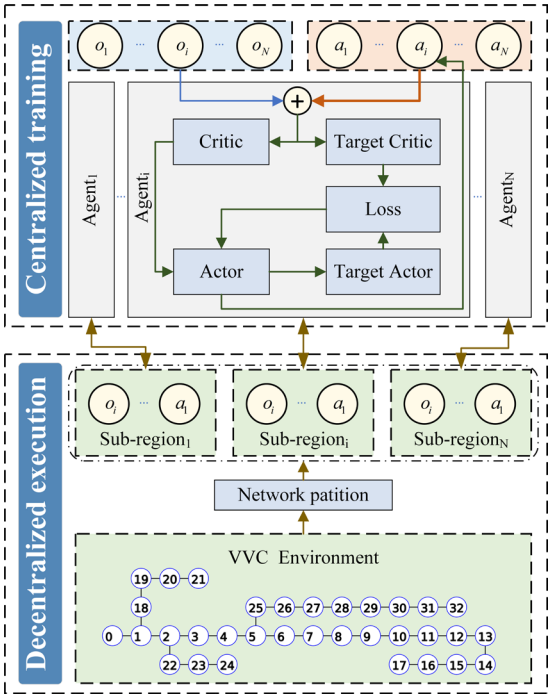

Figure 1. VVC framework based on MASAC.

In the centralized training phase, based on the network partition method, the critic network of each agent receives global state information from all agents, while the actor network corresponding to each agent receives local observations from its respective sub-region. The critic network, utilizing global information, supports the actor network in acquiring coordinated control strategies for multi-PV voltage systems.

During decentralized execution, control strategies are determined by the actor network, requiring the retention of only the actor network. As the actor network has already acquired cooperative control strategies utilizing global information during the training phase, it becomes feasible to deliver more cooperative and resilient control strategies solely relying on local information during execution.

The set of strategies of the MG can be expressed as $\pi = \{\pi_1, \ldots \pi_N\}$, where strategy π_i denotes the action function of agent i. Unlike traditional reinforcement learning, the optimization objective of the MASAC algorithm is to maximize the weighted sum of cumulative returns and entropy, $\sum_{t=0}^{T} E[r_t + \alpha H(\pi_i(\cdot|s_{i,t}))]$, where entropy is denoted as $H(\pi_i(\cdot|s_{i,t}))$; α is the temperature coefficient. Entropy is introduced to improve the efficiency of exploring the action space. To this end, MASAC provides an efficient and principled way for balancing exploration and exploitation [18].

The parameters of the strategy function undergo updates through the utilization of the strategy gradient.

$$\nabla_\mu J(\mu_t) = E_{S_t, A_t \sim D}\left[\nabla_\mu \log(\pi^{\mu_t}(a_{i,t} \mid s_{i,t}))\rho_t(S_t, a_{1,t}, \ldots, a_{N,t})\right] \tag{13}$$

$$\rho_t(S_t, a_{1,t}, \ldots, a_{N,t}) = +Q_i^\pi(S_t, a_{1,t}, \ldots, a_{N,t}) - b\left(S_t, a_{i',t}\right) - \alpha \log(\pi^{\mu_t}(a_{i,t} \mid s_{i,t})) \tag{14}$$

$$b(S_t, a_{i',t}) = E_{a_{i,t} \sim \pi^{\omega_i}(s_{i,t})}\left[Q_i^{\pi}(S_t, (a_{i,t}, a_{i',t}))\right] \tag{15}$$

where $Q_i^{\pi}(S_t, a_{1,t}, \ldots, a_{N,t})$ is the value of the action; b is the base term used to represent the average value of all possible actions of an agent body i in state S_t; and $Q_i^{\pi}(S_t, a_{1,t}, \ldots, a_{N,t}) - b(S_t, a_{i',t})$ is the dominance function representing the value of the current action, which provides an estimate of the relative value of the action for the agent.

The evaluation network $Q_i^{\pi}(\cdot)$ is responsible for calculating the value of an action for an agent. The parameters of the evaluation network are optimized by minimizing the following loss function:

$$L = (y_t - Q_i^{\pi}(S_t, a_{1,t}, \ldots, a_{N,t}))^2 \tag{16}$$

$$y_t = r_{i,t} + \gamma E_{a_{i,t+1} \sim \pi^{\mu_i'}}\left[-\alpha \log\left(\pi^{\mu_i'}(a_{i,t+1} \mid s_{i,t+1})\right) + Q_i'^{\pi}(S_t, a_{1,t}, \ldots, a_{N,t})\right] \tag{17}$$

where y_t is the target value and $\pi^{\mu_i'}$ and $Q_i'^{\pi}$ are the target action and critic function.

Deep neural networks typically demand the input data to be distributed independently and uniformly during training. However, in reinforcement learning, the interaction between agents and the environment often leads to non-independent and non-uniform data distributions, impacting neural network training stability. SAC introduces the experience replay buffer, where each agent interacts with the environment, storing experience data $(s_{j,t}, a_{j,t}, r_{j,t}, s_{j,t+1})$ acquired from these interactions in a playback pool. When updating network parameters, agents randomly draw experience data from this pool, compute gradients, and update corresponding network parameters. This mechanism disrupts data correlations, promoting a distribution closer to independence and uniformity, thereby reducing update variance and enhancing network convergence speed. Moreover, reusing empirical data facilitates more efficient data utilization, particularly in scenarios with limited access to new data.

3. Case Study

This paper conducts arithmetic simulation experiments to assess the performance of the VVC strategy based on the MASAC algorithm, utilizing the IEEE 33-bus [19] and IEEE 141-bus networks [20]. Power flow calculation was performed using PandaPower v2.13.1, and the training and testing of the proposed VVC strategy was performed in Python equipped with the Pytorch library. All experiments are implemented on a workstation equipped with an Intel 12th generation i7-12700H central processor (Intel, Santa Clara, CA, USA) and a NVIDIA GeForce3060 GPU (NVIDIA, Santa Clara, CA, USA).

3.1. The Performance of Network Partition

In this section, simulations are implemented on the IEEE standard network to evaluate the performance of the network partition method. The PV data is obtained from one year's operation of a local grid, and the PV parameters and integration information are shown in Table 1.

Table 1. Capacity and location of PV.

Networks	Capacity	Location
IEEE 33	0.5MW/0.51MVA	11, 17, 21, 24, 29, 32
IEEE 141	0.5MW/0.51MVA	35, 52, 58, 61, 67, 68, 74, 76, 81, 86, 99, 105, 109, 110, 115, 116, 129, 132, 136, 137, 138, 140

Figure 2 illustrates the arrangement of the IEEE 33-bus network topology alongside the outcomes of optimal partitioning. The convergence of the modularity index yields a result of 0.4851, delineating the DN into five distinct sub-regions characterized by strong internal cohesion and minimal inter-regional coupling. The IEEE 141-bus network is divided into nine sub-regions, and the optimal partitioning results are shown in Figure 3. The convergence result of the modularity index is 0.3623, and there are controllable regulation

devices in each region. The proposed partitioning algorithm converges quickly and can be extended to be used in large-scale DN models, which provides the basis for the cooperative VVC strategy.

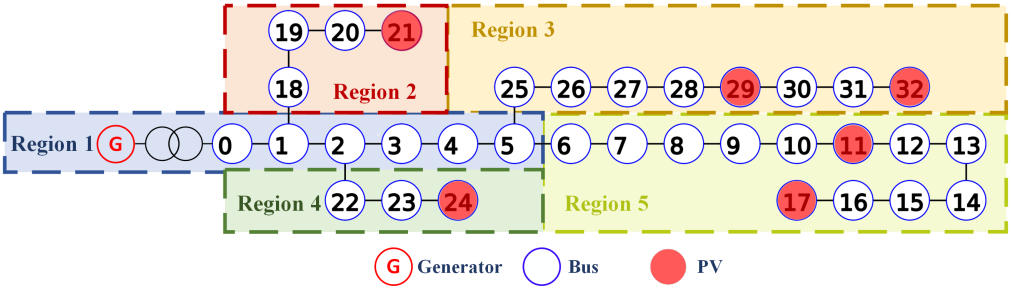

Figure 2. IEEE 33-bus network structure.

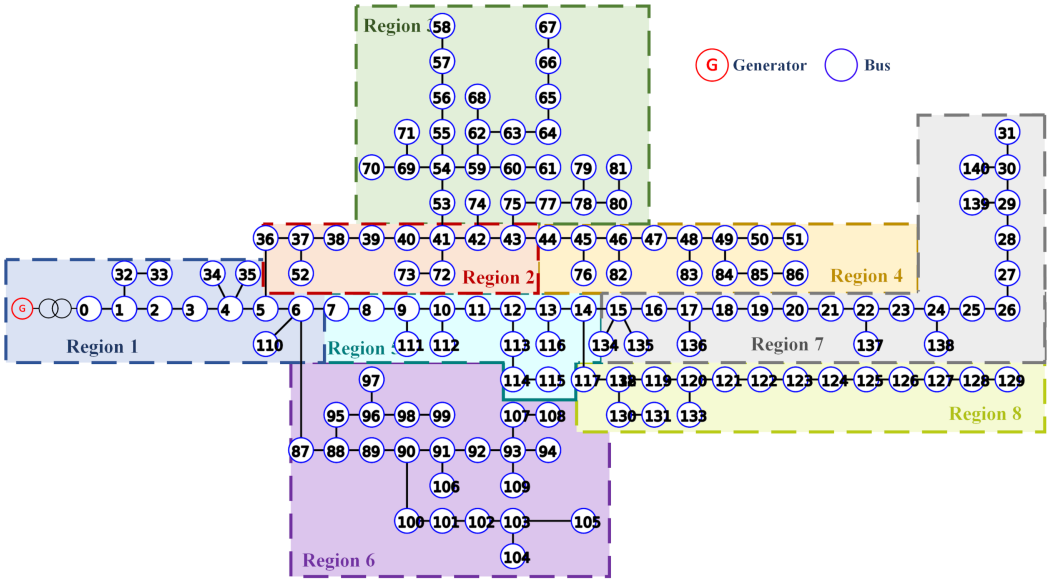

Figure 3. IEEE 141-bus network structure.

3.2. The Performance of Volt-VAR Control Based on MASAC

In this study, every PV inverter is portrayed as a reinforcement learning agent. The IEEE 33-bus system comprises 5 agents, while the IEEE 141-bus system encompasses 22 agents. Within the MASAC algorithm framework, each agent comprises two actor networks and three critic networks, with shared neural network parameters across all agents. GRU, a recurrent neural network, is utilized as an optimizer for the actor networks to tackle the partially observed problem. The critic networks are fashioned using a multilayer perceptron.

Throughout the training phase, we initiate each episode by randomly selecting a day of photovoltaic data as the starting state, with each episode spanning 240 time steps (equivalent to half a day). Testing is conducted every 20 episodes, whereby 10 episodes are randomly chosen for assessment. The training data batch size is configured at 32, while the non-strategy algorithm's buffer size is set to 5000.

The hyperparameter configurations of the MASAC algorithm are presented in Table 2. To assess the control efficacy of the proposed algorithm, comparison experiments involving various MARL algorithms are undertaken. These encompass the following:

1. The counterfactual multi-agent (COMA) method, which uses Q-value decomposition to achieve collaborative decision-making by optimizing local Q-values and influence factors;
2. The MADDPG algorithm, which uses a deep deterministic policy gradient to achieve collaborative decision-making for multi-agents;

3. The multi-agent proximal policy optimization (MAPPO) algorithm, which uses policy optimization and importance sampling to achieve collaborative decision-making for multi-agents;
4. Multi-agent twin delayed deep deterministic (MATD3) method, which uses double-delayed deep deterministic policy gradients to achieve collaborative decision-making for multi-agents;
5. The proposed MASAC method, where SAC algorithm-based agents are trained based on global observations, and each agent controls the inverter reactive output based on local observations within the network in which it is located.

Table 2. Simulation setup.

Parameter	Value
h	0.8
Batch	32
Experience replay buffer	5000
Policy network learning rate	0.001
Critic network learning rate	0.001
λ	0.99

In the experiment, three metrics are established to assess algorithm performance including the following:

1. Rewards: this metric calculates the value of discount rewards received by the agent after executing an action. The agent aims to maximize the discounted rewards, and a higher discounted reward indicates that in the current episode, the agent receives a higher value of the reward from the environment after executing the scheduling action and has a better overall performance in maximizing the trade-off between reducing the voltage excursion and reducing the reactive power loss.
2. Controllable ratio: this metric calculates the ratio of time steps during which all bus voltages are under control during each episode. A higher controllable ratio indicates a better performance of the algorithm in terms of bus control.
3. Reactive power loss: this metric calculates the average value of the reactive power loss of all lines for each time step during each event. A lower reactive power loss can indicate that the algorithm has better performance in reducing power loss.

3.2.1. Test on This IEEE 33-Bus Network

In this study, 400 episodes were trained offline utilizing the proposed method. During the training phase, the initial state for each episode is randomly selected from the dataset, and each episode comprises 240 steps. The size of the experience playback pool is set to 5000 sample sizes, before 5000 sample sizes, the agents' action decisions do not rely on the selection of the strategy function, which enables full exploration of the strategy space. Within this phase, the action state reward function is deposited into the experience playback pool for each agent and the parameters of the neural network remain fixed. After collecting 5000 sample numbers, the capacity of the experience playback pool reaches the upper limit, at which point the neural network parameters of the agents start to be updated. During the training process, the inverter gradually masters the reactive power control strategy. When the training comes to the late stage, the discount reward curve of the agent oscillates in a small range around the fixed value. At this time, the algorithm begins to converge, and the inverter masters the reactive power control strategy of avoiding voltage overruns and reducing reactive power losses. The results of the tests during the training period are given by interquartile shading from 25% to 75% after a sliding average process.

Figure 4 shows the rewards curve. From the analysis of Figure 4, the VVC strategy based on the proposed MASAC algorithm converges faster and has greater discounted rewards as compared with algorithms such as MADDPG, COMA, and MAPPO. This shows

that the MASAC algorithm has a better synthesis ability in learning VVC strategies for PV inverters.

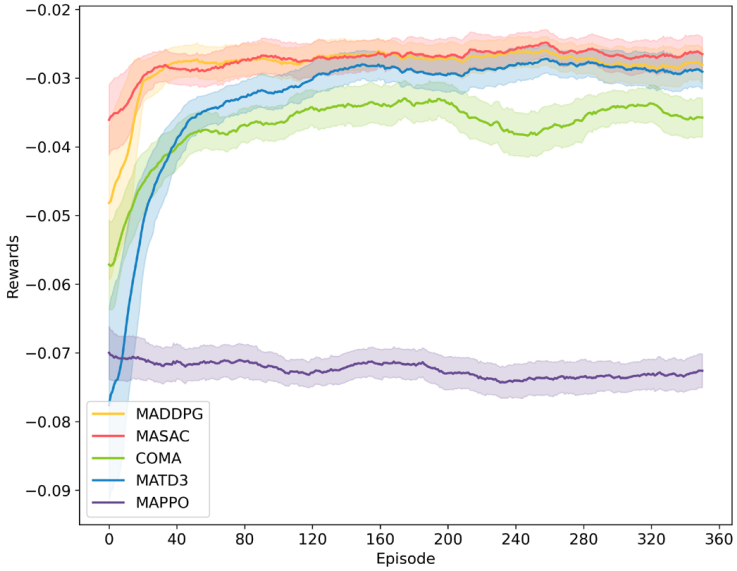

Figure 4. Rewards on the IEEE 33-bus network.

The reactive power loss indicator curve is shown in Figure 5. In this metric, the MASAC algorithm controls the reactive power loss value convergence value of 0.12 MVA. Compared with the other algorithms, the SAC agents learn a better strategy to reduce the reactive power loss.

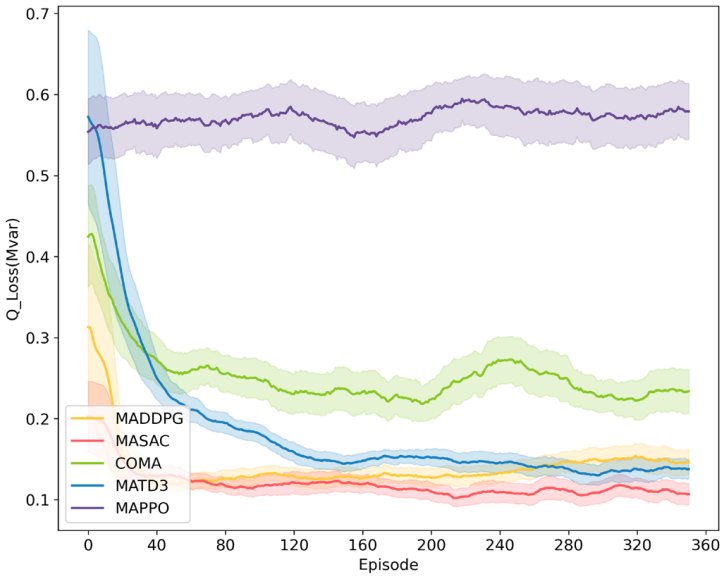

Figure 5. Reactive power loss on the IEEE 33-bus network.

The controllable ratio curve is shown in Figure 6. The MASAC algorithm has high values of final convergence of the controllable ratio and better busbar control capability as compared with the other algorithms. The training trend in the MASAC algorithm shows that it has a faster exploration ability in the initial phase of training.

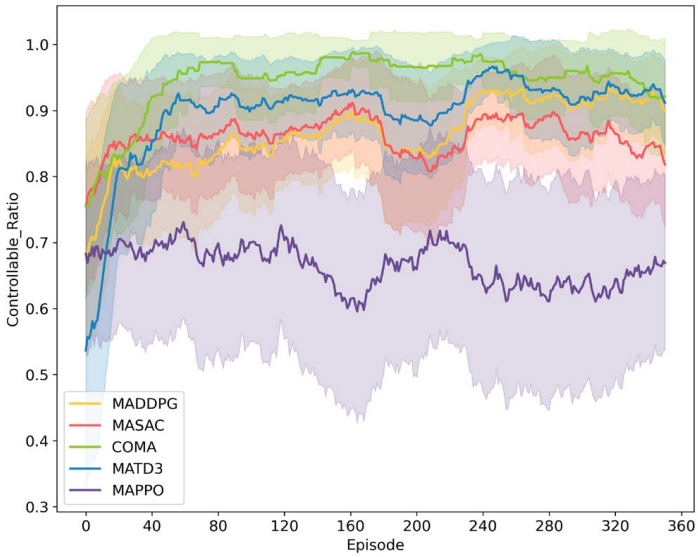

Figure 6. Controllable ratio on the IEEE 33-bus network.

The voltage control effects achieved by different control methods on the test set are shown in Table 3. It can be observed that all control methods are capable of regulating the voltage to a reasonable range. However, the proposed MASAC algorithm achieves an average voltage value of 0.9976 p.u. and a reactive power loss of 0.1233 MVAR, which indicates better voltage control and reactive power loss control compared with the other algorithms. Additionally, the maximum voltage rise is 0.002 p.u. and the maximum voltage drop is 0.001 p.u., both of which remain within the safety range. This demonstrates that the VVC framework based on MASAC effectively mitigates voltage fluctuations at DN nodes.

Table 3. Performance of control strategies on IEEE 33-bus network.

Strategy	Average Voltage	Maximum Voltage Rise	Minimum Voltage Drop	Reactive Power Losses	Controllable Ratio
MASAC	0.9976	0.002	0.001	0.1233	95.37%
MADDPG	0.9944	0.004	0.002	0.1706	96.60%
MATD3	0.9879	0.002	0.001	0.0899	96.03%
MAPPO	1.0096	0.010	0.005	0.6310	60.92%
COMA	0.9972	0.001	0.003	0.2223	98.85%

A comprehensive analysis of the indicators and test results shows that the VVC strategy based on the proposed MASAC algorithm has a strong comprehensive performance. The SAC agents learn fast and converge well. This indicates that the MASAC algorithm has a stronger strategy exploration ability, which verifies that the introduction of the concept of information entropy in the MASAC algorithm and updating the optimization objective of the algorithm have led to a significant improvement in the algorithm's learning ability and convergence effect. When the proposed algorithm training model is applied to the test set, the outcomes reveal that the agents acquire an enhanced cooperative VVC strategy, demonstrating proficient control over the voltage offset, voltage fluctuation, and reactive power loss.

Comparative experiments on the IEEE 33-bus network reveal that the MASAC algorithm outperforms the other MARL algorithms in achieving collaborative VVC of PV inverters under local information-based conditions. This ensures the stable and safe operation of DNs.

3.2.2. Test on the IEEE 141-Bus Network

In order to test the scalability of the proposed MASAC algorithm, this section performs a comparison test on the IEEE 141-bus network. The PV parameters and access information of this network are shown in Table 1, and the rest of the experimental settings and algorithm settings remain unchanged.

Figure 7 shows the rewards curve. From the analysis of Figure 7, the VVC framework based on MASAC for the experiments on the IEEE 141-bus network has more obvious advantages compared with the IEEE 33-bus network. Compared with the MATD3, MADDPG, COMA, and MAPPO algorithms, the proposed method converges faster, gives better convergence results, and has the largest discount reward for convergence.

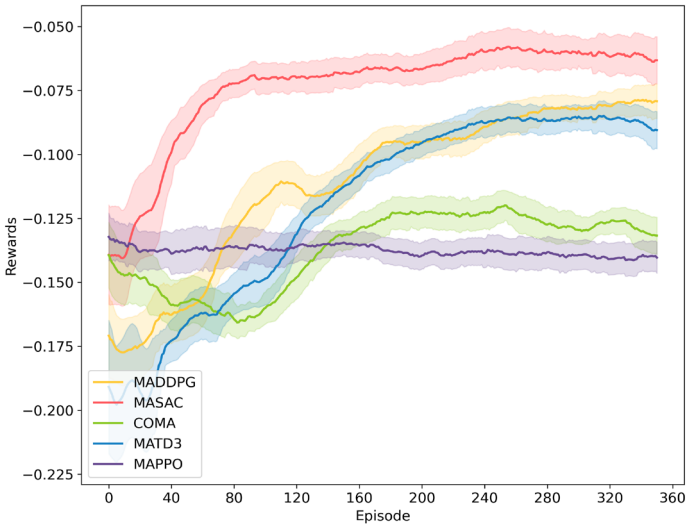

Figure 7. Rewards on the IEEE 141-bus network.

Figure 8 illustrates the reactive power loss curve of the IEEE 141-bus network. An analysis of the convergence results indicates that the MASAC algorithm effectively controls average reactive power loss, achieving a value of 0.49 MVAR. This performance surpasses that of the MADDPG and MATD3 algorithms significantly. The MASAC algorithm demonstrates adeptness in managing the reactive power loss control challenge within large DNs.

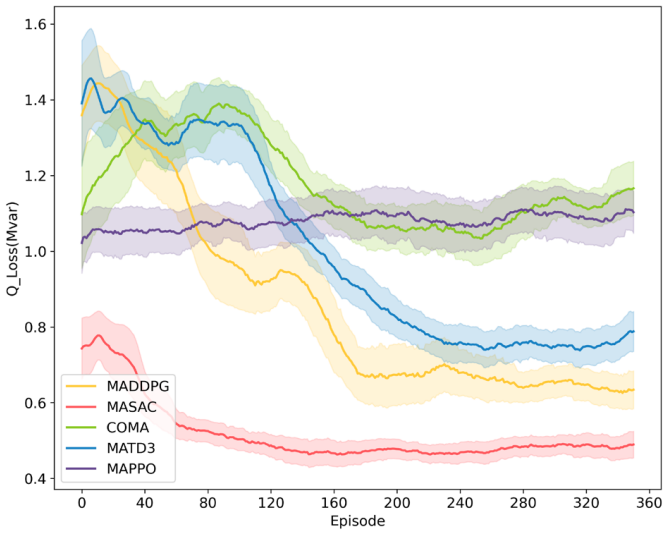

Figure 8. Reactive power loss on the IEEE 141-bus network.

The controllable ratio curve is depicted in Figure 9, which shows that the MASAC algorithm achieves a relatively high level of convergence in the controllable ratio. Additionally, the algorithm demonstrates a higher controllable ratio in the IEEE 141-bus network compared with the IEEE 33-bus network. This indicates that the MASAC algorithm is capable of stabilizing the voltage of nodes in larger-scale networks more effectively.

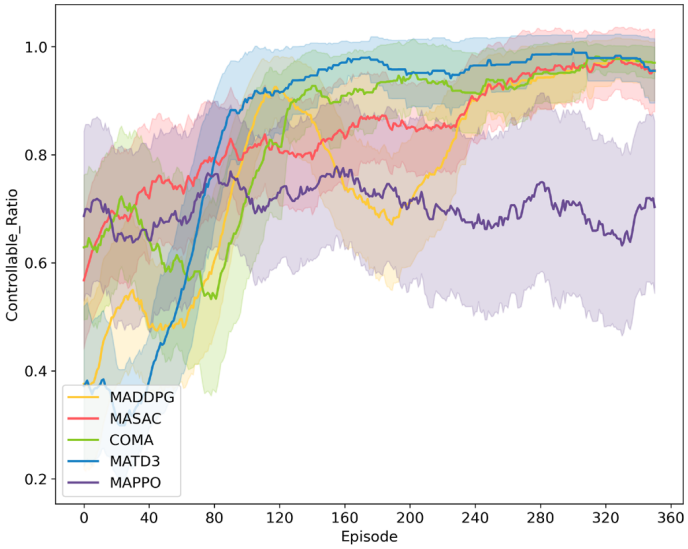

Figure 9. Controllable ratio on the IEEE 141-bus network.

The voltage control results attained on the test set of five control methods in the IEEE 141-bus network are presented in Table 4. The average voltage value of the MASAC algorithm is 0.9957 p.u., the control of reactive power loss is 0.5653 MVAR, and the controllable ratio is 98.41%, which achieves a better control effect compared with the other algorithms.

Table 4. Performance of control strategies on the IEEE 141-bus network.

Strategy	Average Voltage	Maximum Voltage Rise	Minimum Voltage Drop	Reactive Power Losses	Controllable Ratio
MASAC	0.9957	0.003	0.001	0.5653	98.41%
MADDPG	0.9932	0.006	0.002	0.7159	98.50%
MATD3	0.9854	0.005	0.001	0.7940	98.23%
MAPPO	1.0255	0.013	0.004	1.0755	68.73%
COMA	1.0046	0.001	0.003	1.1765	99.99%

By carrying out arithmetic simulations on the IEEE 141-bus network, it was found that the proposed MASAC algorithm has excellent performance in all metrics and can deal with the cooperative VVC problem in large-scale distribution networks with better scalability than the other MARL algorithms.

4. Conclusions

In this paper, we address the VVC problem in ADNs through the utilization of MASAC. Initially, we propose a partitioning strategy for ADN based on reactive voltage sensitivity to divide the network into multiple sub-regions. This partitioning approach serves as the foundation for establishing a CTDE framework relying on MASAC. Subsequently, we formulate the VVC problem involving multiple PV inverters as an MG and apply the MASAC algorithm to tackle the VVC challenges in ADNs. We conduct comparative experiments employing various MARL algorithms on the IEEE 33-bus and IEEE 141-bus

networks. The experimental outcomes substantiate the efficacy of our proposed methodology in enabling cooperative VVC control of multiple PV inverters. Notably, it effectively mitigates voltage deviations and reduces reactive power losses, leveraging solely locally observed information.

Future research work will mainly focus on the following:

1. Broadening the scope of controllable power electronic devices;
2. Improving the scalability of the algorithm to tackle voltage and VVC challenges in larger-scale DNs.

Author Contributions: Conceptualization, S.S.; methodology, H.Z.; software, L.Z.; validation, Q.X. and R.S.; formal analysis, Y.D.; investigation, T.G.; resources, L.W.; data curation, J.Z.; visualization, L.S.; writing—original draft, S.S., H.Z. and L.Z.; writing—review and editing, Q.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Science and Technology Project of Yunnan Power Grid Co., Ltd. (No. YNKJXM20222105).

Data Availability Statement: The data presented in this study are available in this article.

Conflicts of Interest: Authors Shi Su and Qingyang Xie were employed by the company Yunnan Power Grid. The remaining authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Wang, Y.; Qiu, D.; Strbac, G.; Gao, Z. Coordinated Electric Vehicle Active and Reactive Power Control for Active Distribution Networks. *IEEE Trans. Ind. Inform.* **2022**, *19*, 1611–1622. [CrossRef]
2. Liu, H.; Wu, W. Online Multi-Agent Reinforcement Learning for Decentralized Inverter-Based Volt-VAR Control. *IEEE Trans. Smart Grid* **2021**, *12*, 2980–2990. [CrossRef]
3. Lee, X.Y.; Sarkar, S.; Wang, Y. A Graph Policy Network Approach for Volt-Var Control in Power Distribution Systems. *Appl. Energy* **2022**, *323*, 119530. [CrossRef]
4. Dorostkar-Ghamsari, M.R.; Fotuhi-Firuzabad, M.; Lehtonen, M.; Safdarian, A. Value of Distribution Network Reconfiguration in Presence of Renewable Energy Resources. *IEEE Trans. Power Syst.* **2016**, *31*, 1879–1888. [CrossRef]
5. Zhang, L.; Shen, C.; Chen, Y.; Huang, S.; Tang, W. Coordinated Allocation of Distributed Generation, Capacitor Banks and Soft Open Points in Active Distribution Networks Considering Dispatching Results. *Appl. Energy* **2018**, *231*, 1122–1131. [CrossRef]
6. Long, C.; Ochoa, L.F. Voltage Control of PV-Rich LV Networks: OLTC-Fitted Transformer and Capacitor Banks. *IEEE Trans. Power Syst.* **2016**, *31*, 4016–4025. [CrossRef]
7. Minetti, M.; Rosini, A.; Denegri, G.B.; Bonfiglio, A.; Procopio, R. An Advanced Droop Control Strategy for Reactive Power Assessment in Islanded Microgrids. *IEEE Trans. Power Syst.* **2021**, *37*, 3014–3025. [CrossRef]
8. Antoniadou-Plytaria, K.E.; Kouveliotis-Lysikatos, I.N.; Georgilakis, P.S.; Hatziargyriou, N.D. Distributed and Decentralized Voltage Control of Smart Distribution Networks: Models, Methods, and Future Research. *IEEE Trans. Smart Grid* **2017**, *8*, 2999–3008. [CrossRef]
9. Li, Z.; Su, S.; Jin, X.; Xia, M.; Chen, Q.; Yamashita, K. Stochastic and Distributed Optimal Energy Management of Active Distribution Networks within Integrated Office Buildings. *CSEE J. Power Energy Syst.* **2024**, *10*, 504–517. [CrossRef]
10. Xu, T.; Wu, W. Accelerated ADMM-Based Fully Distributed Inverter-Based Volt/Var Control Strategy for Active Distribution Networks. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7532–7543. [CrossRef]
11. Li, P.; Zhang, C.; Wu, Z.; Xu, Y.; Hu, M.; Dong, Z. Distributed Adaptive Robust Voltage/Var Control with Network Partition in Active Distribution Networks. *IEEE Trans. Smart Grid* **2019**, *11*, 2245–2256. [CrossRef]
12. Ma, W.; Wang, W.; Chen, Z.; Hu, R. A Centralized Voltage Regulation Method for Distribution Networks Containing High Penetrations of Photovoltaic Power. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106852. [CrossRef]
13. Kumar Tatikayala, V.; Dixit, S. Multi-Stage Voltage Control in High Photovoltaic Based Distributed Generation Penetrated Distribution System Considering Smart Inverter Reactive Power Capability. *Ain Shams Eng. J.* **2024**, *15*, 102265. [CrossRef]
14. Yang, T.; Guo, Y.; Deng, L.; Sun, H.; Wu, W. A Linear Branch Flow Model for Radial Distribution Networks and Its Application to Reactive Power Optimization and Network Reconfiguration. *IEEE Trans. Smart Grid* **2021**, *12*, 2027–2036. [CrossRef]
15. Liu, H.; Zhang, C.; Guo, Q. Data-Driven Robust Voltage/VAR Control Using PV Inverters in Active Distribution Networks. In Proceedings of the 2020 International Conference on Smart Grids and Energy Systems (SGES), Perth, Australia, 23–26 November 2020; pp. 314–319.
16. Sun, X.; Qiu, J. Two-Stage Volt/Var Control in Active Distribution Networks with Multi-Agent Deep Reinforcement Learning Method. *IEEE Trans. Smart Grid* **2021**, *12*, 2903–2912. [CrossRef]

17. Cao, D.; Zhao, J.; Hu, W.; Ding, F.; Huang, Q.; Chen, Z.; Blaabjerg, F. Data-Driven Multi-Agent Deep Reinforcement Learning for Distribution System Decentralized Voltage Control With High Penetration of PVs. *IEEE Trans. Smart Grid* **2021**, *12*, 4137–4150. [CrossRef]
18. Chen, Y.; Liu, Y.; Zhao, J.; Qiu, G.; Yin, H.; Li, Z. Physical-Assisted Multi-Agent Graph Reinforcement Learning Enabled Fast Voltage Regulation for PV-Rich Active Distribution Network. *Appl. Energy* **2023**, *351*, 121743. [CrossRef]
19. Thurner, L.; Scheidler, A.; Schäfer, F.; Menke, J.-H.; Dollichon, J.; Meier, F.; Meinecke, S.; Braun, M. Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Trans. Power Syst.* **2018**, *33*, 6510–6521. [CrossRef]
20. Zimmerman, R.D.; Murillo-Sánchez, C.E.; Thomas, R.J. MATPOWER: Steady-State Operations, Planning, and Analysis Tools for Power Systems Research and Education. *IEEE Trans. Power Syst.* **2011**, *26*, 12–19. [CrossRef]