International Journal of Innovative Computing, Information and Control
Volume 19, Number 4, August 2023
ICIC International ©2023 ISSN 1349-4198
pp. 1179–1193

RESEARCH ON FREQUENCY OPTIMIZATION STRATEGY OF MULTI VIRTUAL SYNCHRONOUS GENERATOR MICROGRID BASED ON Q-REINFORCEMENT LEARNING

Jiong Jiang[1], Yuhao Lu[2], Tinglong Pan[1,*], Weilin Yang[1], Yachao Wang[3] and Zhenlan Dou[4]

[1]School of Internet of Things Engineering
Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
6201920009@stu.jiangnan.edu.cn; wlyang@jiangnan.edu.cn
*Corresponding author: tlpan@jiangnan.edu.cn

[2]Power China Hubei Electric Engineering Co., Ltd.
No. 1, Xinqiao 4th Road, Wuhan 430040, P. R. China
306546267@qq.com

[3]State Grid Jibei Marketing Service Center (Metrology Center)
No. 1, Dizang'an South Lane, Beijing 100045, P. R. China
lubert_job@163.com

[4]State Grid Shanghai Integrated Energy Service Co., Ltd.
No. 677, Jumen Road, Huangpu District, Shanghai 200001, P. R. China
douzhl@126.com

Received November 2022; revised February 2023

Abstract. *When the virtual synchronous machine control strategy has large capacity load switching, the frequency will exceed the limit which will affect the safe operation of the system. To solve the above problems, this paper proposes a multi-agent distributed frequency modulation control strategy based on reinforcement learning, which detects the frequency deviation of the micro-grid system, defines the local reward and global reward, designs the distributed information interaction process, calculates the control quantity of the secondary frequency modulation of each controller by using Q learning, overcomes the deficiency of centralized frequency modulation, and makes the coordinated output of distributed generators asymptotically eliminate the frequency deviation. The simulation results verify the effectiveness of the proposed control strategy and this control strategy ensures the safe and stable operation of microgrid.*

Keywords: Virtual synchronous machine, Microgrid, Multi-agent, Reinforcement learning, Frequency

1. **Introduction.** State Grid Corporation of China has launched the "Peak Carbon and Carbon Neutral" action plan guided by double carbon targets of 'peak carbon and carbon neutral'. In this context, wind power generation, photovoltaic power generation as the core of the distributed generation system will get rapid and transcendental development [1].

P/Q control [2], V/F control [3], sag control [4], VSG (Virtual Synchronous Generator) control [5] and other methods are applied to existing converter control strategies used in new energy power generation. According to the new IEEE standard 1574-2018, renewable energy generation systems not only should be operated under both grid-connected and independent microgrid status, but also can achieve the seamless handover between the two

DOI: 10.24507/ijicic.19.04.1179

states, so that it can ensure the quality of power supply to critical loads in the microgrid. However, when the virtual synchronous machine control strategy operates in the independent microgrid, the system frequency deviates from the rated value when a large capacity load is switched on, and the primary frequency regulation cannot eliminate the frequency deviation, which can seriously lead to frequency overrun and endanger the stability of the system. Therefore, it is necessary to introduce secondary frequency modulation to enhance the frequency regulation ability under independent microgrid operation.

Based on VSG's research on micro-grid primary frequency regulation offset and out of bounds, many researchers have carried out research on micro-grid secondary frequency regulation. In [6], the necessity for secondary frequency regulation in microgrids is discussed. In [7], authors propose a secondary frequency modulation method of sag inverter in island operation based on integral advance compensation link, which eliminated the static difference of microgrid frequency caused by sag control. On the basis of PI frequency modulation, [8] proposes a secondary self-recovery control method of frequency and voltage in independent microgrid under the control strategy of virtual synchronous machine, which realized the autonomous recovery control of frequency and voltage of converter and power distribution control. In [9], authors provide strategy of virtual synchronous generator differential-free frequency regulation, which is the traditional virtual synchronous generator frequency control loop which is replaced by a controlled frequency integral feedback loop, and enables the microgrid inverter to automatically switch between primary and secondary frequency regulation according to the amount of angular velocity change, in angular velocity enabling the recovery of microgrid frequency and proportional power distribution. [10] builds a scheduling model based on equal micro-increase rate for isolated island microgrid, and uses distributed algorithm to achieve fast frequency recovery at minimum frequency modulation cost. In [11], a secondary regulation of frequency based on the exchange of frequency information among adjacent DGs is proposed, but it relies on the communication of the central controller to achieve. [12] proposes a centralized control mode for secondary frequency regulation, but for multiple VSG FM systems, communication systems need to be relied on to achieve reasonable distribution of load power.

Since the impedance of the circuit in the power grid cannot be completely ignored, serious coupling problems will occur when VSG outputs active and reactive power, and the deployment accuracy of the power grid will be reduced. Therefore, it is necessary to study the parallel power distribution of multiple VSGS [13]. [14] studies the situation that reactive power cannot be evenly distributed in microgrid, and realized that reactive power can be evenly distributed through VSG without being disturbed by the distribution of active power and resistance mismatch in the circuit. For the problem of unbalanced reactive power distribution, [15] adopts the research method combining adaptive control algorithm and output impedance to reduce the influence of resistance in the circuit on power distribution of the grid. [16] improves the droop control scheme, realizing that each module in the system can operate independently without communication and reducing the frequency offset. However, this scheme cannot guarantee the average distribution of reactive power. [17] has optimized the droop controller, enabling it to achieve more accurate power distribution in inverters of different models and on different occasions, and control the voltage within a given range. In [18], this paper adopts Lie derivative to linearize the state feedback of VSG large signal model to simplify the nonlinearity of VSG large signal model. In [19], a nonlinear control strategy based on the projection operator adaptive method with fractional-order sliding-mode backstepping control is proposed as a supplementary control strategy for VSG.

In addition, in the field of power prediction, the continuous development of reinforcement learning based on artificial intelligence technology has led to the main application of reinforcement learning algorithm in power system, power grid fault detection and diagnosis, while the role of reinforcement learning algorithm in deeper fields such as operation control remains to be explored. In [20], authors propose a distributed quadratic optimal control method based on reinforcement learning in-situ feedback method. Aiming at the problem of static deviation of system frequency and voltage caused by primary control sag of distributed power supply in microgrid, the needs of frequency recovery and voltage adjustment can be considered when microgrid uses local information. In [21], an intelligent algorithm based on Q learning embedded in multi-agent hierarchical hybrid control model is proposed to dynamically predict the real-time secondary FM power deficit values for microgrid systems. [22] proposes a reinforcement learning-based frequency coordination control strategy for multiple optical storage virtual synchronizers. Secondary coordination control of the frequency of the microgrid multiple optical storage virtual synchronizers achieved by the central controller of the microgrid monitors the system frequency deviation in real time and uses reinforcement learning algorithms to calculate the power deficit of the microgrid.

The problem of system frequency deviation from the rated value is studied above. Generally, the centralized control VSG frequency coordination strategy is adopted. The centralized control strategy has complex communication and large line investment. Based on VSG system model, combined with distributed control scheme and reinforcement learning algorithm, this paper proposes a multi-agent distributed secondary frequency modulation strategy based on reinforcement learning. The contributions of the paper are mainly as follows.

1) Because of the adoption of Q learning, the strategy does not need to consider the subsequent state and surrounding environment under the current state, which simplifies and accelerates the decision-making process.

2) The proposed control strategy uses the global reward of Q learning as the local feedback, and optimizes and modifies it according to the changes of network topology, so as to realize the fast convergence of reward and the optimization of results, and achieve a good frequency modulation effect.

3) Multi-agent distributed information interaction is adopted to reduce communication delay and improve communication performance.

This paper is organized as follows. In Section 2, the mathematical model of VSG is introduced. In Section 3, the design process of the proposed control strategy is described. In Section 4, the feasibility of the strategy is proved by simulation. In Section 5, the conclusion is presented.

2. The Control Principle and Frequency Modulation Principle of Virtual Synchronous Generator.

2.1. **Control principle of virtual synchronous generator.** Referring to the rotor motion equation and the stator voltage and current relationship of synchronous generator, the mathematical model of VSG can be expressed as

$$\begin{cases} J\dfrac{d\omega}{dt} = T_m - T_e - D(\Delta\omega), & x \leq 1 \\ \dfrac{d\theta}{dt} = \omega, & x > 1 \end{cases} \tag{1}$$

where θ is the phase angle of the excitation electromotive force, T_m is the mechanical torque, T_e is the electromagnetic torque, D is damping coefficient, J is the virtual moment of inertia, and $\Delta\omega$ is the speed deviation.

The main circuit of virtual synchronous generator is grid-connected inverter, and its main structure includes ideal DC power supply, DC/AC converter and filter. The control system is the core of virtual synchronous generator, mainly including active frequency controller, reactive voltage controller, etc. The overall structure of VSG is shown in Figure 1.

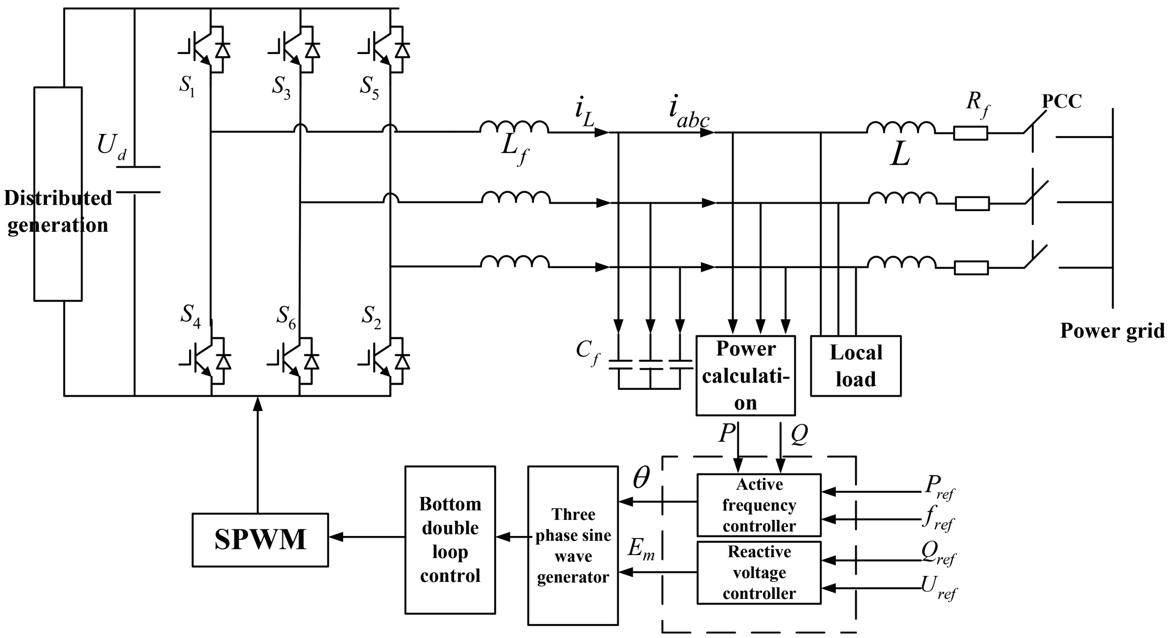

FIGURE 1. Overall structure of VSG

The mathematical model of VSG's active-frequency control link can be obtained by referring to the active frequency droop characteristics of synchronous generator:

$$J\omega_n \frac{d\omega}{dt} = P_{ref} - P_m - (K_f + D\omega_n)(\Delta\omega) \tag{2}$$

where J is the virtual moment of inertia, D is the virtual damping coefficient, ω_n is the rated speed, $\Delta\omega$ is the speed deviation, P_{ref} is the rated active power, and P_m is the mechanical power.

According to Formula (2), the overall structure of the active power control ring of the virtual synchronous generator is shown in Figure 2.

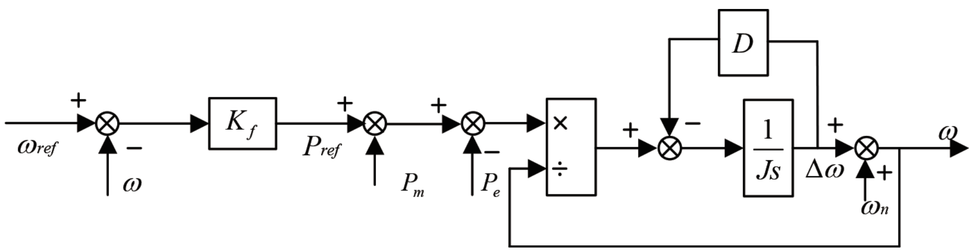

FIGURE 2. Active-frequency control block diagram

Referring to the sagging characteristics of reactive voltage of synchronous generator, the mathematical model of VSG virtual excitation regulator can also be obtained:

$$E = U_0 + \frac{1}{Ks}\left[Q_{ref} + D_Q\left(U_{ref} - U\right) - Q\right] \tag{3}$$

where K is integral coefficient, s is operator variable, D_Q is droop coefficient, E is excitation electromotive force, U_0 is no-load electromotive force, Q_{ref} is rated reactive power, Q is the actual reactive power, U_{ref} is the output voltage set value, and U is the actual output voltage.

According to Formula (3), the overall structure of the reactive power-voltage control ring of the virtual synchronous generator is shown in Figure 3.

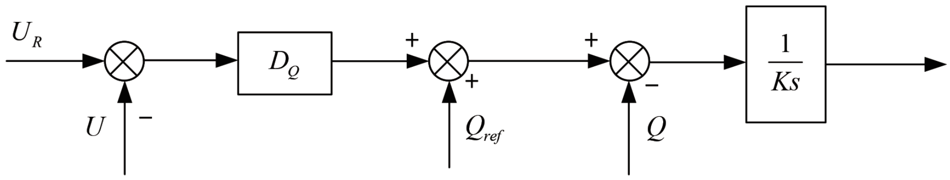

FIGURE 3. The block diagram of reactive power-voltage control

2.2. **Frequency modulation principle of virtual synchronous generator.** According to Formulae (2) and (3), the transient relationship between frequency and power is

$$\omega - \omega_n = \frac{P_{ref} - P_m}{D + K_f}\left(1 - e^{-\frac{D+K_f}{J}t}\right) \tag{4}$$

Formula (4) shows that when the VSG with load of power change, VSG output frequency changes exponentially, the moment of inertia J, damping coefficient D and frequency deviation feedback coefficient K_f jointly determine the dynamic transition time of frequency, the damping coefficient D and frequency deviation feedback coefficient K_f determine the system frequency deviation degree at the steady state. The frequency deviation can be reduced by increasing the sum of D and K_f. However, excessive increase of D and K_f will not only affect the stability of the system, but also accelerate the dynamic transition process of frequency and increase the frequency change rate, which is not conducive to the stability of the system frequency. Therefore, it is necessary to draw lessons from the principle of secondary frequency modulation in power system to restore the frequency to the rated value. If the principle of secondary frequency modulation in the power system is used for reference to make VSG power variation completely offset load variation, frequency no-difference control can be achieved without affecting its stability. The frequency modulation process is shown in Figure 4.

3. **Distributed Frequency Control Strategy Based on Q-Reinforcement Learning.** Q-Reinforcement learning can improve the performance of agents to adapt the environmental changes constantly in the communication and interaction between the agent and the environment. It is an excellent algorithm for the study of agents, so it can be applied in multi-agent automatic control.

3.1. **The principle of Q learning algorithm.** The reinforcement learning has two basic frameworks: environment and agent. The agent consists of three modules: input, reinforcement and strategy. The input module is the environment state setting A, which is entered by the agent. The reinforcement module gives a value to each state quantity, which is the control object of the agent. The maximum cumulative reward value is the learning

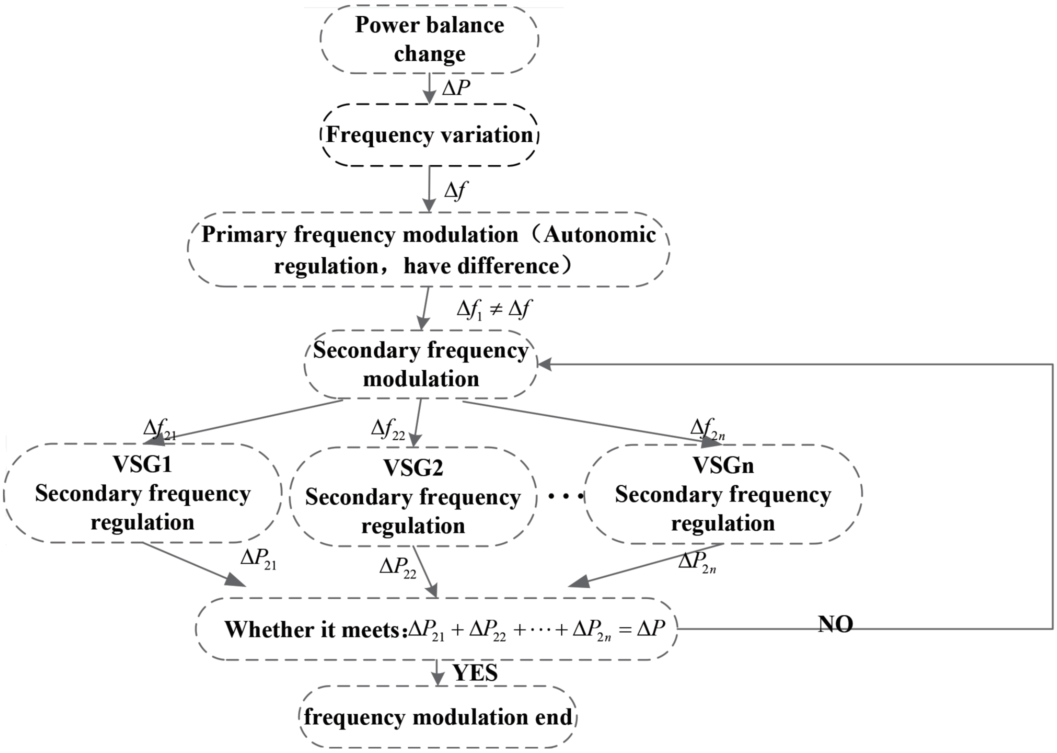

FIGURE 4. The process of VSG participating in frequency modulation

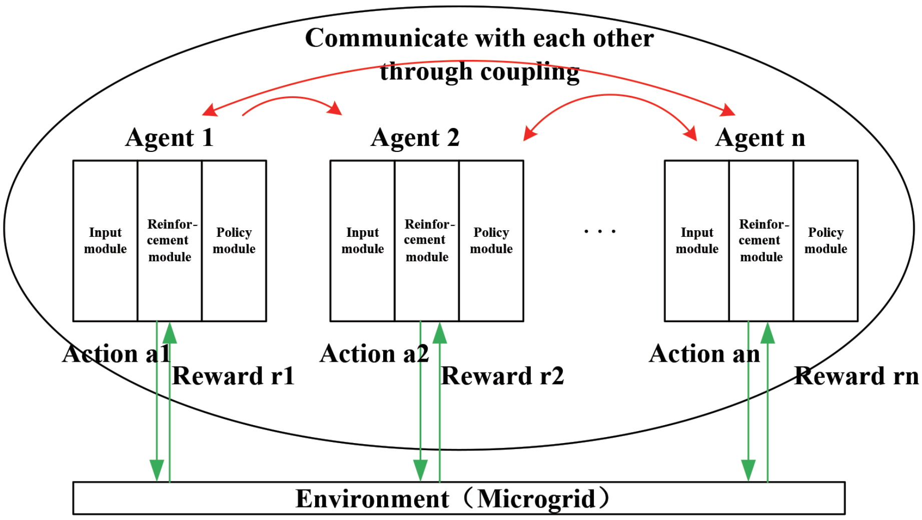

FIGURE 5. The framework of distributed multi-agent reinforcement learning

target. The strategy module is the core. It first updates the agent knowledge base, and then selects the optimal action set S according to certain action selection strategy and carries out the action. Then the state changes and agents communicate with each other through coupling, as shown in Figure 5.

As one of the most basic decision algorithms in reinforcement learning algorithm, the goal of Q learning algorithm is to find an optimal strategy to maximize the expected discount reward value. Because Q learning algorithm does not depend on the environment, it can be used in the model of unknown environment. For the state-action value function $Q(s, a)$ can be directly optimized and iterated to find the optimal strategy online. Define the value function as follows:

$$Q^{\pi}(s_t, a_t) = r(s_t, a_t) + \gamma V^{\pi}(s_{t+1}) \tag{5}$$

where $Q^{\pi}(s_t, a_t)$ is the expected value of taking action a_t to get reward under s_t state in strategy π, and the subscript t represents the time; $r(s_t, a_t)$ is the immediate reward value obtained by taking the action a_t in the s_t state, and $V^{\pi}(s_{t+1})$ is the corresponding value function under the state s_{t+1} in the policy π.

The action obtained each time is obtained by greedy algorithm, and the optimal action of the system can be expressed as

$$\pi^* = \arg\max_{a} Q(s, a) \tag{6}$$

where the function $\arg\max f(x)$ represents the value of x when $f(x)$ is at its maximum value.

If S and A can be divided into g and h discrete intervals respectively, all $Q(s, a)$ compose $g * h$ order Q matrix. The experience of changing the environment from some initial state to a target state is commonly referred to as a scene. No matter what state VSG is in when making decisions, the optimal action strategy is to select the action sequence corresponding to the maximum Q value according to the learned Q matrix for action. Compared with Sarsa and sequential difference learning, the advantage of Q learning is that the current Q table includes all the required information, does not need to know the state transition function, does not involve the subsequent state under the current state, and does not relate to the environmental model, which greatly simplifies and speeds up the decision-making process.

3.2. **The design of Q learning frequency controller.** Micro-grid frequency secondary control can reduce the system frequency deviation by controlling the active power output of each VSG. Environmental state input variable is frequency deviation Δf of micro-grid, and controller output action variable is secondary control active power ΔP. ΔP can be discretized as $\{\Delta P_1, \Delta P_2, \ldots, \Delta P_m\}$, and corresponding actions set $A\{a_1, a_2, \ldots, a_m\}$. If the separation value is too fine, the state-action pair dimension is too high, and the real-time optimization of the system suffers from "dimension disaster", and the algorithm convergence is slow. Too sparse will reduce the precision of the controller. Therefore, it is very important to discretize state space and action space effectively and reasonably. In this paper, the goal of constant frequency secondary control is (50 ± 0.001) Hz. Considering the above factors and the active power-frequency characteristic parameters of each inverter, the frequency deviation state set designed in this paper is $S = \{s_1, s_2, \ldots, s_n\} = \{(-\infty, -0.1); [-0.1, -0.05); [-0.05, -0.01); [-0.01, -0.001); [-0.001, 0.001); [0.001, 0.01); [0.01, 0.05); [0.05, 0.1); [0.1, +\infty)\}$, active action set $A = \{a_1, a_2, \ldots, a_m\} = \{-0.2, -0.15, -0.1, -0.08, -0.06, -0.04, -0.02, -0.01, -0.005, -0.003, -0.0005, 0, 0.0005, 0.003, 0.005, 0.01, 0.02, 0.04, 0.06, 0.08, 0.10, 0.15, 0.2\}$. Second control power change ΔP is calculated according to the following formula:

$$\Delta P = a_k P_i^* \quad k \in A \tag{7}$$

where P_i^* represents the rated active power of the ith VSG, a_k is the element of action concentration, is a dimensionless proportional coefficient, and represents the ratio of active power action to rated power.

Define the local reward R after an agent performs an action as

$$r = R_{\Delta f} + R_U \tag{8}$$

$$R_{\Delta f} = \begin{cases} 0, & |\Delta f| \leq b_1 \\ -\lambda_1|\Delta f|, & b_1 < |\Delta f| \leq b_2 \\ -\lambda_2|\Delta f|, & b_2 < |\Delta f| \leq b_3 \\ -\lambda_3|\Delta f|, & b_3 < |\Delta f| \leq b_4 \\ -\lambda_4|\Delta f|, & |\Delta f| > b_4 \end{cases} \tag{9}$$

$$R_U = \begin{cases} 1, & U_1 \leq U \leq U_2 \\ -1, & \text{otherwise} \end{cases} \tag{10}$$

where $R_{\Delta f}$ is frequency penalty; λ_1, λ_2, λ_3, λ_4 are positive coefficients, set to 5, 15, 30 and 50, respectively; parameters b_1, b_2, b_3, b_4 are set to 0.0001, 0.01, 0.05 and 0.1, respectively; Δf is frequency deviation; R_U is voltage constraint term. U_1 and U_2 are the upper and lower limits of reasonable voltage amplitude U respectively, $U_1 = 0.9U$, $U_2 = 1.1U$.

Based on the local reward definition, the global reward is

$$R_F = \sum_{i=1}^{n} r_{i,f} \tag{11}$$

where sum of $r_{i,f}$ is the local frequency reward, R_F is the global frequency reward, indicating that each learning selects the current action when the global reward value is the maximum to control each inverter.

3.3. **The interaction of multi-agent distributed information.** The distributed control scheme adopted in this paper takes the global reward value that sums local rewards of each agent as the global feedback quantity to control the frequency. According to the above definition of local reward, the distributed information interaction process based on reinforcement learning is described as follows:

$$r_{i,f}^{K+1} = \sum_{j \in N_i} \left(r_{j,f}^{(k)} - r_{i,f}^{(k)} \right) \tag{12}$$

$$\varepsilon_{ij} = \begin{cases} \dfrac{2\rho}{n_i + n_j}, & j \in N_i \\ 1 - \displaystyle\sum_{j \in N_i} \dfrac{2\rho}{n_i + n_j}, & j = i \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

where $r_{i,f}^{K+1}$ is the local frequency reward of the ith agent in the $k+1$ iteration, and ε_{ij} is the communication coupling coefficient between agents i and j; ρ is a constant, $0 \leq \rho \leq 1$; n_i and n_j are the number of agents in the neighborhood of agents i and j respectively, and N_i is the set of agents adjacent to the ith agent.

When the local reward $r_{i,f}^{K+1}$ finally converges to $r_f^{x_1}$, the global reward R_F acquisition process is as follows:

$$R_F = N r_f^{x_1} \tag{14}$$

where N is the total number of intelligent agents.

When any VSG's action changes at a certain moment, it will traverse through the action set, and then calculate the reward expectation of other VSGS respectively. Due to the Q learning, only the state after one action is calculated, so the calculation is not much. After the distributed information interaction, the reward value and Q value are corrected, the status is updated, and the above content is iterated, so that the global optimal VSG action value can be obtained according to the maximum Q value.

3.4. **The strategy flow of distributed frequency control strategy based on Q learning.** The detailed frequency modulation steps are as follows.

Step 1: Initialize all Q values, $k = 0$, and the initial state is f_0.

Step 2: According to the current state, detect the microgrid frequency Δf, and each agent selects the action by greedy strategy.

Step 3: Local reward $r_{i,f}^{K+1}$ is calculated according to Equations (8), (9) and (10).

Step 4: Based on the communication coupling ε_{ij} between them, each agent implements distributed frequency modulation based on Q learning and updates local reward accordingly.

Step 5: As shown in Equation (11), all agents share the global information R_F.

Step 6: Calculate Q value, each agent synchronously control f_i^k according to the action corresponding to the maximum Q value, and transfer to the new state.

Step 7: Return to Step 1, frequency modulation ends when $|\Delta f \leq \varepsilon|$ is satisfied.

The specific flow chart is shown in Figure 6.

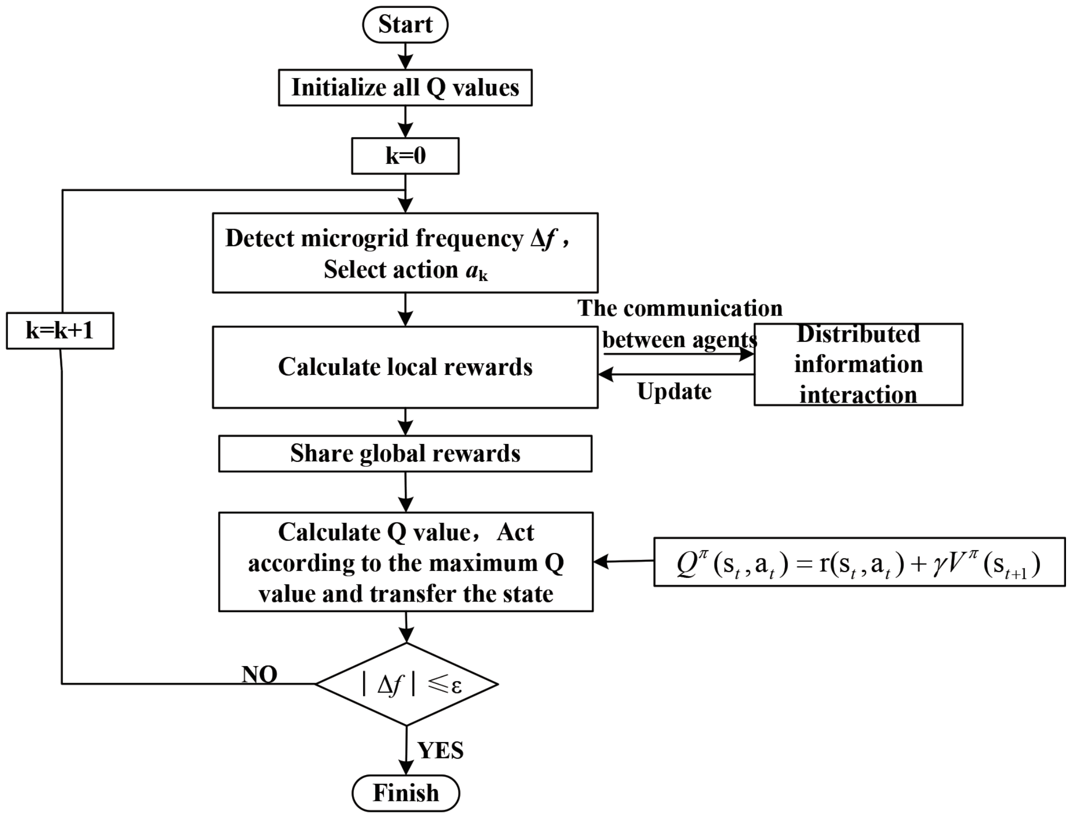

FIGURE 6. Flow chart of distributed frequency modulation strategy based on Q learning

4. Simulation.

4.1. **Simulation platform construction.** In order to verify the effectiveness and correctness of the frequency modulation strategy of the virtual synchronous machine based on reinforcement learning proposed in this paper, a microgrid model is established on the Matlab/Simulink simulation platform, as shown in Figure 7.

The virtual synchronizer parameters in the simulation are shown in Table 1. The microgrid system is selected to be in island mode, switch S1 is disconnected, the power of VSG1 to VSG4 is 9.2 kW, 7.3 kW, 10.4 kW and 13.1 kW respectively, and load 1 to 5 is 5.2 kW, 7.8 kW, 8.1 kW, 8.7 kW and 10.2 kW respectively.

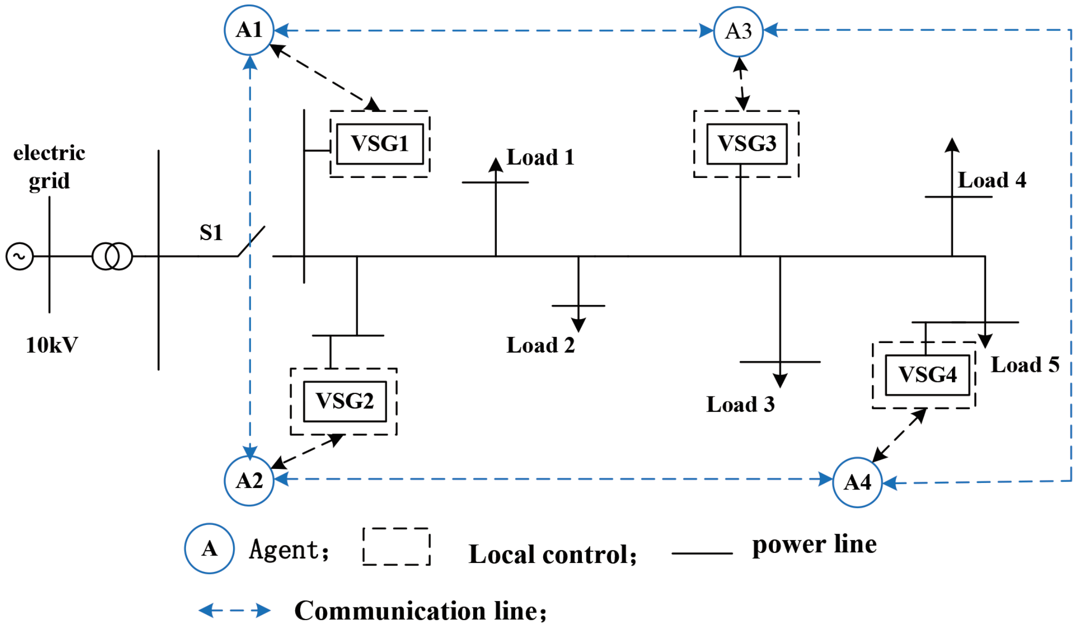

FIGURE 7. Microgrid simulation model

TABLE 1. The main parameters of the virtual synchronization machine

Electrical parameters	Value	Electrical parameters	Value
Voltage supply U_d	700 V	Virtual inertia J	0.213
Inductance L_f	1.3 mH	Damping coefficient D	12.14
The inductor on the electrical network L	1.6 mH	Sag factor D_{Q1}	1746.78
Resistance R_f	0.18 Ω	Sag factor D_{Q2}	0.213
Filter capacitor C_f	20 µF	Sag factor D_{Q3}	0.213
Rated frequency F	50 Hz	Sag factor D_{Q4}	0.213
Integral coefficient K	45.2	Virtual impedance D_1	4 mH

4.2. **The analysis of simulation result.** In order to verify the effectiveness of the proposed frequency modulation strategy based on Q-Reinforcement learning, when $t = 1$ s, load 1 mutates to 8.2 kW and 1.2 s carries out the proposed secondary frequency modulation strategy. The simulation results of system frequency are shown in Figure 8 and Table 2.

As can be seen from Figure 8, in the initial state, the system frequency stabilizes near the rated frequency of 50 Hz. The load changes at 1.0 s, at which point VSG begins to participate in the self-regulation of primary frequency modulation. Due to virtual impedance, the frequency modulation process is relatively smooth. After the end of frequency modulation, the system frequency is stable around 49.96 Hz, Q learning frequency modulation starts from 1.2 s and is performed every 0.1 s. After 5 times of learning, the deviation of system frequency is -0.0008, so the frequency modulation is over. Compared with the traditional PI FM, the proposed frequency modulation speed is faster and time is shorter. The active power output of each VSG is shown in Figure 9 and Table 3.

It can be seen from Figure 9 and Table 3 that VSG's output power changes because of the load mutation at 1.0 s, and VSG participates in primary frequency modulation. Due to the difference in primary frequency modulation, the system frequency cannot be restored after the end of the primary frequency modulation. After starting the second

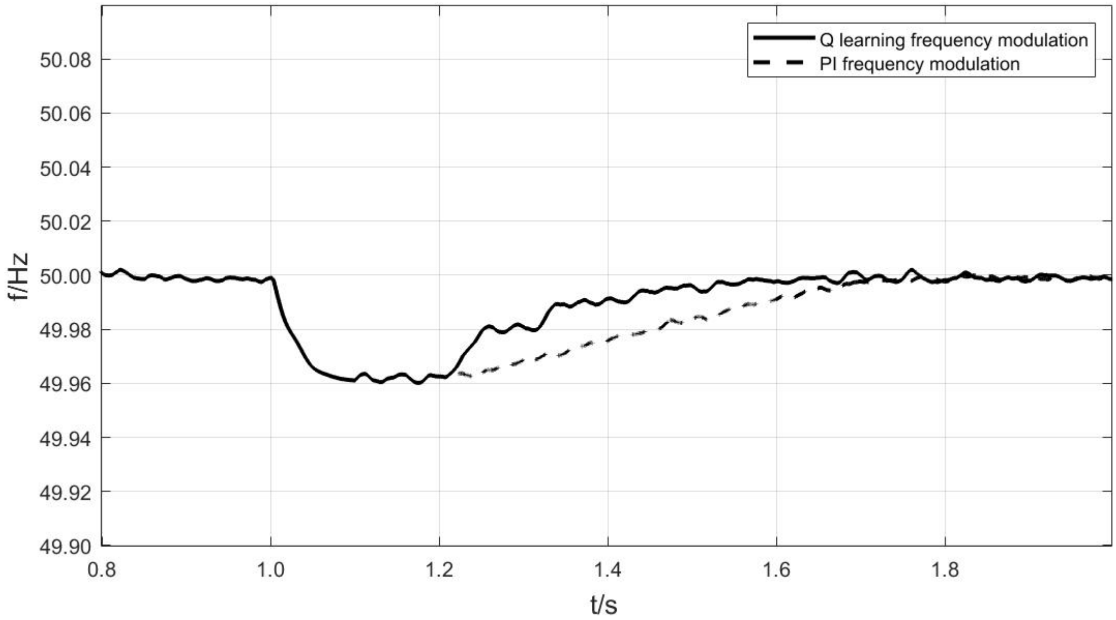

FIGURE 8. System frequency changes under different FM strategies

TABLE 2. System frequency deviation under Q learning

Time t	1.2 s	1.3 s	1.4 s	1.5 s	1.6 s
System frequency deviation Δf	-0.04	-0.02	-0.008	-0.005	-0.0008

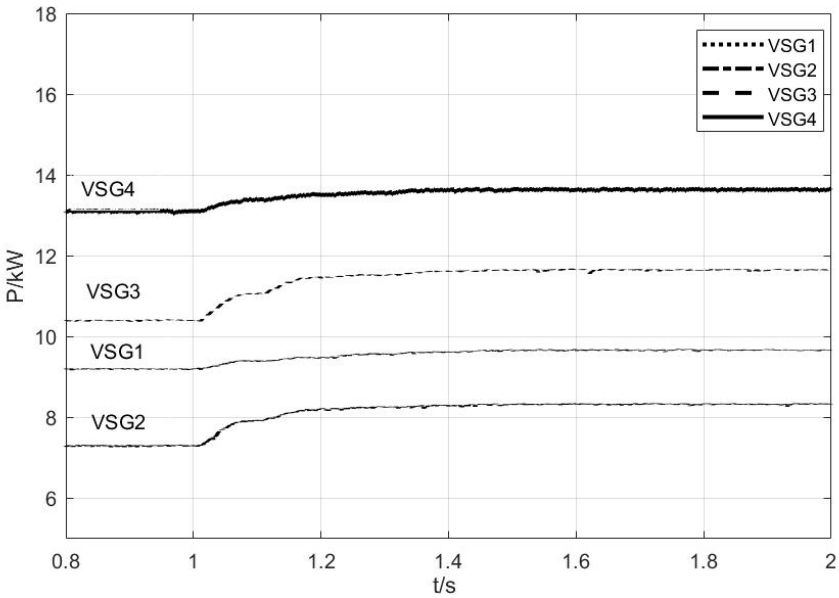

FIGURE 9. Change of VSG's active power

FM strategy at 1.2 s, VSG's output power continued to increase. After 5 adjustments, the output power of each VSG is not adjusted, and the output remains stable. The VSG output power is re-balanced with the load power.

TABLE 3. System frequency deviation under Q learning

Time / VSG	1.0 s	1.2 s	1.3 s	1.4 s	1.5 s	1.6 s
VSG_1	0.08	0.01	0.005	0.005	0.0005	0.0005
VSG_2	0.02	0.01	0.005	0.005	0.0005	0.0005
VSG_3	0.06	0.005	0.005	0.003	0.0005	0.0005
VSG_4	0.02	0.005	0.005	0.003	0.0005	0.0005

4.3. **The analysis of communication delay.** When we study the distributed control of microgrid, communication is one of the most important links. Communication delay is the key factor to be considered in the online control of microgrid, which can be calculated by the following formula:

$$T = \frac{n_{bit}}{V_{com}} n_e n_p n_I \tag{15}$$

where n_{bit} is the number of bits of elements in the information sharing matrix, $n_{bit} = 16$ bit; V_{com} is the speed of data communication; n_e is the number of information exchanges, we set $n_e = 4$; n_p is the number of elements in the information sharing matrix, $n_p = 8$; n_I is the number of iterations.

The communication delay results are shown in Table 4.

TABLE 4. Communication delay

Communication mode	Telecommunication lines	V_{com}	1.2 s	1.3 s	1.4 s
Ethernet	Optical fiber	10 Mbit/s	0.0003 s	0.0005 s	0.0008 s

The communication delay is in microsecond level, which has little impact on the real-time control of this paper. Even when the communication speed is as low as 1 Mbit/s, the communication delay value can generally be ignored. For the actual online FM communication in the microgrid, this value level is also acceptable, which effectively shows that the control strategy proposed in this paper has good communication performance.

5. **Conclusions.** Based on the problem that the system frequency deviates from the rated value under the condition of load switching by virtual synchronous machine during the operation of independent microgrid, a multi-agent distributed frequency control strategy based on Q-Reinforcement learning is proposed by this paper to realize the secondary frequency modulation control of microgrid system. After simulation and verification, the results are as follows.

1) The multi-agent distributed frequency control strategy based on Q-Reinforcement learning uses the overall situation reward as local feedback. According to the changes of network topology, it is optimized and adjusted to achieve fast convergence of rewards and optimal results, and achieve good frequency modulation effect.

2) The control strategy overcomes the disadvantages of centralized control strategy, such as multiple communication sets and large line investment. The strategy realizes global coordination control, frequency recovery and voltage regulation through adjacent communication, and ensures the flexibility and stability of distributed control of microgrid.

In the future, the research will further consider the enrichment of multi-agent and specify local rewards for agents. At the same time, the economic and environmental benefits are taken into consideration in the process of frequency modulation to enhance the applicability of the proposed scheme, solve various conditions that may exist in the operation of micro-grid, and optimize the control strategy.

Acknowledgment. This work is partially supported by the project of National Natural Science Foundation of China (Grant No. 61903158). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] Y. Zhang, H. Dai, X. Wu, R. Chen and N. Zhang, Development trends and key issues of China's integrated energy services, *Electric Power*, vol.54, no.2, pp.1-10, 2021.

[2] S. Y. M. Mousavi, A. Jalilian, M. Savaghebi and J. M. Guerrero, Autonomous control of current- and voltage-controlled DG interface inverters for reactive power sharing and harmonics compensation in islanded microgrids, *IEEE Transactions on Power Electronics*, vol.33, no.11, pp.9375-9386, 2018.

[3] R. H. Lasseter, Z. Chen and D. Pattabiraman, Grid-forming inverters: A critical asset for the power grid, *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol.8, no.2, 925925, 2020.

[4] Z. Li, K. Chan, J. Hu et al., Adaptive droop control using adaptive virtual impedance for microgrids with variable PV outputs and load demands, *IEEE Transactions on Industrial Electronics*, vol.68, no.10, pp.9630-9640, 2021.

[5] X. Hou, Y. Sun, X. Zhang et al., Improvement of frequency regulation in VSG-based AC microgrid via adaptive virtual inertia, *IEEE Transactions on Power Electronics*, vol.35, no.2, pp.1589-1602, 2020.

[6] X. Yang, *Research on Microgrid Inverter and Coordinated Control Strategies of Multi-Inverters*, Ph.D. Thesis, Hefei University of Technology, 2011.

[7] J. Dong, Z. Wang, G. Zhu and J. Bao, Secondary frequency regulation method for droop inverters in island operation, *Electric Power Automation Equipment*, vol.42, no.5, pp.40-46, 2022.

[8] J. Ma, Y. Li, J. Li, G. Li and P. He, A secondary control strategy for frequency and voltage recovery and power sharing control method in island microgrid, *Modern Electric Power*, vol.39, no.3, pp.363-370, 2022.

[9] L. Sun, X. Yang, L. Sun, Y. Liu and C. Liu, Frequent deviation-free control for microgrid multi-inverters based on improving a virtual synchronous generator, *Power System Protection and Control*, vol.49, no.11, pp.18-27, 2021.

[10] Z. Wang, W. Wu and B. Zhang, A fully distributed power dispatch method for fast frequency recovery and minimal generation cost in autonomous microgrids, *IEEE Transactions on Smart Grid*, vol.7, no.1, pp.19-31, 2015.

[11] L. Y. Lu and C. C. Chu, Consensus-based secondary frequency and voltage droop control of virtual synchronous generators for isolated AC microgrids, *IEEE Journal on Emerging & Selected Topics in Circuits & Systems*, vol.5, no.3, pp.443-455, 2015.

[12] R. P. S. Chandrasena, F. Shahnia, A. Ghosh et al., Secondary control in microgrids for dynamic power sharing and voltage/frequency adjustment, *Power Engineering Conference*, Perth, WA, Australia, pp.5363-5374, 2014.

[13] B. Li, L. Zhou, Y. Cao et al., Improved power decoupling control strategy based on virtual synchronous generator, *IET Power Electronics*, vol.10, no.4, pp.462-470, 2017.

[14] W. Wang, L. Jiang, X. Yu et al., A parameter alternating VSG controller of VSC-MTDC systems for low frequency oscillation damping, *IEEE Transactions on Power Systems*, vol.35, no.6, pp.4609-4621, 2020.

[15] B. Qin and Y. Xu, Modal analysis of multi-virtual synchronous machine grid-connected power-frequency oscillation, *Proc. of the CSEE*, vol.41, no.19, pp.6570-6581, 2021.

[16] H. Miao, Y. Yang, F. Mei et al., A seamless mode switching control strategy in virtual synchronous generator, *Acta Energiae Solaris Sinica*, vol.41, no.19, pp.6570-6581, 2021.

[17] M. Zhao, H. Yin, Y. Xue et al., Coordinated damping control design for power system with multiple virtual synchronous generators based on prony method, *IEEE Open Access Journal of Power and Energy*, pp.316-328, 2021.

[18] D. Xu, Z. Cheng, W. Yang and W. Zhang, Robust nonlinear control for virtual synchronous generator based on exact feedback linearization, *International Journal of Innovative Computing, Information and Control*, vol.18, no.4, pp.1133-1145, 2022.
[19] Q. Teng, D. Xu and W. Yang, An adaptive fractional-order sliding mode control for virtual synchronous generator in microgrid, *ICIC Express Letters*, vol.16, no.9, pp.983-991, 2022.
[20] J. Shen, W. Liu, H. Li, N. Li et al., Reinforcement learning based distributed secondary optimal control for multiple microgrids, *Automation of Electric Power Systems*, vol.44, no.5, pp.198-206, 2020.
[21] M. Mao, Y. Xi, L. Zhang et al., On-line adaptive control of microgrid secondary frequency based on Q-learning, *Automation of Electric Power Systems*, vol.39, no.20, pp.26-31, 2015.
[22] H. Zhang, C. Mou, M. Zhao and T. Yao, Frequency coordination control strategy of multiple photovoltaic-battery virtual synchronous generators based on reinforcement learning, *Electric Driver*, vol.51, no.19, pp.36-42, 2021.

Author Biography

Jiong Jiang received his B.Sc. degree in electrical engineering from Beihua University, Jilin, China, in 2020. He is currently pursuing his M.Sc. degree in electrical engineering in Jiangnan University. His main research interest is new energy power generation technology.

Yuhao Lu received his B.Sc. degree in electrical engineering and automation from North China Electric Power University, 2016; the Master degree in Jiangnan University, 2022. He is currently a full-time electrical engineer at Power China Hubei Electric Engineering Co., Ltd. His research interests include new energy source technology and energy storage technology.

Tinglong Pan received his B.Eng. degree in industrial automation from China University of Mining and Technology, Xuzhou, China, in 1999, and the Ph.D. degree in power electronics and power drive from China University of Mining and Technology, Xuzhou, China, in 2004. He is currently a professor at Jiangnan University, where his research interests include microgrid control technology, power conversion technology, power drive system and its intelligent control technology.

Weilin Yang received his B.Eng. degree in machine design & manufacture and their automation from University of Science and Technology of China, Hefei, China, in 2009, and the Ph.D. degree in mechanical engineering from City University of Hong Kong, Hong Kong, in 2013. He was a postdoctoral researcher at Masdar Institute of Science and Technology (now Khalifa University), Abu Dhabi, UAE, 2013-2016. He was a research engineer of General Electric (GE) Global Research, Shanghai, 2016-2017. He joined Jiangnan University in July 2017, where he is currently an associate professor. His research interests include modeling and control of energy systems, robust model predictive control, and dada-driven control.

Yachao Wang received a doctorate in mechanical engineering from Beijing University of Technology, Beijing, in 2018. He is a senior engineer, and is currently working in the State Grid Jibei Marketing Service Center (Metrology Center). His research interest is power system and its automation.

Zhenlan Dou received a doctorate in electrical engineering from Shanghai Jiao Tong University, Shanghai, in 2013. She is currently working in State Grid Shanghai Integrated Energy Service Co., Ltd. Her research interest is new energy generation technology.