Energy Proceedings
Vol 39, 2024
ISSN 2004-2965

# Energy-Saving and Thermal Comfort Control of Electric Vehicle Air Conditioning Systems with Deep Reinforcement Learning

Shuai Dai[1,2], Kuining Li[1,2*], Jiangyan Liu[1,2], Yi Xie[3]

1 Key Laboratory of Low-Grade Energy Utilization Technologies and Systems, Chongqing University, Ministry of Education, Chongqing, 400044, China

2 School of Energy and Power Engineering, Chongqing University, Chongqing, 400044, China

3 College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing, 400044, China
(*Kuining Li: leekn@cqu.edu.cn)

## ABSTRACT

This study employs deep reinforcement learning algorithms, including Deep Q-Network, Deep Deterministic Policy Gradient, Twin Delayed Deep Deterministic Policy Gradient, and Soft Actor-Critic, to control the air conditioning system of electric vehicles to improve thermal comfort and reduce energy consumption. Additionally, random adjustments to environmental temperature and solar radiation intensity during the training process are made to enhance the algorithms' applicability. The results demonstrate that these algorithms significantly reduce energy consumption while maintaining thermal comfort. Notably, the Deep Deterministic Policy Gradient algorithm achieves an impressive 37.6% reduction in energy consumption. Comparative analysis among the algorithms reveals that Deep Q-Network, Deep Deterministic Policy Gradient, and Twin Delayed Deep Deterministic Policy Gradient exhibit relatively stable control behaviors. In contrast, the Soft Actor-Critic algorithm's compressor control curve exhibits more significant fluctuations, potentially leading to mechanical wear. Deep Q-Network, Deep Deterministic Policy Gradient, and Twin Delayed Deep Deterministic Policy Gradient algorithms consistently demonstrate effective thermal comfort control and energy-saving performance in various operating conditions.

**Keywords:** deep reinforcement learning, electric vehicle, air conditioning system, thermal comfort

## NONMENCLATURE

| *Abbreviations* | |
|---|---|
| DQN | Deep Q-Network |
| DDPG | Deep Deterministic Policy Gradient |
| TD3 | Twin Delayed Deep Deterministic Policy Gradient |
| SAC | Soft Actor-Critic |
| PMV | Predicted Mean Vote |

## 1. INTRODUCTION

With the increasing severity of global energy and environmental issues, the electric vehicle industry is experiencing rapid growth. However, the challenge of electric vehicle range has persisted, primarily due to the limited energy density of electric vehicle power batteries, typically ranging from 110 to 160 watt-hours per kilogram [1]. Under the current battery energy density conditions, devising rational control strategies to reduce energy consumption across various systems of electric vehicles has become a practical approach to improving their range. The electric vehicle air conditioning system is considered one of the highest energy consumers among these systems. Studies indicate that activating the electric vehicle air conditioning system significantly affects the vehicle's range, sometimes causing a reduction of over 30% [2]. Electric vehicle air conditioning systems rely solely on battery power and cannot generate additional power like internal combustion engines in conventional vehicles, so designing efficient control strategies for electric vehicle air conditioning systems has become paramount. Research indicates that activating the air conditioning system can significantly impact the driving range of electric vehicles. In certain conditions, it can even reduce over 30% in driving range [2]. This is primarily because the electric vehicle's air conditioning system relies solely

# This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

on the battery for energy input, unlike traditional combustion engine vehicles that use the engine to power the air conditioning. Hence, designing an efficient control strategy for electric vehicle air conditioning systems becomes paramount.

Electric vehicle air conditioning systems typically employ on/off or PID control strategies. While the on/off strategy is straightforward, the frequent start-stop cycles of the compressor lead to increased energy consumption in the air conditioning system and result in frequent fluctuations in cabin temperature [3]. PID control strategies are widely used in air conditioning system control due to their simplicity and reliability, adaptability, and robustness [4]. However, setting PID parameters requires expertise or engineering experience, making their control often suboptimal and challenging to meet cabin thermal comfort and low energy consumption requirements simultaneously. Model-based optimization control methods, such as dynamic programming [5] and model predictive control [6], have been extensively researched because they can achieve optimal results while satisfying air conditioning system constraints. However, these methods require highly accurate and simplified dynamic models of the air conditioning system and have high hardware computational demands, limiting their practical applicability.

Reinforcement Learning (RL) offers an improved control methodology for electric vehicle air conditioning systems. It is oriented towards system control objectives, continuously enhancing control strategies through interactions with the controlled system to achieve optimal control performance. Researchers such as Kasbi [7] and Brusey [8] have demonstrated better results in temperature control and energy consumption for electric vehicle air conditioning systems using the SARSA(State-Action-Reward-State-Action) algorithm compared to traditional control methods. However, traditional RL methods often require the discretization of observations and control values, making them less suitable for continuous air conditioning systems. Choi et al. [9] used DQN to control compressors and blowers, and tested the generalization ability of DQN algorithm. While Joo et al. [10] utilized the SAC algorithm to regulate the expansion valve and compressor, achieving an energy consumption as low as 53% compared to PID control during cabin cooling. Nevertheless, research that comprehensively considers external environmental changes and the trade-off between thermal comfort and energy consumption remains relatively limited. Comparative studies of different deep reinforcement learning algorithms regarding rewards, thermal comfort, and energy consumption are exceedingly scarce.

This paper introduces a deep reinforcement learning control approach that considers environmental factors. By incorporating environmental temperature and solar radiation intensity into observations and defining thermal comfort metric PMV and energy consumption in the reward function, we iteratively train the model to obtain the optimal control strategy. Furthermore, we compare the control performance of four deep reinforcement learning algorithms, DQN, DDPG, TD3, and SAC, and discuss their reward convergence, thermal comfort, and energy efficiency. This study aims to facilitate the application of deep reinforcement learning algorithms in electric vehicle air conditioning system control.

## 2. AIR CONDITIONING SYSTEM OF EV AND CONTROL PROBLEM DESCRIPTION

### *2.1 Air conditioning system*

As shown in Fig. 1, the electric vehicle air conditioning system is a complex, nonlinear thermal system. To investigate control strategies, this paper requires establishing a simplified yet sufficiently accurate dynamic air conditioning model. The moving-boundary lumped-parameter modelling method is widely employed in the dynamic model of the air conditioning system's heat exchanger because it can provide concise and precise results [11]. Therefore, this paper adopts the moving boundary method to construct the dynamic air conditioning model. The system consists of a compressor, condenser, expansion valve, and evaporator.

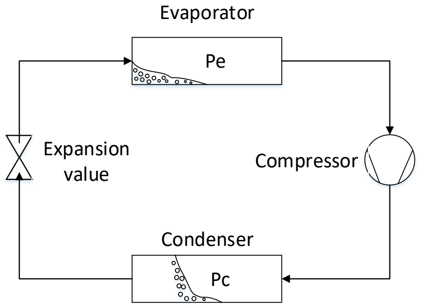


Fig. 1 Air conditioning system

The dynamic process of the compressor can be represented as follows:

$$\dot{m}_{comp} = N_{comp} V_d \rho_{ref} \eta_{vol} \tag{1}$$

$$h_{oc} = \left(h_{is} - h_{ic}\right) / \eta_a + h_{ic} \tag{2}$$

$N_{comp}$ is the mass flow rate of the refrigerant, and $N_{comp}$ is the compressor speed. $V_d$ is the displacement. $\rho_{ref}$ is the refrigerant density, and $\eta_{vol}$ is the volumetric efficiency. Moreover $h_{is}$ and $h_{ic}$ are the enthalpy values at the compressor's inlet and outlet. $\eta_a$ is the isentropic efficiency.

The dynamic process of the expansion valve can be represented as follows:

$$\dot{m}_v = C_v a_v \sqrt{\rho_v \left(P_c - P_e\right)} \tag{3}$$

$C_v$ is the flow coefficient, and $a_v$ is the opening degree of the expansion valve. $\rho_v$ is the refrigerant density. $P_c$ is the condensing pressure, and $P_e$ is the evaporating pressure.

R134a undergoes three phases in the evaporator: liquid, two-phase liquid-gas, and superheated vapor. The heat exchange process mainly occurs in the evaporation region, including the liquid and two-phase liquid-gas phases. According to conservation of energy, the length of the evaporation region can be expressed as:

$$\rho_{le} h_{lge} A_e \left(1 - \overline{\gamma_e}\right) \frac{dl_e}{dt} = \dot{m}_v \left(h_{ge} - h_{ie}\right) - a_{ie} \pi D_{ie} l_e \left(T_{we} - T_{re}\right) \tag{4}$$

The first term on the right-hand side of the equation is the enthalpy change in the two-phase region of the refrigerant, where $\dot{m}_v$ is the mass flow rate of the refrigerant, $h_{ge}$ and $h_{ie}$ are the enthalpy values of the refrigerant vapor state and the inlet refrigerant enthalpy of the evaporator, respectively. The next two terms on the right-hand side are used to describe the heat exchange between the refrigerant and the inner wall surface of the evaporator. Here, $a_{ie}$ is the heat transfer coefficient, $D_{ie}$ is the inner diameter of the flattened tube, $l_e$ is the length of the two-phase region, $T_{we}$ is the equivalent temperature of the tube wall and fins, and $T_{re}$ is the saturation temperature of the refrigerant at the evaporator pressure.

The dynamic variation of the evaporating pressure can be expressed as follows:

$$A_e L_e \frac{d\rho_{ge}}{dP_e} \frac{dP_e}{dt} = \dot{m}_v \frac{(h_{ie} - h_{le})}{h_{\mathrm{lg}e}} - \dot{m}_{comp} - \frac{a_{ie} \pi D_{ie} l_e \left(T_{we} - T_{re}\right)}{h_{\mathrm{lg}e}} \tag{5}$$

$\rho_{ge}$ is the density of refrigerant vapor, and $h_{le}$ is the enthalpy of the refrigerant liquid.

The evaporator surface temperature can be expressed as follows:

$$\left(C_p m\right)_{we} \frac{dT_{we}}{dt} = a_{oe} A_{oe} \left(T_{ae} - T_{we}\right) - a_{ie} \pi D_{ie} l_e \left(T_{we} - T_{re}\right) \tag{6}$$

$C_p$ is the specific heat capacity of the evaporator material, and $m$ is the mass of the evaporator. $a_{oe}$ is the heat transfer coefficient between the evaporator and the ambient air. $A_{oe}$ corresponds to the frontal area of the evaporator, and $T_{ae}$ is the ambient air temperature around the evaporator.

The modeling method for the condenser is the same as that for the evaporator:

$$\begin{aligned} \left(C_p m\right)_{wc} \frac{dT_{wc}}{dt} &= a_{ic} \pi D_{ic} l_c \left(T_{rc} - T_{wc}\right) - a_{oc} A_{oc} \left(T_{wc} - T_{ac}\right) \\ &+ a_{ish} \pi D_{ic} \left(L_c - l_c\right) \left(\frac{T_\pi + T_{ic}}{2} - T_{wc}\right) \end{aligned} \tag{7}$$

Throughout the refrigeration cycle, the total mass of the refrigerant remains constant, and the following equation is used to describe refrigerant mass conservation:

$$\begin{aligned} m_{total} - m_{pipe} = A_e &\left[\rho_t l_e \left(1 - \gamma_e\right) + \rho_{gs} l_e \gamma_e + \rho_{gt} \left(L_e - l_e\right)\right] \\ &+ A_c [\rho_c l_e \left(1 - \gamma_c\right) + \rho_{gc} l_e \gamma_c + \rho_{ge} \left(L_t - l_c\right)] \end{aligned} \tag{8}$$

The left side of the equation is used to describe the total mass inside the evaporator and condenser, where $m_{total}$ is the total refrigerant mass, and $m_{pipe}$ is the refrigerant mass outside the heat exchanger.

### *2.2 Cabin thermal model*

To balance computational complexity with model accuracy, a lumped-parameter thermal model for the passenger compartment is established in this study.

The temperature variation in the cabin is described by the following equation:

$$\frac{dT_c}{dt} = \frac{Q_{cab} + Q_{AC}}{M_a C_a} \tag{9}$$

$Q_{AC}$ is the heat exchange between the passenger compartment air and the evaporator, $M_a$ is the mass of air inside the cabin, $C_a$ is the specific heat capacity of air, $Q_{cab}$ is the total heat load in the cabin, including convective heat load $Q_{conv}$, solar radiation heat load $Q_{solar}$, ventilation heat load $Q_{vent}$, electrical equipment heat load $Q_e$, and human body heat load $Q_{human}$.

The heat exchange between the cabin air and the evaporator:

$$Q_{AC} = h_{ec} A_e (T_{ec} - T_c) \tag{10}$$

$h_{ec}$ is the heat transfer coefficient between the evaporator surface and the air, $A_e$ is the convective heat transfer area of the evaporator, and $T_{ec}$ is the evaporator surface temperature.

The convective heat exchange between the interior walls and the cabin:

$$Q_{conv} = \sum h_i A_i (T_{si} - T_c) \tag{11}$$

$T_{si}$ is the temperature of the cabin enclosure structure, which includes the front windshield, the front end of the car, the roof, the rear windshield, the side panels, side windows, and seats. $h_i$ is the convective heat transfer coefficient between the interior air and the cabin enclosure structure, and $A_i$ is the surface area of the enclosure structure.

The solar radiation exchange:

$$Q_{solar} = \sum I\eta S_i \tag{12}$$

$I$ is the solar radiation intensity, $\eta$ is the coefficient of solar light passing through the windows, and $S_i$ is the effective area of the windows perpendicular to the direction of sunlight. For the sake of convenience in the study, $Q_{vent}$, $Q_e$, and $Q_{human}$ are considered as constant values.

### *2.3 Air conditioning control objectives*

The air conditioning control objectives consist of two aspects: firstly, to maintain a comfortable environment within the cabin, and secondly, to reduce the energy consumption of the air conditioning system. In this study, cabin thermal comfort is evaluated using the Predicted Mean Vote (PMV).

The control objectives are represented by the following equation:

$$f = \min \int_0^{t_{final}} \{ \left| PMV_{cabin}(t) - PMV_{target} \right| + \lambda \left[ P_{comp}(t) + P_{fan}(t) \right] \} dt \tag{13}$$

$PMV_{cabin}(t)$ is the thermal comfort evaluation of the cabin at time $t$, $PMV_{Target}$ is the target thermal comfort evaluation of the cabin, $P_{comp}$ is the compressor power, $P_{fan}$ is the fan power, $\lambda$ is the energy consumption weighting factor used to balance the temperature control and energy consumption objectives, and $t_{final}$ is the duration of the vehicle test cycle.

## 3. DEEP REINFORCEMENT LEARNING ALGORITHM

### *3.1 Electric vehicle air conditioning MDP model*

The reinforcement learning task for electric vehicle air conditioning control can be constructed as a Markov Decision Process (MDP), which can be described using a quadruple {S, A, R, P}, where the S is the state space composed of various state variables of the electric vehicle air conditioning system, the A is the action space consisting of control parameters, the R is the reward obtained from the electric vehicle air conditioning system after taking action, and the P is the transition probabilities between different states of the electric vehicle air conditioning system.

The parameters of the Markov Decision Process (MDP) model for the electric vehicle air conditioning system are as follows:

1) State representation

The selected state variables include cabin temperature, ambient temperature, solar radiation intensity, cabin temperature rate of change, and evaporator surface temperature. Among these, cabin and evaporator surface temperatures represent the heat exchange status between the passenger cabin and the evaporator. The rate of change of passenger cabin temperature is used to describe the dynamic variations in the passenger cabin's thermal environment, while ambient temperature and solar radiation intensity are used to describe the environmental thermal conditions.

2) Action representation

The action should align with the actual control variables of the electric vehicle air conditioning system. This study selects compressor speed and blower fan speed as actions. Since the DQN algorithm can only handle discrete action spaces, the action space is discretized with equal intervals, with a difference of 500 r/min.

3) Transition probability

The transition probability P reflects the dynamic characteristics of the electric vehicle air conditioning system. P is unknown for the electric vehicle air conditioning system, and in this paper, Monte Carlo methods are used to obtain an unbiased estimate of P.

4) Reward function

Based on the optimization objective, the reward function is defined as follows:

$$r(t) = \begin{cases} \alpha \left| PMV_{cabin}(t) - PMV_{target} \right| & \text{if } \left| PMV_{cabin}(t) - PMV_{target} \right| \geq 0.5 \\ 10 & \text{if } \left| PMV_{cabin}(t) - PMV_{target} \right| < 0.5 \end{cases} \tag{14}$$
$$+ \beta [P_{comp}(t) + P_{fan}(t)]$$

$\alpha$ and $\beta$ are weighting coefficients, where $E_{use}(t)$ is the power consumption of the air conditioning system, including the compressor power $P_{comp}(t)$ and the blower fan power $P_{fan}(t)$. In this paper, $PMV_{target}$ is set to 0. The first part of the equation calculates the deviation between the actual PMV in the cabin and the target PMV, while the second part calculates the energy consumption of the air conditioning system.

### *3.2 Deep reinforcement learning algorithm*

#### 3.2.1 DQN

Deep Q-learning is a reinforcement learning algorithm based on deep learning, aimed at approximating the action-value function (Q-value) within Markov decision processes [12]. DQN utilizes deep neural networks to approximate the Q-value function and employs techniques like experience replay and fixed target networks to enhance training stability and convergence. The fundamental concept behind DQN involves storing the agent's experience samples in an experience replay buffer and randomly sampling from it during training to address sample correlation issues. The update equation for the Q-network is as follows:

$$L(w) = \frac{1}{N} \sum_i \frac{1}{2} [r + \gamma \max Q(s_{t+1}, a_{t+1}, w') - Q(s_t, a_t, w)]^2 \tag{15}$$

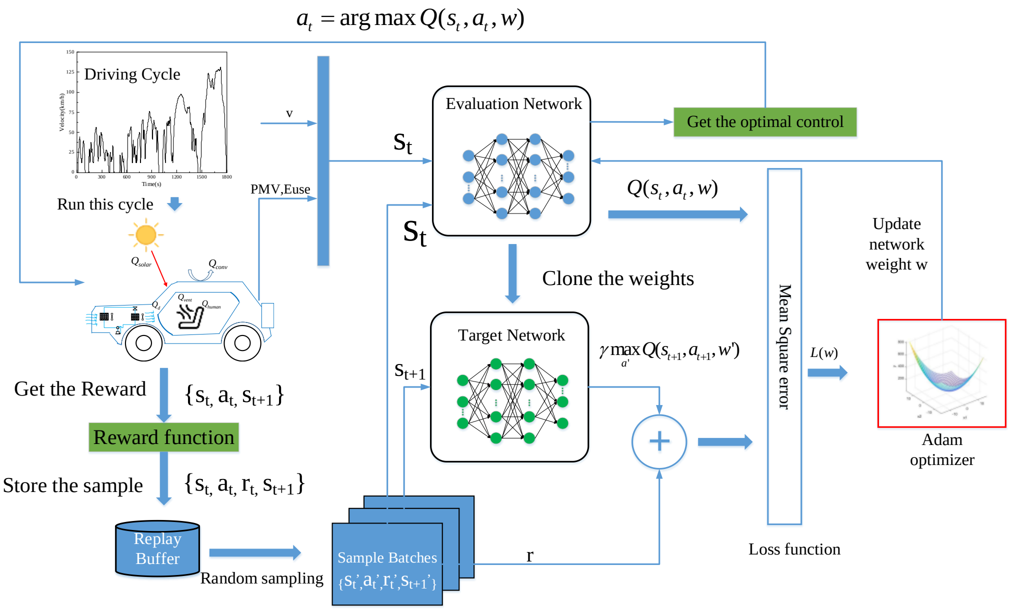


Fig. 2 Illustrates the DQN training architecture

$$w \leftarrow w - \alpha \bullet \frac{\partial L(w)}{\partial w} \tag{16}$$

$$w' \leftarrow \tau \bullet w + (1-\tau) w' \tag{17}$$

$Q(s_t, a_t, w)$ is the value network, where $w$ denotes the parameters of the value evaluation network. $\alpha$ stands for the learning rate, while $Q(s_{t+1}, a_{t+1}, w')$ is the target network, with $w'$ being the parameters of the target network. $\tau$ is the soft update rate.

### 3.2.2 DDPG

Deep Deterministic Policy Gradient is a deep reinforcement learning algorithm designed for continuous action space [13]. DDPG learns both deterministic policy networks and action value networks. The policy network outputs actions directly, while the value network evaluates the quality of the policy. DDPG uses empirical replay and target network technology to enhance the stability and convergence of the algorithm. It is good at solving continuous control problems and shows strong generalization ability. The following formula represents DDPG network update:

$$L(w) = \frac{1}{N}\sum_i \left[\left(r + \gamma Q\left(s_{t+1}, \mu'\left(s_{t+1}\right)\right) - Q\left(s_t, a_t\right)\right)\right]^2 \tag{18}$$

$$\nabla_{\vartheta^\mu} J = \frac{1}{N}\sum_i \left[\nabla_a Q\left(s_t, \mu\left(s_t\right)\right) \nabla_{\vartheta^\mu} \mu\left(s_t | \vartheta^\mu\right)\right] \tag{19}$$

$$\vartheta' \leftarrow \varepsilon\vartheta + (1-\varepsilon)\vartheta', \vartheta^{\mu'} \leftarrow \varepsilon\vartheta^\mu + (1-\varepsilon)\vartheta^{\mu'} \tag{20}$$

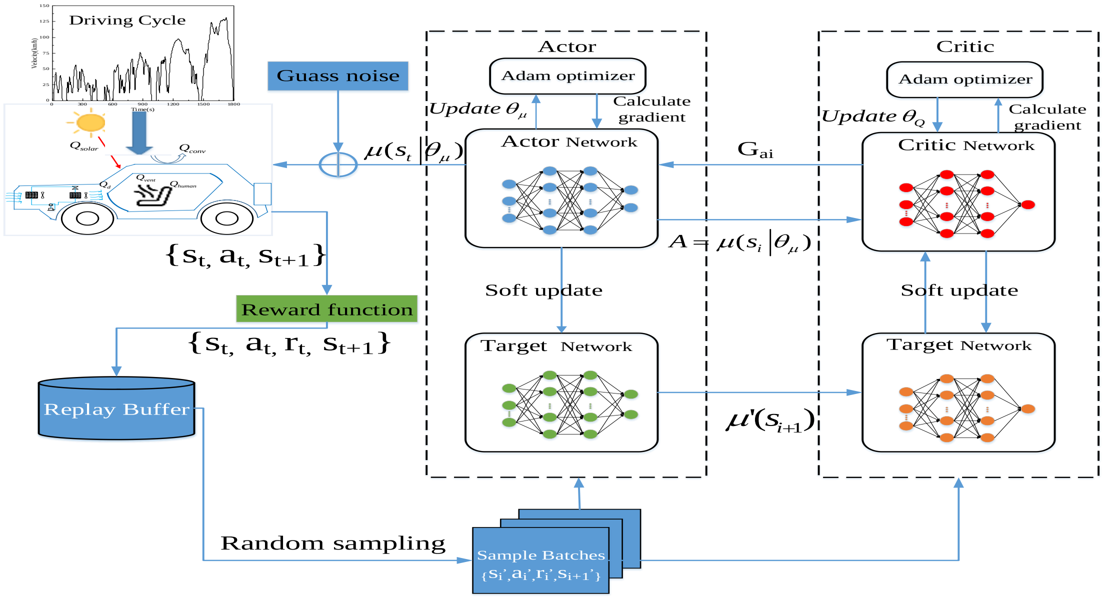


Fig. 3 Illustrates the DDPG training architecture

### 3.2.3 TD3

Twin Delayed Deep Deterministic Policy Gradient is an improved version of DDPG [14]. TD3 enhances algorithm stability by introducing twin Critic networks and target policy smoothing regularization. Specifically, TD3 employs two independent value function networks to mitigate overestimation, and the following equation represents the target value estimation:

$$y = r + \gamma \min_{i=1,2} Q_i^{'}\left(s_{t+1}, \mu'\left(s_{t+1}\right)\right) \tag{21}$$

Additionally, target policy smoothing regularization is applied, which involves introducing perturbations to the action in the next state:

$$a_{t+1} = \mu'\left(s_{t+1}\right) + \varepsilon, \varepsilon \sim clip\left(N\left(0, \sigma^2\right), -c, c\right) \tag{22}$$

The loss function is computed as follows:

$$L_i = \left(Q_i\left(s_t, a_t\right) - y\right)^2 \tag{23}$$

The stability of the algorithm is further improved by employing the delayed update of the action value network. In other words, the action value network is updated after multiple updates of the Critic network.

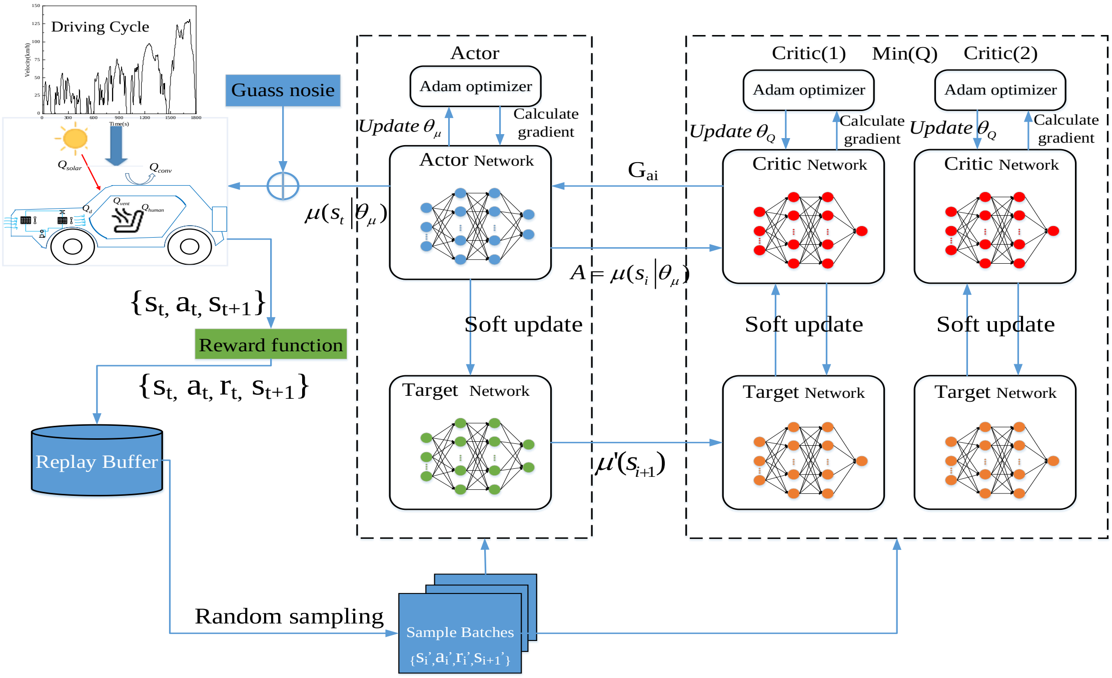


Fig. 4 Illustrates the TD3 training architecture

### 3.2.4 SAC

Soft Actor-Critic is a deep reinforcement learning algorithm based on maximum entropy theory [15]. SAC introduces entropy regularization to strike a balance between reward return and entropy. Entropy measures the randomness of the policy; increasing entropy implies greater policy randomness, which encourages more exploration and thus accelerates learning. This strategy considers the reward signal and emphasizes exploration and maintaining policy diversity. The computation of the target Q-value in SAC is as follows:

$$\begin{aligned} y = r + \gamma(&\min_{i=1,2} Q_{\mu,i}(s_{t+1}, \pi(\bullet|s_{t+1})) \\ &-\alpha \log(\pi(\pi(\bullet|s_{t+1})|s_{t+1}))) \end{aligned} \tag{24}$$

$-\alpha \log(\pi(\pi(\bullet|s_{t+1})|s_{t+1}))$ is the entropy of the policy, where $\alpha$ denotes the weight of the entropy. SAC also employs two value function networks to evaluate the policy's value and uses maximum entropy regularization to enhance policy exploration, thus improving the algorithm's exploration efficiency.

## 4. SIMULATION EXPERIMENT

### 4.1 *Parameter settings*

To enhance the generalization ability of deep reinforcement learning algorithms, the following parameter settings were used for the electric vehicle air conditioning system simulation experiments in this paper. The training environment temperature for the electric vehicle air conditioning system follows a uniform distribution in the range of [25°C, 45°C], and the solar radiation intensity follows a uniform distribution in the range of [500W/m2, 1000W/m2]. The initial temperature of the passenger compartment is set to 50°C. During training, the Worldwide Harmonized Light vehicles Test Cycle (WLTC) is used as the training scenario, and 1000 training epochs are conducted. The sampling time is set to

10 seconds, and the action outputs are averaged over a 10-second time window to avoid abrupt changes in action outputs. To ensure a fair comparison of the performance of different deep reinforcement learning algorithms, the same hyper parameters were used for all four deep reinforcement learning algorithms. The hyper parameter settings are shown in the table below:

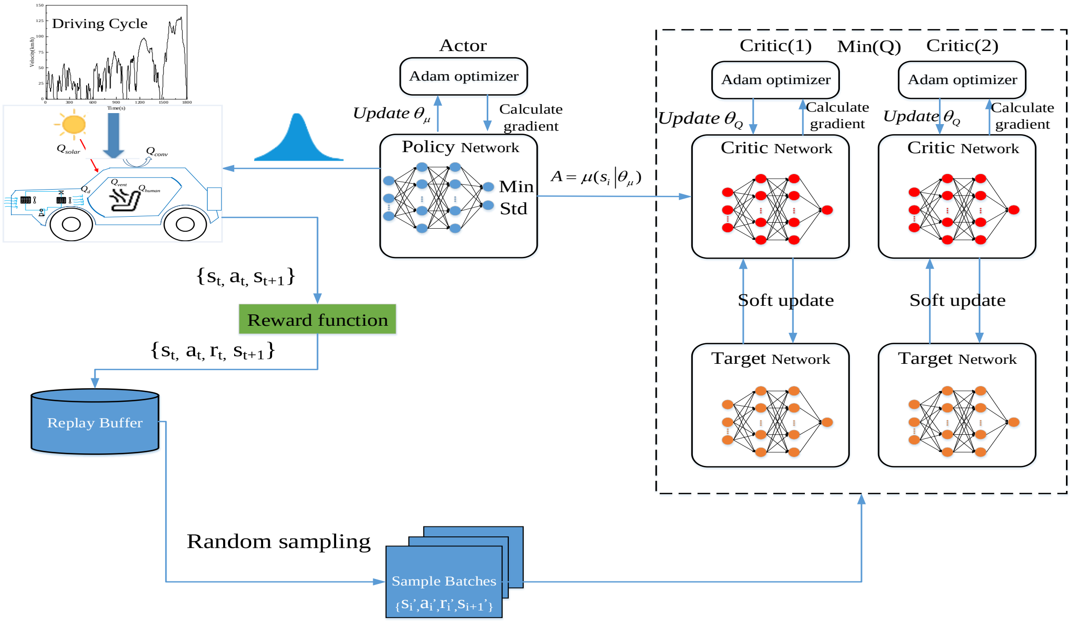


Fig. 5 Illustrates the SAC training architecture

Table 1 Hyper parameter settings

| Hyper parameter | Value |
|---|---|
| Discount factor | 0.99 |
| Learning rate | 0.0001 |
| Number of neurons | 256 |
| Minibatch size | 64 |
| Memory pool size | $10^6$ |

## 4.2 *Experimental results*

### 4.2.1 Convergence results

Fig. 6 displays four deep reinforcement learning algorithms' sliding average reward convergence trajectories during the training process, with a sliding window length set to 20. Several key observations can be made by examining the curves:

Firstly, in terms of convergence speed, the DQN algorithm exhibits the fastest convergence rate. It reaches a stable average reward after only 156 training rounds. In contrast, the SAC algorithm demonstrates the slowest training speed, requiring 583 rounds of training to achieve convergence.

Secondly, regarding convergence rewards, the DDPG algorithm achieves the highest average reward. Compared to the DQN, TD3, and SAC algorithms, DDPG experiences an increase in average rewards by 15.6%, 8.0%, and 44.8%, respectively.

Lastly, concerning training stability, the DDPG algorithm shows the smoothest variation in average rewards after convergence. This suggests that the DDPG algorithm exhibits excellent adaptability to different external environments.

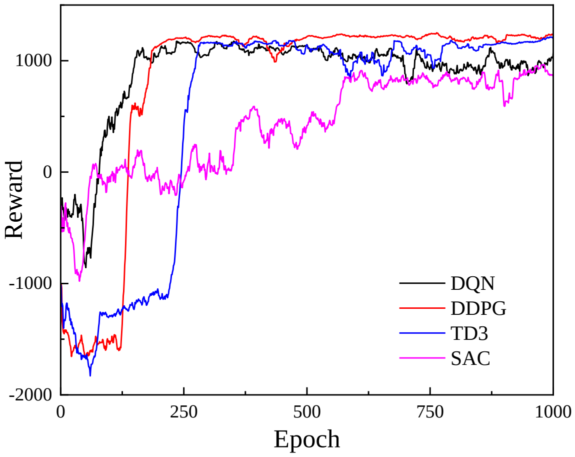


Fig. 6 Average reward

Table 2 Convergence results

| Agent | Convergence epochs | Convergence reward | Reward STDEVA |
|---|---|---|---|
| DQN | 156 | 1037.3 | 80.5 |
| DDPG | 188 | 1198.8 | 38.3 |
| TD3 | 290 | 1110.1 | 72.6 |
| SAC | 583 | 827.7 | 63.9 |

### 4.2.2 Energy performance of deep reinforcement learning

The energy consumption of the electric vehicle air conditioning system directly affects its driving range. The table below shows the energy performance of on/off control and different deep reinforcement learning algorithms at different environmental temperatures (solar radiation intensity is 750W/m2). The control logic for on/off control is as follows: when the passenger compartment PMV is more significant than 0.5, both the compressor speed and fan speed are set to the maximum; when the passenger compartment PMV is less than 0.5, the compressor and fan speeds are set to 0. When the passenger compartment PMV is between -0.5 and 0.5, the compressor and fan speeds remain unchanged from the previous time step.

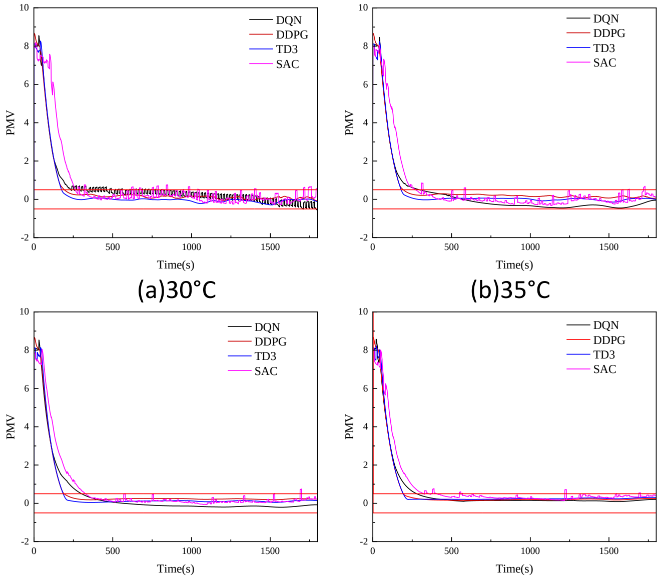
Table 3 illustrates that control strategies based on deep reinforcement learning result in varying degrees of energy improvement compared to on/off control. Among them, the control strategy based on DDPG shows better energy performance. Compared to the traditional on/off control strategy at different environmental temperatures, the energy consumption of the DDPG-based control strategy is reduced by 8.4%, 38.2%, 53.2%, and 50.7%, with an average reduction of 37.6%. This indicates that control strategies based on deep reinforcement learning have significant energy-saving potential.

Table 3 Energy Consumption of the Air Conditioning System

| Control strategy | Energy consumption (kW·h) | | | |
|---|---|---|---|---|
| | 30℃ | 35℃ | 40℃ | 45℃ |
| on/off | 0.83 | 1.02 | 1.24 | 1.50 |
| DQN | 0.90 | 0.67 | 0.63 | 1.47 |
| DDPG | 0.76 | 0.63 | 0.58 | 0.74 |
| TD3 | 0.83 | 0.69 | 0.63 | 0.79 |
| SAC | 0.83 | 0.84 | 0.72 | 0.79 |

### 4.2.3 Performance of Deep Reinforcement Learning in Thermal Comfort Control

Fig. 7 shows the performance of different deep reinforcement learning algorithms in passenger compartment thermal comfort control. It can be seen that under different external environmental temperatures, all deep reinforcement learning algorithms can transition the passenger compartment from the initial state to a thermally comfortable state in approximately 250 seconds. At an environmental temperature of 30°C, both SAC and DQN algorithms experienced thermal comfort violations after transitioning to a thermally comfortable state, while DDPG and TD3 thermal comfort control remained relatively stable.



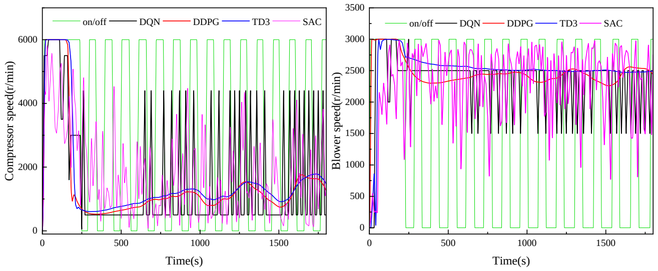
(a)30°C (b)35°C

(c) 40°C (d)45°C

Fig. 7 Cabin thermal comfort control

### 4.2.4 Performance of deep reinforcement learning in compressor/fan Control

Fig. 8 shows the compressor/blower control performance of each deep reinforcement learning control strategy when the ambient temperature is 40℃, and the solar radiation intensity is 750W/m2. It can be seen that in the first 250s, in order to achieve rapid cooling of the crew cabin, each control strategy adopts a higher compressor/blower speed. Due to the random strategy adopted by SAC, the compressor/blower control curve fluctuates wildly, and frequent and violent compressor speed fluctuations can lead to wear of mechanical components. In contrast, the compressor control curve of other depth strengthening algorithms is relatively gentle, which can adapt to the actual frequency conversion control of the compressor.

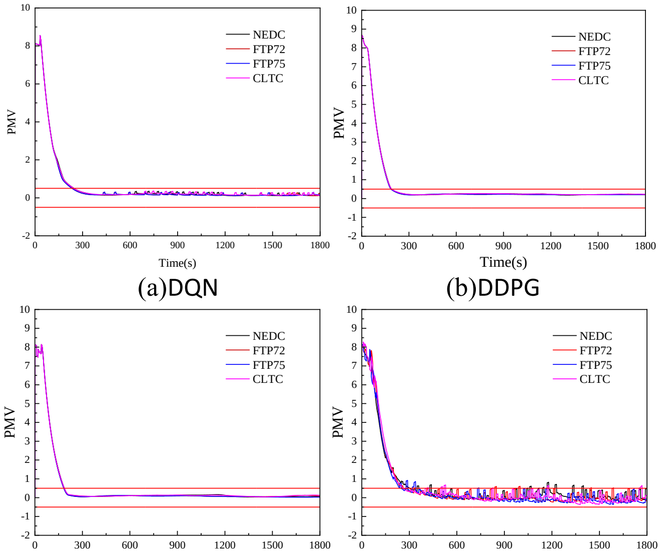


(a) Compressor speed (b) Blower fan voltage

Fig.8 Control curves

### 4.2.5 Generalization performance

This study selected NEDC, FTP72, FTP75, and CLTC as the test driving cycles to test the generalization abilities of different deep reinforcement learning algorithms. The environmental conditions for these tests were set at 40°C with a solar radiation intensity of 750 W/m2. Fig. 9 shows the thermal comfort control performance of the four deep reinforcement learning-based control strategies when facing non-training driving cycles. It can be observed that all four deep reinforcement learning algorithms achieve good thermal comfort performance under these conditions. Table 4 provides their energy consumption performance. Compared to traditional on/off control, DQN reduces energy consumption by an average of 47.1% in the four non-training driving cycles and 34.3% in the training cycle WLTC. DDPG reduces energy consumption by an average of 50.4% in the four non-training driving cycles and 53.2% in the training cycle WLTC. TD3 reduces energy consumption by an average of 44.8% in the four non-training driving cycles and 49.2% in the training cycle WLTC. SAC reduces energy consumption by an average of 27.1% in the four non-training driving cycles and 41.9% in the training cycle WLTC. This indicates that even under different test conditions, DQN, DDPG, and TD3 still perform well, while SAC's performance is relatively inferior.



(a)DQN (b)DDPG

(c)TD3 (d)SAC

Fig 9 Thermal comfort control under test cycle

Table 4 Energy consumption of the air conditioning system under test conditions

| control strategy | Energy consumption (kW·h) | | | |
|---|---|---|---|---|
| | NEDC | FTP72 | FTP75 | CLTC |
| on/off | 1.07 | 1.01 | 1.05 | 1.11 |
| DQN | 0.57 | 0.56 | 0.53 | 0.58 |
| DDPG | 0.53 | 0.53 | 0.50 | 0.54 |
| TD3 | 0.59 | 0.59 | 0.56 | 0.60 |
| SAC | 0.75 | 0.75 | 0.75 | 0.84 |

## 5. CONCLUSION

After researching deep reinforcement learning-based control strategies for electric vehicle air conditioning, this paper draws the following conclusions:

This study employs four deep reinforcement learning algorithms to control the electric vehicle air conditioning system, considering both passenger cabin temperature control and energy consumption of the air conditioning system. The algorithm's applicability is enhanced by randomly setting environmental temperature and solar radiation intensity during training.

Deep reinforcement learning-based control strategies for electric vehicle air conditioning significantly reduce the energy consumption of the air conditioning system. Compared to traditional on/off control, DDPG achieves an average energy consumption reduction of 37.6% within an environmental temperature range of 25-45°C.

When comparing control strategies of different deep reinforcement learning algorithms, it was observed that DQN and TD3 exhibit relatively stable control, whereas SAC, based on a stochastic policy, shows significant fluctuations in compressor control, which could lead to severe mechanical wear. DQN, DDPG, and TD3 control strategies, on the other hand, demonstrate smoother operation.

Regarding the generalization capability of control strategies, this paper tested four different operating conditions and found that DQN, DDPG, and TD3 all achieve reasonable thermal comfort control and energy-saving results under different conditions.

## ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China under Grant 52072052.

## REFERENCE

[1] Gao S W, Gong X Z, Liu Y, et al. Energy Consumption and Carbon Emission Analysis of Natural Graphite Anode Material for Lithium Batteries. Materials Science Forum, 2018, 913: 985-990.

[2] Farrington R, Rugh J. Impact of vehicle air-conditioning on fuel economy, tailpipe emissions, and electric vehicle range. National Renewable Energy Lab.(NREL), Golden, CO (United States), 2000.

[3] Huang X, Li K, Xie Y, et al. A novel multistage constant compressor speed control strategy of electric vehicle air conditioning system based on genetic algorithm. Energy, 2022, 241: 122903.

[4] Xie Y, Liu Z, Liu J, et al. A Self-learning intelligent passenger vehicle comfort cooling system control strategy.

Applied Thermal Engineering, 2020, 166: 114646.

[5] Zhang Q, Li S E, Deng K, et al. Modeling air conditioning system with storage evaporator for vehicle energy management. Automotive Air Conditioning: Optimization, Control and Diagnosis, 2016: 247-266.

[6] Wang H, Kolmanovsky I, Amini M R, et al. Model predictive climate control of connected and automated vehicles for improved energy efficiency. 2018 Annual American Control Conference (ACC), 2018: 828-833.

[7] Kasbi M, Sallans B, Russ G. A New Approach in Controlling the Compressor of the Vehicle Air Conditioning System. 2006 IEEE Intelligent Vehicles Symposium, 2006: 484-491.

[8] Brusey J, Hintea D, Gaura E, et al. Reinforcement learning-based thermal comfort control for vehicle cabins. Mechatronics, 2018, 50: 413-421.

[9] Choi W, Kim J W, Ahn C, et al. Reinforcement Learning-based Controller for Thermal Management System of Electric Vehicles. IEEE Vehicle Power and Propulsion Conference (VPPC), 2022.

[10] Joo S, Lee D, Kim M, et al. Multi-Agent Reinforcement Learning Based Actuator Control for EV HVAC Systems. Ieee Access, 2023, 11: 7574-7587.

[11] Grald E W, Macarthur J W. A moving-boundary formulation for modeling time-dependent two-phase flows. International Journal of Heat and Fluid Flow, 1992, 13(3): 266-272.

[12] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015, 518(7540): 529-533.

[13] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. arXiv preprint arXiv:1509.02971, 2015.

[14] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods. International conference on machine learning, 2018: 1587-1596.

[15] Haarnoja T, Zhou A, Abbeel P, et al. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. International conference on machine learning, 2018: 1861-1870.