The Aeronautical Journal (2026), **130**, pp. 508–525
doi:10.1017/aer.2025.10080

RESEARCH ARTICLE

Collaborative penetration guidance strategy for hypersonic vehicles based on double-DQN

B. Teng[1], J. Song[1] and M. Wei[2]

[1]School of Astronautics, Beihang University, Beijing, China
[2]China Aerospace Science and Industry Corp, Beijing Electro-Mechanical Engineering Institute, Beijing, China
Corresponding author: J. Song; Email: songjia@buaa.edu.cn

Received: 19 March 2025; **Revised:** 15 September 2025; **Accepted:** 16 September 2025

Keywords: multi-hypersonic vehicles; predictor-corrector method; reinforcement learning

Abstract

The penetration strategy of hypersonic vehicles in hostile environments is a critical factor in determining their effectiveness in completing reconnaissance or strike missions. Reinforcement learning (RL), as an end-to-end method, exhibits inherent advantages in addressing complex problems. However, existing research indicates that to enhance the efficiency of RL-based strategies, further advancements are necessary to reduce training costs and improve generalisation capabilities. This paper introduces a RL-based cooperative guidance law for multi-hypersonic vehicles, incorporating the estimated remaining time-of-flight and the absolute value of the bank angle obtained through a predictor-corrector method. The observation space and reward function are specifically designed to simplify the complex decision-making problem into a single-value decision problem, thereby reducing computational complexity and training costs. The proposed guidance law integrates the observation space, reward function and action space within the reinforcement learning framework to control flight trajectories, flight time and penetration of no-fly zones, ensuring compliance with multiple constraints. Model training and simulation tests conducted under multiple constraints demonstrate that the proposed approach reduces the training iterations required for the reinforcement learning agent and improves decision-making efficiency. Furthermore, simulations under different no-fly zone distributions confirm the proposed guidance approach's high generalisation ability.

Nomenclature

RL	reinforcement learning
CAV	common aero vehicle
DNN	deep neural networks
DQN	deep Q network
DDPG	deep deterministic policy gradient
PPO	proximal policy optimisation

1.0 Introduction

Hypersonic vehicles, distinguished by their high manoeuverability, precision and extended endurance, excel in penetration missions [1]. However, with the development of anti-missile systems, the complexity of no-fly zones has increased, and the penetration capability requirements of hypersonic vehicles have also increased. As a result, the survivability of a single hypersonic vehicle has been significantly constrained. To address these challenges, the cooperative penetration strategy of multi-hypersonic vehicles has become an important research focus. Cooperative penetration can reduce the probability of interception within no-fly zones, thereby enhancing deterrence against the target, and perform special

© The Author(s), 2025. Published by Cambridge University Press on behalf of Royal Aeronautical Society.

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

tasks such as 'silent' strike and target saturation attack, so as to improve the overall strike efficiency. Nevertheless, the complex flight environment puts forward higher adaptability and accuracy requirements for the existing time-cooperative penetration strategies. Therefore, the design of an efficient time cooperative guidance strategy for multi-hypersonic vehicles has become a key issue in recent research. A major challenge is how to quickly and accurately plan the feasible or optimal flight trajectories of these multi-vehicle formations [2].

The advancement of anti-missile technologies has brought heightened focus to the escalating difficulty of penetrating restricted airspace or no-fly zones. The tentacle feedback method has shown certain potential in addressing this issue [3], and can also have a good performance in the complex flight environment of hypersonic vehicles [4]. However, as a local planning algorithm, it is prone to falling into local optimal state under the condition of complex no-fly zone, highlighting the need for more effective algorithms to achieve effective penetration. For this reason, researchers have carried out many studies, and proposed targeted guidance methods specifically designed to account for the strong nonlinearity and strong coupling characteristics of hypersonic vehicles [5, 6]. A common approach is to decouple the longitudinal and lateral guidance, and design strategies for each separately [7]. In addition, progress has been made in the development of integrated guidance and control laws for hypersonic vehicles [8, 9]. Time-cooperative guidance exhibits significant advantages in enhancing the strike effectiveness against designated targets. Estimating the remaining time-of-flight and the reversal strategy of the bank angle have always been the key to the time-cooperative guidance problem [10]. While sliding mode algorithm performs well in the time-cooperative problem of the traditional vehicles [11], its chattering characteristics make it unsuitable for hypersonic vehicles. Consequently, it is necessary to design time-cooperative guidance strategies tailored to the characteristics of hypersonic vehicles. One approach decomposes the time coordination problem into two stages: coarse adjustment and fine adjustment. The coarse adjustment is realised by the lateral guidance strategy, and then the fine adjustment of the longitudinal guidance strategy can realise the time coordination of the hypersonic vehicle [12]. Furthermore, many scholars have conducted targeted research on this problem [13, 14], contributing to the development of advanced solutions.

However, the above researches primarily focus on individual problems, and fail to simultaneously consider the no-fly zone constraint and time coordination constraint. In view of this problem, numerous scholars have carried out some researches and proposed some time coordination penetration strategies. For instance, by designing the linear relationship between energy and bank angle, the remaining time-of-flight and bank angle reversal strategy can be determined simultaneously [15]. With the development of neural networks, their ability to model complex functions and end-to-end output characteristics have been widely used in many fields, including the field of hypersonic vehicle cooperative guidance [16–18]. Combining neural network with traditional methods has shown to enhance guidance effectiveness significantly [19]. However, in the aforementioned studies, the application of deep learning algorithms to guidance algorithm design is heavily dependent on large-scale dataset training, restricting their functionality to classification tasks based on existing data while lacking the adaptive capability to recognise and classify unseen categories. The introduction of reinforcement learning method provides a significant opportunity to overcome these shortcomings.

Reinforcement learning methods are widely applied across various domains due to their unique characteristics, particularly their ability to self-learn and continuously adapt. These attributes make them highly versatile in tackling complex problems. Wang designed a hypersonic vehicle attitude control method based on reinforcement learning [20]. Xu used the reinforcement learning method to realise the cooperative landing of UAVs [21]. Moreover, the inherent characteristics of reinforcement learning – such as its focus on long-term rewards and effectiveness in parameter optimisation – align closely with the requirements of time-coordinated guidance tasks for hypersonic vehicle. Therefore, utilising reinforcement learning methods for the research of time-coordinated guidance for hypersonic vehicle provides a viable approach [10]. The deep Q network (DQN) algorithm and the deep

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

deterministic policy gradient (DDPG) algorithm have been applied to the time cooperative attack on both fixed and dynamic targets, respectively [16, 22]. Additionally, Zhuang applied the proximal policy optimisation (PPO) algorithm to solve the penetration problem of a six-degree-of-freedom unmanned aerial vehicle UAV under the interceptor constraint [23], further demonstrating the potential of reinforcement learning in complex guidance scenarios. However, these methods have notable limitations such as extensive training time, which often requires tens of thousands of training iterations to achieve converge, or poor adaptability, which requires multiple trainings according to different tasks. Thus, more practical, efficient and generalisable coordinated entry guidance methods for multi-hypersonic vehicles still need more attention and investigation.

A time-cooperative guidance method based on reinforcement learning is proposed in this paper to tackle the time-coordinated penetration task for hypersonic vehicles. Here, 'time-cooperative guidance' refers to a strategy in which a single hypersonic vehicle adjusts its trajectory to meet time-related constraints, such as synchronised arrival, without involving explicit coordination or communication among multiple agents. The reinforcement learning algorithm is therefore applied to optimise the guidance behaviour of an individual agent, rather than in a multi-agent cooperative setting. To overcome issues such as prolonged training times and limited algorithm adaptability, a predictor-corrector algorithm is introduced to predict the remaining time-of-flight and calculate the absolute value of the bank angle, which is used as the observation of the reinforcement learning agent. A carefully designed reward function is proposed to transform the complex decision-making process into a simplified single-value selection problem, thereby significantly reduces the training cost. The effectiveness of the proposed method is validated through training and simulations under multiple constraints. Furthermore, the generalisation ability of the algorithm is demonstrated through simulations conducted under various no-fly zone distributions.

The remainder of this paper is organised as follows. The hypersonic vehicle reentry model, the multi-constraint model and the problem description is given in section 2. Then a reinforcement learning-based time-coordinated penetration algorithm for multi-hypersonic vehicles is proposed in section 3. Following that, simulations under multiple constraints and different no-fly zone distributions are shown in section 4. Finally, the conclusion of this paper is provided in section 5.

2.0 Problem statement

With the complexity and informatisation of the future battlefield environment, it is difficult for conventional tactical missiles to complete combat missions. Hypersonic vehicles have the structural advantages of high lift-to-drag ratio and excellent manoeuverability, enabling it to fly in unconventional trajectories. Therefore, the survivability under complex battlefield conditions has been greatly improved, and the task completion rate has also been improved accordingly. The mission objective of this paper is the cooperative strike of multi-hypersonic vehicles on the target point under the constraints of heat flow, overload, dynamic pressure, control quantity, terminal velocity and height, terminal longitude and latitude, no-fly zone and time coordination. Firstly, the modeling of multi-hypersonic vehicles is carried out, and then the models of various constraints are analysed one by one. The environmental model is shown in Fig. 1.

The reentry phase of the hypersonic vehicle is divided into the initial descent phase, the glide phase and the end phase. The glide phase has the largest span and the longest flight time, so the probability of detection and interception is also the largest. In this paper, glide phase time coordination and no-fly zone penetration are studied.

2.1 Multi-hypersonic vehicles model

According to the kinematics and dynamics equations of the hypersonic vehicle, the motion equation of the glide phase of multi-hypersonic vehicles is established within the track coordinate system as

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

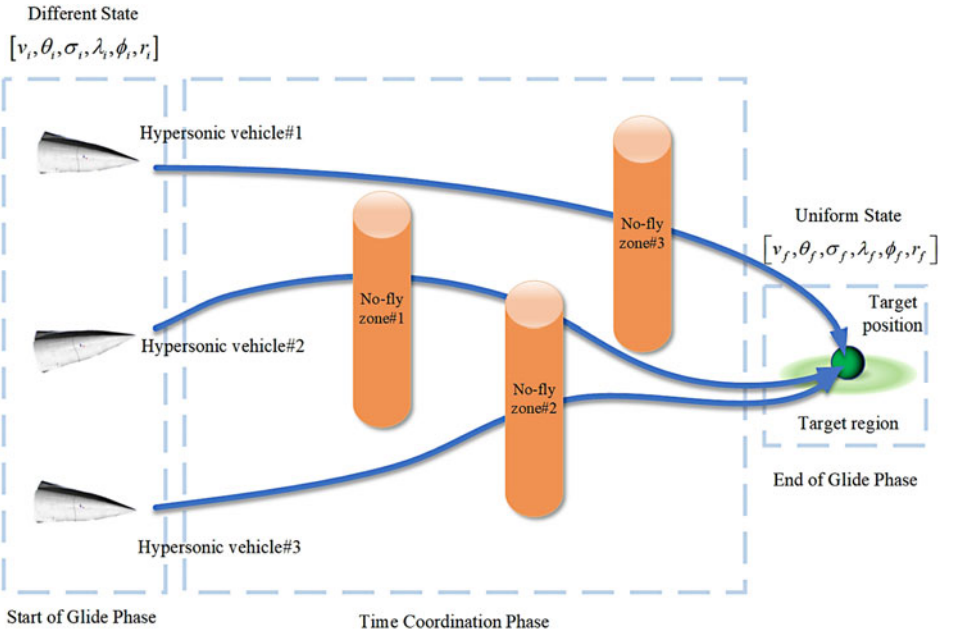


Figure 1. *Schematic diagram of time cooperative penetration scenario for multi-hypersonic vehicles.*

follows [13, 25]:

$$\begin{cases} \dfrac{dv_{\mathrm{i}}}{dt} = -D_{\mathrm{i}} + g'_{\mathrm{r}} \sin\theta + C'_{\mathrm{v_i}} \\ \dfrac{d\theta_{\mathrm{i}}}{dt} = \dfrac{L_{\mathrm{i}} \cos\gamma_{\mathrm{i}}}{v_{\mathrm{i}}} + \dfrac{g'_{\mathrm{r}} \cos\theta_{\mathrm{i}}}{v_{\mathrm{i}}} + \dfrac{v_{\mathrm{i}} \cos\theta_{\mathrm{i}}}{r_{\mathrm{i}}} + C'_{\theta_{\mathrm{i}}} + C''_{\theta_{\mathrm{i}}} \\ \dfrac{d\sigma_{\mathrm{i}}}{dt} = \dfrac{L_i \sin\gamma_i}{v_i \cos\theta_i} + \dfrac{v_i \tan\phi_i \cos\theta_i \sin\sigma_i}{r_i \cos\theta_i} + C'_{\sigma_{\mathrm{i}}} + C''_{\sigma_{\mathrm{i}}} \\ \dfrac{d\lambda_i}{dt} = \dfrac{v_i \cos\theta_i \sin\sigma_i}{r_i \cos\phi_i} \\ \dfrac{d\phi_i}{dt} = \dfrac{v_i \cos\theta_i \cos\sigma_i}{r_i} \\ \dfrac{dr_i}{dt} = v_i \sin\theta_i \end{cases} \tag{1}$$

The subscript i represents the state of the ith hypersonic vehicle. v_i represents the velocity of the hypersonic vehicle relative to the surface of the Earth; θ_i is the local trajectory inclination; σ_i is the azimuth angle of the hypersonic vehicle; r_i represents the radial distance between hypersonic vehicle's centre of mass and the centre of the earth. λ_i and ϕ_i are the geographical longitude and latitude of the current position of the hypersonic vehicle, respectively. γ_i denotes the bank angle; g'_r is the gravity of the earth; D_i and L_i correspond to the aerodynamic drag and lift forces acting on the vehicle, respectively. $C'_{\mathrm{v_i}}$, $C'_{\theta_{\mathrm{i}}}$, $C''_{\theta_{\mathrm{i}}}$, $C'_{\sigma_{\mathrm{i}}}$, $C''_{\sigma_{\mathrm{i}}}$ are the supplementary factors caused by earth's rotation, and their values could be

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

calculated using the subsequent equations:

$$\begin{cases} C'_{v_i} = -\omega_e^2 r_i(\cos\phi_i \sin\phi_i \cos\sigma_i \cos\theta_i - \cos^2\phi_i \sin\theta_i) \\ C'_{\theta_i} = \dfrac{\omega_e^2 r_i}{v_i}(\cos\phi_i \sin\phi_i \cos\sigma_i \sin\theta_i + \cos^2\phi_i \cos\theta_i) \\ C''_{\theta_i} = 2\omega_e \cos\phi_i \sin\sigma_i \\ C'_{\sigma_i} = \dfrac{\omega_e^2 r_i \cos\phi_i \sin\phi_i \sin\sigma_i}{v_i \cos\theta_i} \\ C''_{\sigma_i} = -\dfrac{2\omega_e(\cos\phi_i \cos\sigma_i \sin\theta_i - \sin\phi_i \cos\theta_i)}{\cos\theta_i} \end{cases} \tag{2}$$

where ω_e is the angular velocity of the Earth's rotation. D_i and L_i could be calculated as follows:

$$\begin{cases} D_i = 0.5\rho\ (h_i)\ v_i^2 C_D(v_i, \alpha_i) S_{ref}/m \\ L_i = 0.5\rho\ (h_i)\ v_i^2 C_L(v_i, \alpha_i) S_{ref}/m \end{cases} \tag{3}$$

where $\rho_i(h_i)$ is the atmospheric density at the current height; the relationship is $\rho_i(h_i) = \rho_0 e^{(-h_i/H)}$, ρ_0 is the atmospheric density at sea level, h_i is the height of the vehicle, and $H = 7200$m is the equivalent density height of the earth's atmosphere. S_{ref} and m denotes the aerodynamic reference area and mass of hypersonic vehicle, respectively; $C_D(v_i, \alpha_i)$ and $C_L(v_i, \alpha_i)$ represents drag coefficient and lift coefficient of the hypersonic vehicle, respectively.

2.2 Complex multi-constraint modeling

Hypersonic vehicle is usually limited by heat flux constraints $\dot{Q}_{max}$, overload constraints n_{Tmax} and dynamic pressure constraints q_{max} during the course of the flight. In addition, the periodic oscillation of the gliding process is usually suppressed by the pseudo-equilibrium gliding condition [16], the above four constraints are all path constraints, their expressions are as follows:

$$\begin{cases} \dot{Q}_i = k_Q \rho(h_i)^{0.5} v_i^{3.15} \le \dot{Q}_{max} \\ n_i = \dfrac{1}{mg} |L_i \cos\alpha_i + D_i \sin\alpha_i| \le n_{Tmax} \\ q_i = \dfrac{1}{2}\rho_i v_i^2 \le q_{max} \\ L_i \cos\gamma_{EQ} - \left(\dfrac{\mu_M}{r_i^2} - \dfrac{v_i^2}{r_i}\right) + 2\omega_e v_i \ge 0 \end{cases} \tag{4}$$

where $\dot{Q}_i$, n_i and q_i denote the heat flux, overload and dynamic pressure acting on the vehicle, respectively. K_Q is a constant related to the shape structure of hypersonic vehicle; μ_M is the gravitational coefficient of the earth; γ_{EQ} is the assumed bank angle under the condition of balanced gliding; α_i is the hypersonic vehicle's angle-of-attack.

The terminal speed, altitude, longitude and latitude of the hypersonic vehicle can determine whether the vehicle can switch from the glide phase to the terminal guidance phase. Therefore, the speed, altitude, longitude and latitude of the terminal are constrained, their expressions are as follows:

$$\begin{cases} h_i\left(t_f\right) = h_f^*, v_i\left(t_f\right) = v_f^* \\ \lambda_i\left(t_f\right) = \lambda_f^*, \phi_i\left(t_f\right) = \phi_f^* \end{cases} \tag{5}$$

where t_f denotes the arrival time of the hypersonic vehicle, h_f^*, v_f^*, λ_f^*, and ϕ_f^* represent the anticipated terminal height, speed value, longitude and latitude value, respectively.

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

The no-fly zone constraint is mainly the area where there is an enemy radar or missile defense system. In this paper, it is modeled as a hemispherical area or an infinitely high cylindrical area:

The expression of the cylindrical no-fly zone is:

$$(\lambda_i - \lambda_j)^2 + (\phi_i - \phi_j)^2 \geq (R_j / r_i)^2 \tag{6}$$

The expression of the hemispherical no-fly zone is:

$$((\lambda_i - \lambda_j) \times r_i)^2 + ((\phi_i - \phi_j) \times r_i)^2 + (r_i - r_e)^2 \geq R_j^2 \tag{7}$$

where i denotes the ith hypersonic vehicle, j denotes the jth no-fly zone, and R_j denotes the radius of the no-fly zone.

At the same time, it is also necessary to constrain the bank angle and angle-of-attack of hypersonic vehicle and the amplitude of their change rate. The expression is as follows:

$$\begin{cases} \alpha_i \in [\alpha_{min}, \alpha_{max}] & |\dot{\alpha}_i| \leq \dot{\alpha}_{max} \\ \gamma_i \in [\gamma_{min}, \gamma_{max}] & |\dot{\gamma}_i| \leq \dot{\gamma}_{max} \end{cases} \tag{8}$$

where $[\alpha_{min}, \alpha_{max}]$, $[\gamma_{min}, \gamma_{max}]$ represent the allowable ranges of the angle-of-attack and bank angle, respectively, whereas $\dot{\alpha}_{max}$ and $\dot{\gamma}_{max}$ indicate their corresponding maximum rates of change.

The time coordination constraint is that the time of multi-hypersonic vehicles reaching the target point is consistent:

$$\left[T_f^1 \cdots T_f^i \cdots T_f^N\right] = T_f^* \tag{9}$$

Given the long-range nature of the mission and the fact that the vehicles are launched from different initial positions, sufficient spatial separation is maintained throughout the entire flight. Moreover, the terminal phase is defined as a circular area with a 100 km radius centred around the target point, which provides a large enough region to ensure that the probability of collision at the endpoint is extremely low. Therefore, intra-group collisions between vehicles are not considered in this study.

3.0 Derivation of guidance strategy

Addressing the challenge of time cooperative guidance for hypersonic vehicles adhering to no-fly zone constraints, this paper decouples the guidance process into longitudinal and lateral components, with each component designed independently. This decoupled approach ensures that the vehicle's trajectory satisfies complex process constraints, terminal constraints and no-fly zone limitations. The longitudinal guidance strategy establishes a correlation between the angle-of-attack and the hypersonic vehicle's speed, and realises the satisfaction of the terminal velocity and height constraints. The lateral guidance is achieved through the bank angle reversal strategy. In this paper, the predictor-corrector method is used to estimate the absolute value of the bank angle and the remaining time-of-flight. The double-DQN method is used to implement the bank angle reversal, so that it can meet the penetration requirements of the no-fly zone and the terminal position and time coordination constraints. The algorithm structure and process are shown in Fig. 2.

3.1 Longitudinal guidance strategy design

The flight states required for the longitudinal guidance of hypersonic vehicles can be directly obtained from the current flight cycle. By focusing on the design of the absolute value of the bank angle and the angle-of-attack, longitudinal guidance objectives can be achieved. Accordingly, the longitudinal guidance strategy is divided into two components: the computation of the absolute value of the bank angle and the formulation of the angle-of-attack.

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

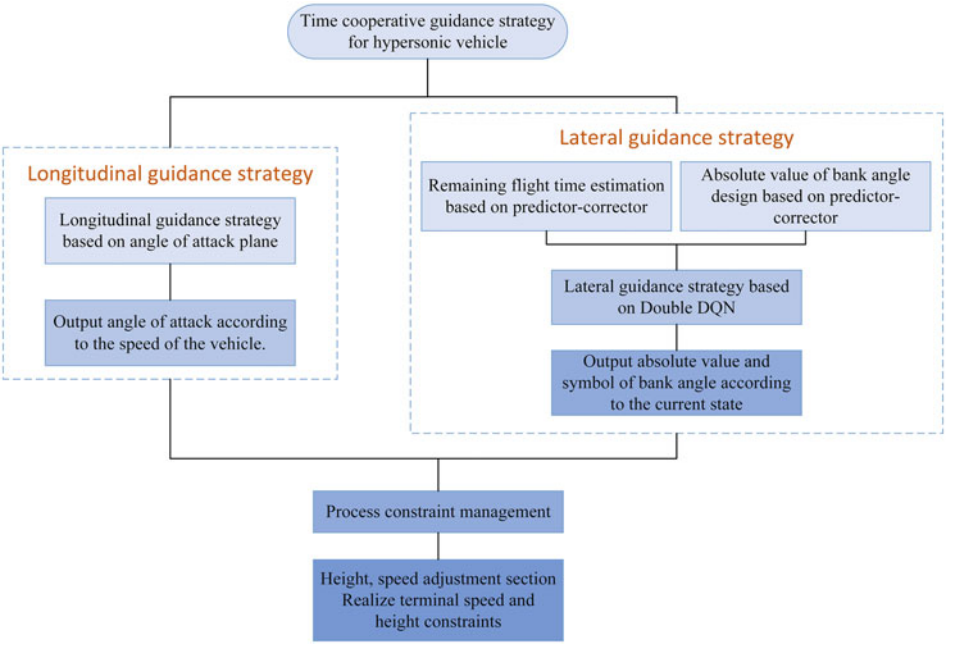


Figure 2. *Lateral and longitudinal maneuver strategy design.*

3.1.1-. Angle-of-attack plane design

In order to streamline the design process of the control sequence, the hypersonic vehicle's angle-of-attack command designed in this paper is designed as a fixed angle-of-attack plane related to speed, which is expressed as:

$$\alpha_i = \begin{cases} \alpha_{max,i} & v_i \geq V_{1,i} \\ \dfrac{\alpha_{maxL/D,i} - \alpha_{max,i}}{V_{2,i} - V_{1,i}}(v_i - V_{1,i}) + \alpha_{max,i} & V_{2.i} \leq v_i \leq V_{1.i} \\ \alpha_{maxL/D,i} & v_i \leq V_{2,i} \end{cases} \tag{10}$$

where α_{max} and $\alpha_{maxL/D}$ are the angles of attack corresponding to the maximum angle-of-attack and the maximum lift-to-drag ratio, and $V_{1,i}$ and $V_{2,i}$ are the given velocity values. In the initial stage, due to the large speed, in order to meet the thermal protection characteristics of hypersonic vehicles, a large angle-of-attack design is adopted to prevent the aircraft from excessive heat flow. In the middle and late range, in order to ensure the flight range of hypersonic vehicles, the angle-of-attack decreases gradually with the decrease of speed to the maximum lift drag ratio angle-of-attack [13, 24].

3.2 Lateral guidance strategy design

The time control of hypersonic vehicle is mainly to control its lateral manoeuvre. By designing the bank angle inversion strategy, the control of lateral manoeuvre can be realised. The flight states required for the lateral guidance of hypersonic vehicles can be directly obtained from the current flight cycle. The remaining time-of-flight and the absolute value of the bank angle of the hypersonic vehicles are calculated by the predictor-corrector method. Therefore, the role of the reinforcement learning algorithm in this paper is to obtain the bank angle symbol at the next moment according to the current state of the hypersonic vehicle and the remaining time-of-flight and the absolute value of the bank angle obtained by the predictor-corrector method. This reformulates the problem addressed by the reinforcement learning algorithm, simplifying it from a complex bank angle decision-making task to a straightforward bank

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

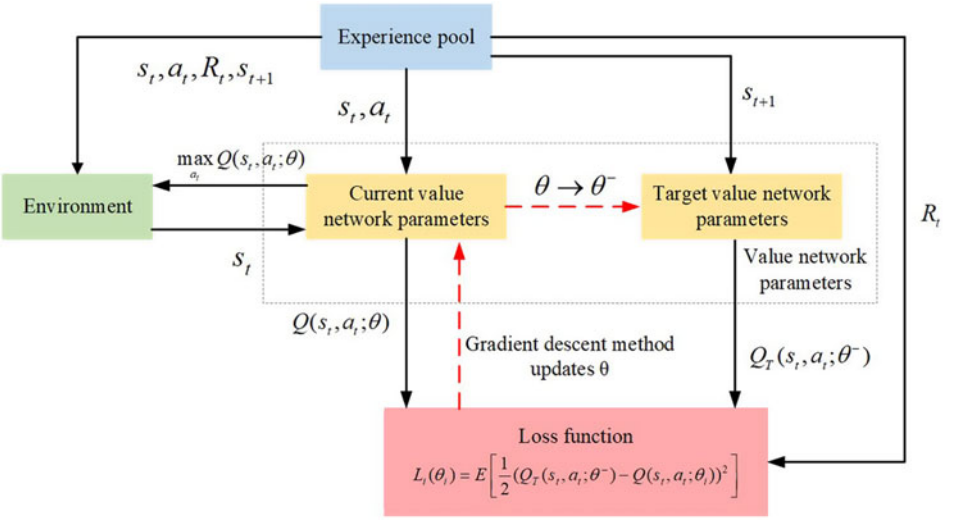


Figure 3. *Principle diagram of double-DQN algorithm.*

angle sign selection task. In this process, the input of the reinforcement learning agent is a continuous state, and the output is a discrete bank angle symbol. The double-DQN algorithm features a continuous observation space and a discrete action space. Its straightforward structure ensures ease of implementation while maintaining the capability to handle decision-making tasks in complex environments. Therefore, the double-DQN algorithm is selected to design the bank angle reversal strategy.

3.2.1 Double-DQN algorithm

The double-DQN algorithm builds upon the DQN algorithm by introducing two neural networks, each assigned distinct roles: one for selecting the optimal action and the other for calculating the target Q-value to update the parameters of the training network [26, 27]. The introduction of an additional network also modifies the update rule for the target Q-value:

$$Q_T(s_t, a_t; \theta^-) = \begin{cases} R_t s_{t+1} = s_{t_f} \\ R_t + \gamma \max\limits_{a_{t+1}} Q(s_{t+1}, a_{t+1}; \theta^-) s_{t+1} \neq s_{t_f} \end{cases} \tag{11}$$

where s_t and α_t represent the current state and action, respectively, and θ^- represents the network parameters; R_t is the expected value at the current time; s_{t_f} is the terminal time state.

The training procedure of the double-DQN algorithm is shown in Fig. 3. During the training process, the DQN network will traverse all the actions that can be taken at this moment, and estimate the value after taking the action, output the action with the greatest value, and then update the network parameter θ according to the error between the actual value and the estimated value after taking the action with the greatest value. At the same time, in order to enable the network to explore the unknown area, the exploration rate ϵ is designed, so that the network has a certain probability to adopt a strategy that has not been adopted at each selection.

According to the Bellman equation form of $Q(s_t, a_t; \theta)$, the loss function of the ith iteration in the training process of the reinforcement learning network is defined as:

$$L_i(\theta_i) = E\left[\frac{1}{2}(Q_T(s_t, a_t; \theta^-) - Q(s_t, a_t; \theta_i))^2\right] \tag{12}$$

where θ_i denotes the network parameter at the ith iteration.

After a certain step of learning, the gradient descent method is used to update the network parameter θ^- of the objective value function, which can partially diminish the correlation between the current Q

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

value and the target Q value, enhancing the stability of the algorithm:

$$\theta_{i+1} = \theta_i + \nabla\theta_i \cdot L_i(\theta_i) = \theta_i + E\left[\left(Q_T(s_t, a_t; \theta^-) - Q(s_t, a_t; \theta_i)\right)\nabla\theta_i \cdot Q(s_t, a_t; \theta_i)\right] \tag{13}$$

where $\nabla\theta_i$ denotes the gradient of θ. According to the strategy of Equation (12), the network parameters are continuously updated to obtain the network parameter θ that determines the sign of the bank angle according to the current state and the target state, that is, to find the optimal strategy.

3.2.2 Remaining time-of-flight estimation

To complete the guidance task that satisfies the time coordination constraint, it is necessary to make a more accurate estimate of the remaining time-of-flight. Therefore, this section adopts the predictor-corrector method to achieve a more accurate estimate of the remaining time-of-flight through integral operation.

Since the flight height of the aircraft is significantly less than the radius of the Earth, that is, $r/R_e \approx 1$, combined with Equation (1) yields:

$$\frac{dv_i}{dt} = -\frac{1}{R_e}\frac{D_i}{L_i}\frac{g_0 R_e - {v_i}^2}{\cos\gamma_i} \tag{14}$$

that is:

$$dv_i = -\frac{1}{R_e}\frac{D_i}{L_i}\frac{g_0 R_e - {v_i}^2}{\cos\gamma_i}dt \tag{15}$$

Integral on both sides of the above formula:

$$\int_t^{t_f} dt = \int_v^{v_f} -R_e\frac{L}{D}\frac{\cos\gamma_i dv}{g_0 R_e - {v_i}^2} \tag{16}$$

from this, the analytical estimation expression of the remaining time-of-flight can be obtained:

$$t_{go,i} = \frac{\sqrt{R_e}}{2\sqrt{g_0}}\frac{L_i}{D_i}\cos\gamma_i \ln\left(\frac{\sqrt{g_0 R_e} + v_i\sqrt{g_0 R_e} - v_f}{\sqrt{g_0 R_e} - v_i\sqrt{g_0 R_e} + v_f}\right) \tag{17}$$

3.2.3 Absolute value of bank angle calculation

In this paper, the predictor-corrector method is also used to determine the absolute value calculation of the bank angle. Firstly, the energy of the hypersonic vehicle is defined as $e_i = \mu_M/r - v_i^2/2$, where μ_M is the earth constant, and the terminal constraint of the mission is transformed into:

$$\begin{cases} s_{f,i} = s\left(e_{f,i}\right) = 0 \\ e_{f,i} = \dfrac{\mu_M}{r_{f,i}} - \dfrac{v_{f,i}^2}{2} \end{cases} \tag{18}$$

From the kinematic equation, the remaining range s_i of the longitudinal range of the aircraft and the energy e_i of the aircraft satisfy the following relationship:

$$\frac{ds_i}{dt_i} = v_i\cos\theta_i,\ \frac{de_i}{dt_i} = -\dot{v}_i v_i = -D_i v_i \Rightarrow \frac{ds_i}{de_i} = -\frac{\cos\theta_i}{D_i} \tag{19}$$

s_i is the longitudinal distance of the current aircraft relative to the target point:

$$s_{res,i} = R_0 \cdot \arccos\left[\sin\phi i \sin\phi_f + cos\,\phi i \cos\phi f \cos\left(\lambda_f - \lambda_i\right)\right] \tag{20}$$

where R_0 denotes the radius of the earth and $[\lambda_f, \phi_f]$ is the longitude and latitude of the target point. The angle-of-attack planing surface has been given by the previous article, so the variables that can affect the voyage are only the bank angle instructions γ_i. Therefore, it is necessary to optimise the inclination angle in the solution set, and the absolute value of the inclination angle that meets the terminal constraints can be obtained. By using the secant method, the bank angle γ_i is adjusted iteratively. When the remaining

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

range $s_{res,i}$ of the aircraft is close to the given remaining range $s^*_{res,i}$, the bank angle command γ_i^* satisfying the terminal constraint in the current guidance period can be obtained. The process is as follows:

$$\begin{aligned} F_i(\gamma_i) &= s_{res,i}(\gamma_i) - s^*_{res,i} \\ |\gamma_i|^{k+1} &= |\gamma_i|^k - \frac{F_i^k}{F_i^k - F_i^{k-1}}\left(|\gamma_i|^k - |\gamma_i|^{k-1}\right) \end{aligned} \tag{21}$$

where k is the number of iterations, which is usually related to the allowable error value.

After the inclination angle command is obtained, it is also necessary to constrain it to meet the heat flow, overload and dynamic pressure constraints in the process. The expression is as follows:

$$\left.\begin{aligned} |\gamma_{Q,i}(v_i)| &\le \cos^{-1}\left[\left(g - \frac{v_i^2}{r_i}\right)\frac{2m_i v_i^{4.3}}{C_{L,i}S_{ref,i}}\left(\frac{k_Q}{\dot{Q}_{max,i}}\right)^2\right] = |\gamma_{Q\max,i}(v_i)| \\ |\gamma_{n,i}(v_i)| &\le \cos^{-1}\left[\left(g - \frac{v_i^2}{r_i}\right)\frac{\sqrt{1+\left(C_{D,i}/C_{L,i}\right)^2}}{g_0 n_{max,i}}\right] = |\gamma_{n\max,i}(v_i)| \\ |\gamma_{q,i}(v_i)| &\le \cos^{-1}\left[\left(g - \frac{v_i^2}{r_i}\right)\frac{m}{C_{L,i}Q_{max}S_{ref,i}}\right] = |\gamma_{q\max,i}(v_i)| \end{aligned}\right\} \tag{22}$$

where $|\gamma_{Q\max,i}(v_i)|$, $|\gamma_{n\max,i}(v_i)|$ and $|\gamma_{q\max,i}(v_i)|$ correspond to the maximum bank angle under the maximum heat flux, overload and dynamic pressure limits, respectively.

In order to avoid the periodic oscillation of the flight trajectory caused by the bank angle command obtained by the predictive correction method, introducing the height feedback to suppress the large change of the flight trajectory. The expression is as follows:

$$\left|\gamma^*_{cmd,i}\right| = \left|\gamma_i^*\right| + K_{hp,i}(h_{QEGC,i} - h_i) + K_{hd,i}(\dot{h}_{QEGC,i} - \dot{h}_i) \tag{23}$$

where $\left|\gamma^*_{cmd,i}\right|$ presents the command bank angle after introducing feedback, $h_{QEGC,i}$ and $\dot{h}_{QEGC,i}$ are the height and height change rate of balanced gliding, and $K_{hp,i}$ and $K_{hd,i}$ are two feedback coefficients.

In summary, the bank angle command can be obtained as follows:

$$\left|\gamma^*_{c,i}(v_i)\right| = \min\left\{\left|\gamma^*_{cmd,i}(v_i)\right|, \left|\gamma_{Q\max,i}(v_i)\right|, \left|\gamma_{n\max,i}(v_i)\right|, \left|\gamma_{q\max,i}(v_i)\right|\right\} \tag{24}$$

where $\left|\gamma^*_{c,i}(v_i)\right|$ is the absolute value instruction of the bank angle, which is determined in a way that ensures compliance with both the trajectory path and the terminal constraints of the flight.

3.2.4 Lateral guidance strategy based on double-DQN algorithm

Firstly, the double-DQN network structure is established. Because DQN has the problem of overestimation of Q value, the double-DQN algorithm uses two Q networks, one is the training network and the other is the target network. In this design, the two networks are the same.

A basic framework of reinforcement learning is the Markov decision process. In reinforcement learning, agents learn the optimal strategy by interacting with the environment, and this process can usually be expressed as a Markov decision process. The decision-making process of bank angle reversal of hypersonic vehicle has Markov property, that is, the state of the next moment is only related to the current state of the hypersonic vehicle and the action taken, and has nothing to do with the previous state. Therefore, the bank angle symbol decision process is modeled as a Markov decision-making process. Through the control of the lateral manoeuvre to achieve the adjustment of the flight time, it is necessary to comprehensively consider the remaining range, the remaining time-of-flight, the current flight state, etc., and synthesise it into the state space S:

$$S = \left[v_i\ \theta_i\ \sigma_i\ \lambda_i\ \phi_i\ h_i\ \alpha_i\ L_i\ \Delta\sigma_i\ \mu_i\ T_f^i\right] \tag{25}$$

where v_i, θ_i, σ_i, λ_i, ϕ_i, h_i, L_i represent the velocity, velocity inclination, track yaw angle, longitude, latitude, flight height and remaining range of the hypersonic vehicle, respectively. μ_i represents the

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

absolute value of the bank angle estimated by the predictor-corrector method, and T_f^i represents the flight time of the hypersonic vehicle.

Since the bank angle symbol has only positive and negative options, the action space is designed as:

$$A = [1, -1] \tag{26}$$

The design of the reward function is directly related to whether the reinforcement learning can complete the target task. The training goal of this paper is to enable the hypersonic vehicle to achieve time-cooperative penetration through the no-fly zone under the condition of satisfying the terminal state constraint. Therefore, the design of the reward function should allow the hypersonic vehicles to approach the target point while avoiding entering the no-fly zone. Since the focus of the hypersonic vehicles differs at different stages, the reward function is designed as two segmented functions, $R_1(s_{t_f})$ and $R_2(s_{t_f})$. $R_1(s_{t_f})$ controls the movement of the hypersonic vehicles towards the target point, with the primary consideration being the remaining range. To prevent excessive deviation that could cause trajectory divergence, the impact of the vehicle's track yaw angle is also considered. Additionally, to achieve time-coordinated guidance, the ratio of remaining time-of-flight to the set flight time is included in the reward function, and all three should decrease continuously when hypersonic vehicles fly correctly, making the reward inversely proportional to them. $R_2(s_{t_f})$ controls the hypersonic vehicle to penetrate through the no-fly zone, thus $R_2(s_{t_f})$'s only dependent variable is the distance of the hypersonic vehicle from the no-fly zone. The reward decreases as it approaches the no-fly zone and increases as it moves away from it. The expression for the reward function $R(s_{t_f})$ is as follows:

$$R_1(s_{t_f}) = \begin{cases} -0.4 - K1 \times \left(\frac{L_i}{6\times10^6}\right) - K2 \times |\Delta\sigma_i| - K3 \times \left(\frac{t_i}{T}\right) L_i \geq 4 \times 10^6 \\ -0.3 - K1 \times \left(\frac{L_i}{6\times10^6}\right) - K2 \times |\Delta\sigma_i| - K3 \times \left(\frac{t_i}{T}\right) 3 \times 10^6 \leq L_i \leq 4 \times 10^6 \\ -0.2 - K1 \times \left(\frac{L_i}{6\times10^6}\right) - K2 \times |\Delta\sigma_i| - K3 \times \left(\frac{t_i}{T}\right) 2 \times 10^6 \leq L_i \leq 3 \times 10^6 \\ -0.1 - K1 \times \left(\frac{L_i}{6\times10^6}\right) - K2 \times \left|\frac{\Delta\sigma_i}{2}\right| - K3 \times \left(\frac{t_i}{T}\right) 1 \times 10^6 \leq L_i \leq 2 \times 10^6 \\ 0.5 - K1 \times \left(\frac{L_i}{2\times10^6}\right) - K2 \times \left|\frac{\Delta\sigma_i}{5}\right| - K3 \times \left(\frac{t_i}{T}\right) 5 \times 10^5 \leq L_i \leq 1 \times 10^6 \\ 0.8 - K1 \times \left(\frac{L_i}{1\times10^6}\right) - K2 \times \left|\frac{\Delta\sigma_i}{6}\right| - K3 \times \left(\frac{t_i}{T}\right) 3 \times 10^5 \leq L_i \leq 5 \times 10^5 \\ 1.0 - K1 \times \left(\frac{L_i}{6\times10^6}\right) - K2 \times \left|\frac{\Delta\sigma_i}{6}\right| - K3 \times \left(\frac{t_i}{T}\right) 1 \times 10^5 \leq L_i \leq 3 \times 10^5 \end{cases} \tag{27}$$

$$R_2(s_{t_f}) = \begin{cases} R_{i,j} - 0.14 \times 10^5 \leq s_i \leq 5 \times 10^5, s_i \leq s'_i \\ R_{i,j} + 0.124 \times 10^5 \leq s_i \leq 5 \times 10^5, s_i \geq s'_i \\ R_{i,j} - 0.23 \times 10^5 \leq s_i \leq 4 \times 10^5, s_i \leq s'_i \\ R_{i,j} + 0.243 \times 10^5 \leq s_i \leq 4 \times 10^5, s_i \geq s'_i \end{cases} \tag{28}$$

where $K1$, $K2$ and $K3$ denote the weighting factors corresponding to the rewards related to the remaining range, the track yaw angle deviation and the remaining time-of-flight, respectively. And $R_1(s_{t_f})$ represents the reward that works all the way, which is mainly related to the remaining range, the remaining time-of-flight and the deviation of the track yaw angle. $R_2(s_{t_f})$ represents the reward that starts to work after entering a certain range of the no-fly zone, which is defined as the warning area of the no-fly zone. In this paper, the warning area of the cylindrical no-fly zone is a hollow cylinder with a thickness of 200 km around the no-fly zone, and the warning area of the hemispherical no-fly zone is a hemispherical shell with a thickness of 200 km. Since the risk of inter-group collisions is beyond the scope of this study, the reward function does not include any components to account for such risks. L_i represents the remaining range, $\Delta\sigma_i$ represents the track yaw angle deviation, t_i represents the flight flight, and s_i represents the distance of the hypersonic vehicle relative to the no-fly zone, respectively. After training, the trained

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

Table 1. *Initial states of hypersonic vehicles*

	Velocity (km/s)	Height (km)	Velocity dip angle (°)	Track interval angle (°)	Coordinate	Range (km)
CAV-H#1	6.79	56.08	0.0688	89.93	(10.24° E, 0°)	4811.5
CAV-H#2	6.79	56.09	0.0688	89.15	(10.28° E, 4.93° N)	4832.0
CAV-H#3	6.79	56.09	0.0688	90.70	(10.28° E, 4.92° S)	4832.0

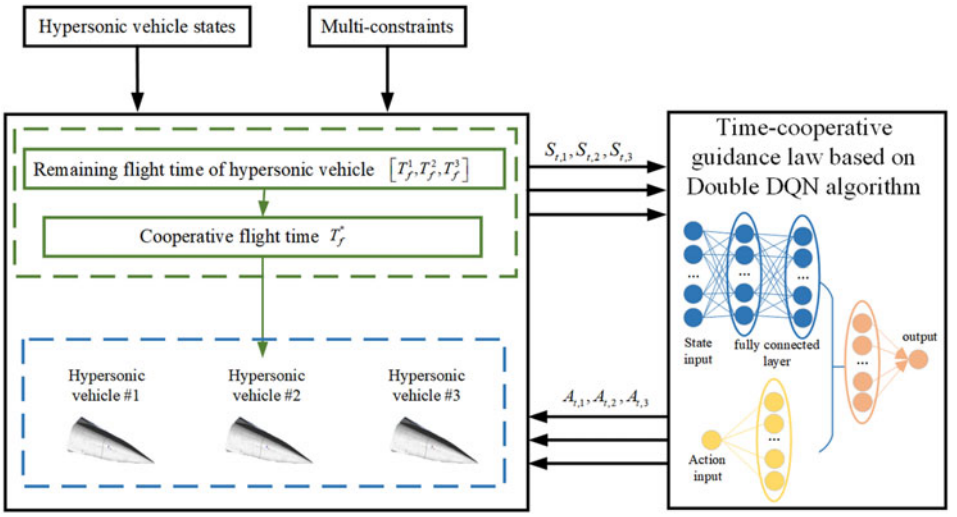


Figure 4. *Time cooperative penetration algorithm for multi-hypersonic vehicles.*

agent is applied to the bank angle reversal decision to achieve time cooperative guidance. The guidance process is depicted in Fig. 4.

The reinforcement learning guidance law in the figure is the agent obtained by offline training, and the bank angle symbol in the current state can be obtained by entering the state of the hypersonic vehicle into the agent.

4.0 Guidance law performance

The hypersonic vehicle model selects the CAV-H aircraft. The objective is to chart the reentry flight paths for three hypersonic vehicles. These vehicles take off from distinct points: (10° E, 0), (10° E, 5° N) and (10° E, 5° S). Following their launch, they undergo a period of time, the vehicles reach the reentry phase, at which point the algorithm designed in this paper comes into play. The initial state of each vehicle at the commencement of the reentry phase is shown in Table 1. The coordinate of the target point is (53.4° E, 0). When the remaining range of the hypersonic vehicle is less than 100 km, the hypersonic vehicle enters the terminal guidance phase and the reentry phase guidance ends. The specific constraint values of the hypersonic vehicle are shown in Table 2. The hardware platform used in the simulation verification process of this article is: CPU: Ryzen 7 8500H with Radeon Graphics; GPU: RTX 3060, the system platform is Windows 11 and the simulation software used is matlab2023b.

The guidance method designed in this paper is for the cooperative guidance of the reentry phase of the hypersonic vehicles. Therefore, the terminal height is set to 30 km, the distance from the target point is set to 100 km, and the remaining flight phase is the descent attack phase. The terminal height and the radius from the target are set to ensure that the vehicles have the same initial conditions after entering the descent phase, so as to carry out the cooperative guidance of the descent phase.

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

Table 2. *Hypersonic vehicle flight constraints*

Constraint variables	Value
Maximal heat flux	1300 KW/m^2
Maximum overload	6 g
Maximum dynamic pressure	300 KPas
Terminal velocity	2000 m/s
Terminal height	30 km
Position of the target point	(53.4° E, 0), $R = 100$ km
Angle-of-attack constraint	$\alpha_{max} = 25°, \alpha_{min} = 0°$
Bank angle constraint	$\gamma_{max} = 75°, \gamma_{min} = -75°$
Time coordination constraint	$T_1 = T_2 = \cdots = 1150s$

Table 3. *Double-DQN network parameters*

Parameter	Value
Sampling time	1
Maximum number of training rounds	3000
Maximum number of steps per round	1500
Target network update frequency	4
Experience pool size	100000
Number of samples taken each time	256
Exploration scheme	$\epsilon - greedy$
Learning rate	5×10^{-5}
Initial exploration rate	0.99
Exploration rate decay rate	1×10^{-5}
Annealing point	0.1

The network parameter settings are shown in Table 3.

4.1 Multi-hypersonic vehicles cooperative guidance under no-fly zone constraint

The initial state of the hypersonic vehicle is consistent with that without the no-fly zone constraint, and the parameters of the no-fly zone are shown in Table 4. With the implementation of no-fly zone constraints, the vehicles must continuously assess their distance relative to the no-fly zone during the flight. In order to accomplish the flight mission, it is essential to consider the impact of no-fly zone constraints when approaching the vicinity of the no-fly zone. Once the vehicles move away from the no-fly zone, the primary objective shifts towards achieving the time-coordinated objectives. Therefore, this paper introduces a concept of a 'warning zone'. When vehicles enter the warning zone, the no-fly zone's influence on the reward function, donated as $R_2(s_{tf})$, comes into play, guiding the multi-hypersonic vehicles to navigate through the no-fly zone. Once the vehicles exit the warning zone, only $R_1(s_{tf})$ remains in effect, guiding multi-hypersonic vehicles to achieve time coordination. The simulation results are shown in Figs. 5 and 6, and Table 5.

The training reward curve is shown in Fig. 5, where the blue curve represents the reward value for a single training session, and the red curve represents the average reward value for the previous 200 training sessions, from which it can be seen that under the no-fly zone constraints, due to the introduction of the remaining time-of-flight and the absolute value of the bank angle predicted by the predictor-corrector algorithm, the reward value can be maintained at the maximum value of the reward around 200 times of training, and the average reward curve converges at 1000 times of training. Compared with using reinforcement learning alone to achieve time coordination among multiple hypersonic vehicles,

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

Table 4. *No-fly zone constraint parameters setting*

Definition	Longitude	Latitude	Radius
No-fly zone#1	25° E	2° S	300 km
No-fly zone#2	35° E	5° N	300 km
No-fly zone#3	45° E	1.5° N	300 km

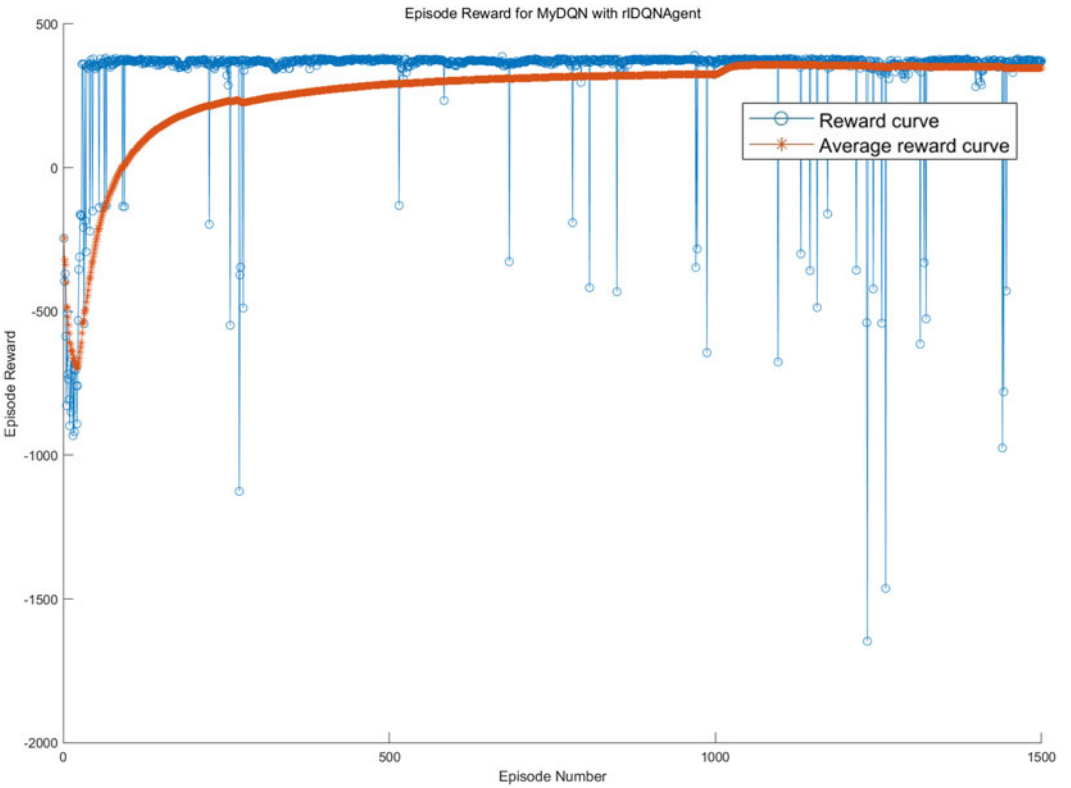


Figure 5. *Training reward curve under no-fly zone.*

the algorithm proposed in this paper incorporates no-fly zone constraints and still achieves a reduction in training iterations by two orders of magnitude. This highlights the superior efficiency of the proposed method in significantly reducing training requirements and improving training performance [28]. The curves in Fig. 6a shows that the heat flow, overload, and dynamic pressure meet the specified constraints. This ensures that the vehicle does not sustain any mechanical or structural damage during reentry. It can be seen from the figure that it can meet the constraints set above. So as to ensure that the vehicle will not produce mechanical or structural damage. The trajectory curves depicted in Figs. 6b and 6c demonstrate the algorithm's effectiveness in steering multiple hypersonic vehicles towards their designated targets while simultaneously avoiding no-fly zones, as necessitated by the constraints of these restricted areas. The change curve of the bank angle during the flight is shown in 6e. The time coordination and the penetration of the no-fly zone are realised by the change of the bank angle. Because the closer to the target point, the change of the line-of-sight angle is more intense, so the change frequency of the bank angle is also increased. The curve in Fig. 6f represents the designed angle-of-attack planing surface, the terminal altitude of the hypersonic vehicle is related to the angle-of-attack and the bank angle, through the design of the angle-of-attack plane, the terminal altitude of the hypersonic vehicle can be kept within a certain range. Therefore, by designing the angle-of-attack planing surface, the terminal height constraint in the

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

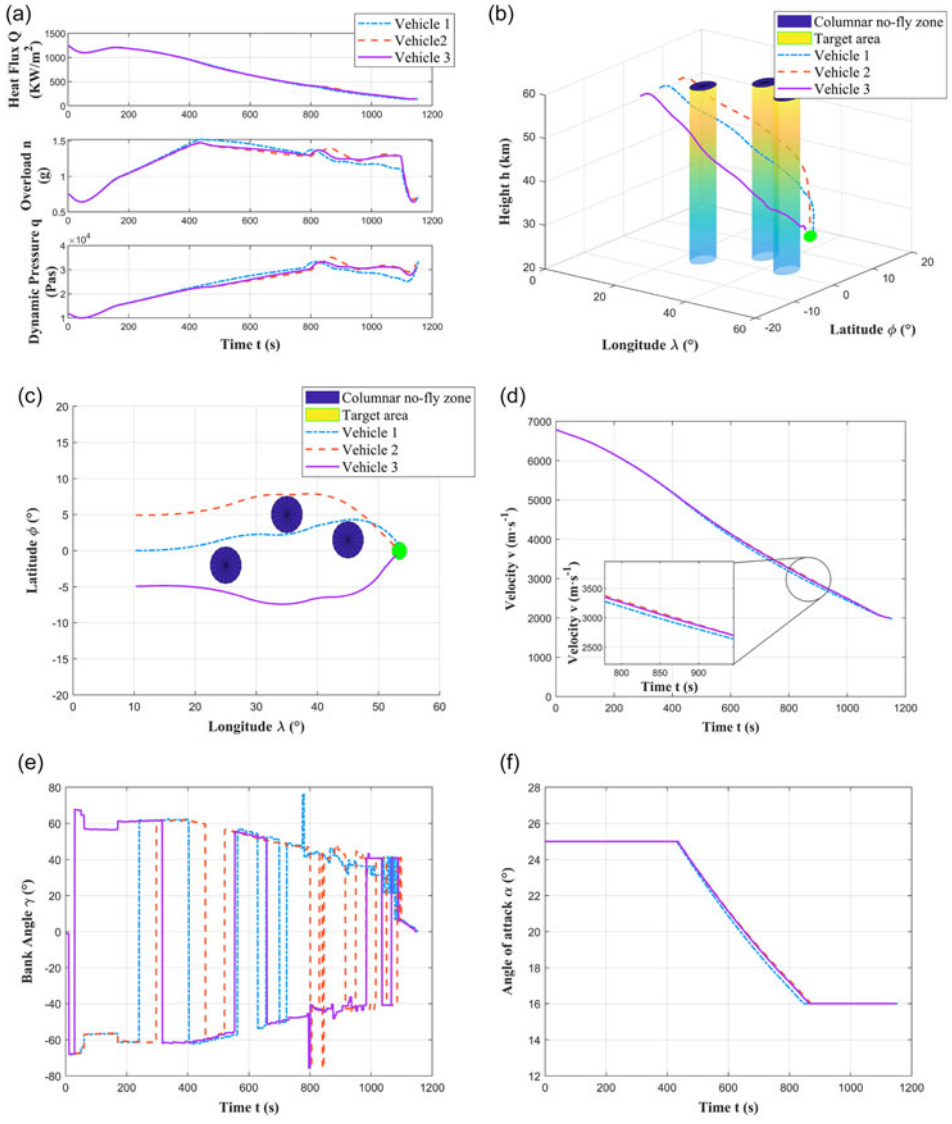


Figure 6. *Time co-simulation results of multi-hypersonic vehicles under no-fly zone.*

flight mission can be satisfied. The information presented in Table 5 demonstrates that the algorithm successfully enables time-coordinated guidance for multiple hypersonic vehicles within these restricted areas. This evidences the generalisation ability of the proposed algorithm.

4.2 Multi-hypersonic vehicles cooperative guidance under different no-fly zone constraint

In order to verify the universality of the designed method, and its performance under unseen conditions, the distribution of the no-fly zone is modified after training on the scenario described in 4.1, and the simulation verification is carried out. The simulation results are shown in Fig. 7 and Table 6.

The penetration performance of multi-hypersonic vehicles under three different no-fly zone distributions is shown in Fig. 7. The data in Table 6 show the performance of the designed algorithm under different no-fly zone distributions. It can be seen from the figure that when the distribution of the no-fly zone changes, the designed algorithm can successfully realise the time-coordinated penetration task of

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

Table 5. *Terminal state and error of multi-hypersonic vehicles*

	Velocity error (m/s)	Height error (km)	Terminal coordinate	Flight time error (s)
CAV-H#1	7.54(0.38%)	1.57(5.23%)	(52.87° E, 0.07° N)	8(0.70%)
CAV-H#2	5.50(0.28%)	1.41(4.70%)	(52.77° E, 0.06° N)	1(0.09%)
CAV-H#3	5.50(0.28%)	1.44(4.80%)	(52.83° E, 0.06° S)	3(0.26%)
Average	6.18(0.31%)	1.47(4.90%)	–	4(0.35%)

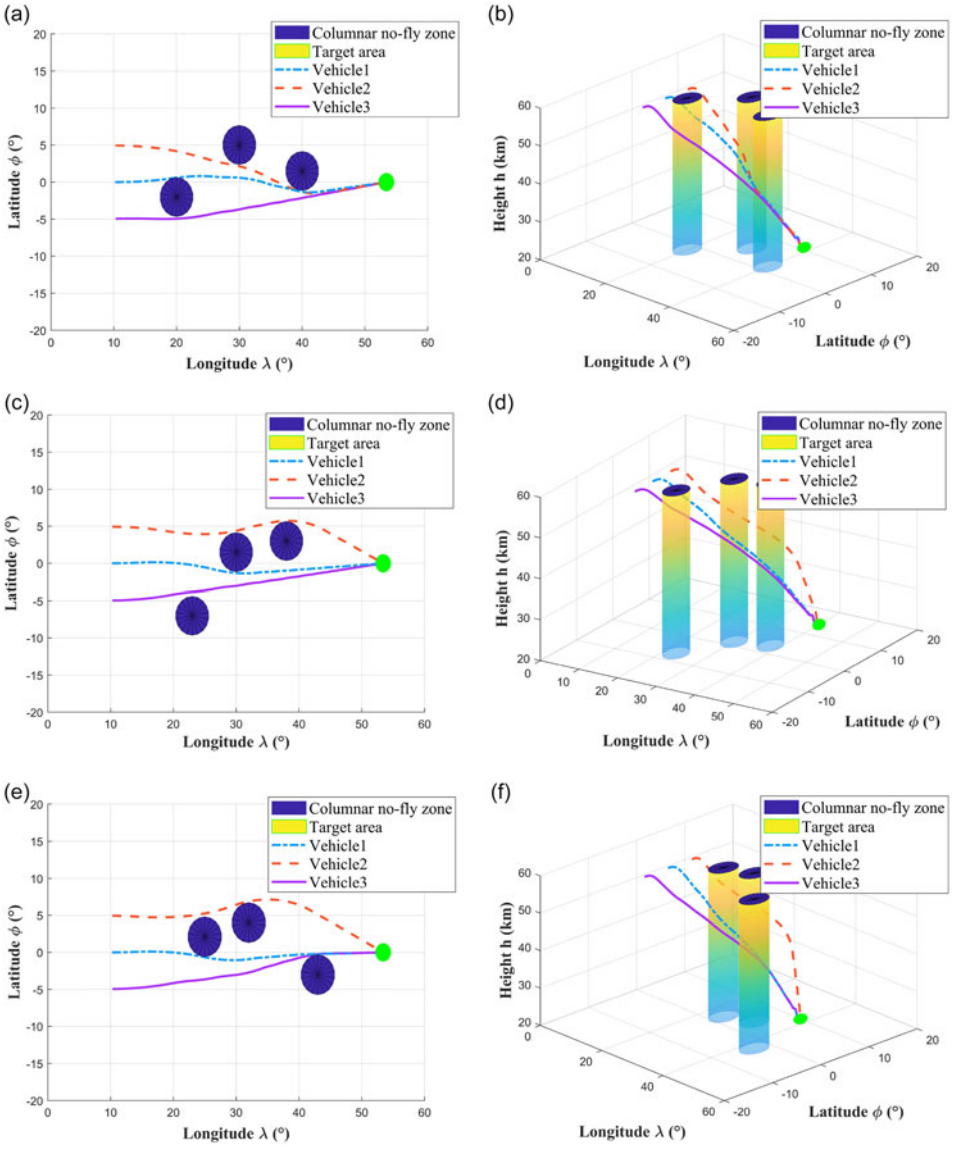


Figure 7. *Time co-simulation results of multi-hypersonic vehicles under different no-fly zones.*

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

Table 6. *Terminal state and error of multi-hypersonic vehicles under different no-fly zones*

	Definition	CAV-H#1	CAV-H#2	CAV-H#3
Case 1	Velocity error	0.23%	0.10%	0.16%
	Height error	3.80%	2.90%	4.00%
	Flight time error	0.43%	0.35%	0.43%
	Terminal coordinate	(52.51° E, 0.12° N)	(52.52° E, 0.15° S)	(52.52° E, 0.13° S)
Case 2	Velocity error	0.24%	0.09%	0.26%
	Height error	3.58%	3.41%	3.50%
	Flight time error	0.35%	0.70%	0.61%
	Terminal coordinate	(52.51° E, 0.05° S)	(52.60° E, 0.37° N)	(52.51° E, 0.10° S)
Case 3	Velocity error	0.07%	0.10%	0.08%
	Height error	4.02%	3.39%	4.32%
	Flight time error	0.43%	0.17%	0.70%
	Terminal coordinate	(52.51° E, 0.01° S)	(52.60° E, 0.38° N)	(52.52° E, 0.03° S)

different no-fly zones after one training. Compared with Ref. (23), it shows that the algorithm designed in this paper has better universality.

5.0 Conclusion

In this paper, a time-cooperative guidance method based on reinforcement learning for hypersonic vehicles under no-fly zone and other multi-constraint conditions is designed. In order to solve the problems of long training times and limited applicability in reinforcement learning time cooperative guidance, the predictor-corrector method is introduced to estimate the remaining time-of-flight and calculate the absolute value of the bank angle, which are then used as the observations of the reinforcement learning agent. By introducing the predictor-corrector method, the complex decision-making behaviour of the agent is simplified, requiring only the judgement of the bank angle sign at each time step, which significantly improves the training efficiency and reduces the number of training iterations. At the same time, reward function and action space are specifically designed for the tasks of no-fly zone penetration and time coordination, followed by mathematical simulations. Simulations under multiple constraint conditions and different no-fly zone distributions concluded that the proposed algorithm had good applicability, adaptability and robustness in completing the time cooperative no-fly zone penetration task. In the face of increasingly complex flight environments in the future, the operational space for aircraft will become further constrained. Therefore, incorporating dynamic no-fly zones or interceptors in future research would further enhance the applicability and robustness of the proposed algorithm.

Acknowledgements. This work was supported in part by the Key Laboratory of Spacecraft Design Optimization and Dynamic Simulation Technologies, Ministry of Education, in part by the Science and Technology on Complex System Control and Intelligent Agent Cooperation Laboratory, Beijing, 100074, China.

References

[1] Ding, Y., Yue, X. and Chen, G. Review of control and guidance technology on hypersonic vehicle, *Chin. J. Aeronaut.*, 2022, **35**, pp 1–18.

[2] Jeon, I-S., Lee, J-I. and Tahk, M-J. Homing guidance law for cooperative attack of multiple missiles, *J. Guid. Control Dyn.*, 2010, **33**, pp 275–280.

[3] Liang, Z. and Ren, Z. Tentacle-based guidance for entry flight with no-fly zone constraint, *J. Guid. Control Dyn.*, 2018, **41**, pp 996–1005.

[4] Gao, Y., Cai, G., Yang, X., et al. Improved tentacle-based guidance for reentry gliding hypersonic vehicle with no-fly zone constraint, *IEEE Access*, 2019, **7**, pp 119246–119258.

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press

[5] Wang, Z. Optimal trajectories and normal load analysis of hypersonic glide vehicles via convex optimization, *Aerosp. Sci. Technol.*, 2019, **87**, pp 357–358.
[6] Halbe, O., Raja, R.G. and Padhi, R. Robust reentry guidance of a reusable launch vehicle using model predictive static programming, *J. Guid. Control Dyn.*, 2014, **37**, pp 134–148.
[7] Zang, L., Lin, D., Chen, S., et al. An on-line guidance algorithm for high L/D hypersonic reentry vehicles, *Aerosp. Sci. Technol.*, 2019, **89**, pp 150–162.
[8] Zhao, T., Wang, P., Liu, L., et al. Integrated guidance and control with L2 disturbance attenuation for hypersonic vehicles, *Adv. Space Res.*, 2016, **57**, pp 2519–2528.
[9] Zhu, J., Liu, L., Tang, G., et al. Three-dimensional robust diving guidance for hypersonic vehicle, *Adv. Space Res.*, 2016, **57**, pp 562–575.
[10] Fang, K., Zhang, Q., Ni, K., et al. Time-coordinated reentry guidance law for hypersonic vehicle, *Acta Aeronaut. Astronaut. Sin*, 2018, **39**, pp 1–16.
[11] Zhang, Y., Fang, G., et al. Time-cooperative guidance law for multiple UAVs with angle constraints, 2019 *IEEE International Conference on Unmanned Systems (ICUS)*, 2019, pp 189–194.
[12] Li, Z., He, B., Wang, M., et al. Time-coordination entry guidance for multi-hypersonic vehicles, *Aerosp. Sci. Technol.*, 2019, **89**, pp 123–135.
[13] Yu, J., Dong, X., Li, Q., et al. Cooperative guidance strategy for multiple hypersonic gliding vehicles system, *Chin. J. Aeronaut.*, 2020, **33**, pp 990–1005.
[14] Hong, H., Maity, A. and Holzapfel, F. Free final-time constrained sequential quadratic programming–based flight vehicle guidance, *J. Guid. Control Dyn.*, 2021, **44**, pp 181–189.
[15] Yu, W., Chen, W., Jiang, Z., et al. Analytical entry guidance for coordinated flight with multiple no-fly-zone constraints, *Aerosp. Sci. Technol.*, 2019, **84**, pp 273–290.
[16] Zhan, Y., Xv, J., Yao, K., et al. A new UAV swarm pursuit task scheme driven by DDPG algorithm, *Acta Aeronaut*, 2020, **41**, pp 314–326.
[17] Luo, Y., Song, J., Zhao, K., et al. UAV-cooperative penetration dynamic-tracking interceptor method based on DDPG, *Appl. Sci.*, 2022, **12**, p 1618.
[18] Li, Z., Guo, J., Tang, S., et al. A deep learning-based approach to time-coordination entry guidance for multiple hypersonic vehicles, *Aeronaut. J.*, 2023, **127**, pp 604–626.
[19] Song, J., Xu, X., Tong, X., et al. A time cooperation guidance for multi-hypersonic vehicles based on LSTM network and improved artificial potential field method, *Aerospace*, 2022, **9**, p 562.
[20] Wang, Z., Wu, T., Zhu, Z. and Ma, C., Reinforcement learning–based adaptive attitude control method for a class of hypersonic flight vehicles subject to Nonaffine structure and unmatched disturbances, *J. Aerosp. Engin*, 2024, **37**, p 04024003.
[21] Xu, Y., Li, J., Wu, B., et al. Cooperative landing on mobile platform for multiple unmanned aerial vehicles via reinforcement learning, *J. Aerosp. Engin*, 2024, **37**, p 04023095.
[22] Chen, H., Yu, J. and Dong, X. A cooperative guidance law for multiple missiles based on reinforcement learning, 2022 *IEEE International Conference on Unmanned Systems (ICUS)*, 2022, pp 1473–1478.
[23] Zhuang, X., Li, D., Wang, Y., et al. Optimization of high-speed fixed-wing UAV penetration strategy based on deep reinforcement learning, *Aerosp. Sci. Technol.*, 2024, **148**, p 109089.
[24] Sana, K., and Hu, W. Hypersonic reentry trajectory planning by using hybrid fractional-order particle swarm optimization and gravitational search algorithm, *Chin. J. Aeronaut.*, 2021, **34**, pp 50–67.
[25] Zhao, J., and Zhou, R. Reentry trajectory optimization for hypersonic vehicle satisfying complex constraints, *Chin. J. Aeronaut.*, 2013, **26**, pp 1544–1553.
[26] Mnih, V., Kavukcuoglu, K., Silver, D., et al. Human-level control through deep reinforcement learning, *Nature*, 2015, **518**, pp 529–533.
[27] Van Hasselt, H., Guez, A., and Silver, D. Deep reinforcement learning with double q-learning, *Proceedings of the AAAI Conference on Artificial Intelligence*, 2016, **30**, pp 2094–2100.
[28] Zhang, W., Yu, W., Li, J., and Chen, W. Intelligent cross-range maneuver reentry cooperative guidance of aircraft based on longitudinal analytical solution, *Acta Armamentarii*, 2021, **42**, pp 1400–1411.

Cite this article: Teng B., Song J. and Wei M. (2026). Collaborative penetration guidance strategy for hypersonic vehicles based on double-DQN. *The Aeronautical Journal*, **130**, 508–525. https://doi.org/10.1017/aer.2025.10080

https://doi.org/10.1017/aer.2025.10080 Published online by Cambridge University Press