

Article

Control and Communication Co-Optimization Method with Handshake Frequency Hopping for Multi-AGVs

Jisong Yu ^{1,2}, Changqing Xia ^{1,*} , Yang Xiao ^{1,2}, Yueqi Li ^{1,2}, Chi Xu ¹  and Xi Jin ³ 

¹ State Key Laboratory of Robotics and Intelligent Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China; yujisong@sia.cn (J.Y.); xiaoyang24@mails.ucas.ac.cn (Y.X.); liyueqi@sia.cn (Y.L.); xuchi@sia.cn (C.X.)

² University of Chinese Academy of Sciences, Beijing 100049, China

³ School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China; xijin@ieee.org

* Correspondence: xiachangqing@sia.cn

Abstract

In dynamic, high-interference industrial and logistics environments, multi-AGV cooperative tasks are often affected by communication delays and data loss, leading to information staleness and reduced control accuracy. Traditional handshake frequency hopping communication strategies introduce additional overhead in high-load environments, and channel selection strategies struggle to adapt to dynamic changes. To address challenges related to communication delay, task coordination, and real-time information exchange, we propose a control and communication co-optimization method based on a nonlinear Age of Information (AoI) penalty and an adaptive handshake frequency hopping mechanism. The method constructs a coupled control-communication model, designs an adaptive handshake period and multi-channel frequency hopping strategy to reduce channel conflicts, and introduces a nonlinear AoI penalty function that prioritizes the update of critical timely information, improving communication success rates and path control accuracy. Furthermore, by integrating the differential dynamics model, state estimation under communication delay and control error modeling, we propose a cooperative optimization algorithm for perception control and communication based on nonlinear AoI optimization (PPO-CCBNA). The algorithm achieves efficient solution based on approximate policy optimization (PPO). Simulation results demonstrate that PPO-CCBNA significantly outperforms benchmark algorithms in communication success rates, control stability, and energy efficiency, validating its effectiveness and feasibility in complex multi-AGV cooperative tasks.

Keywords: multi-AGVs collaboration; handshake frequency hopping; nonlinear age of information; cooperative control; multi-agent deep reinforcement learning

MSC: 93C20; 94A15; 05C80



Academic Editor: Shih-Wei Lin

Received: 21 October 2025

Revised: 2 November 2025

Accepted: 12 November 2025

Published: 13 November 2025

Citation: Yu, J.; Xia, C.; Xiao, Y.; Li, Y.; Xu, C.; Jin, X. Control and Communication Co-Optimization Method with Handshake Frequency Hopping for Multi-AGVs. *Mathematics* **2025**, *13*, 3639. <https://doi.org/10.3390/math13223639>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In modern industrial and intelligent logistics systems, multi-autonomous guided vehicles (AGVs) rely on stable wireless communication networks to enable dynamic environmental sensing, real-time information sharing, and precise collaborative control, which are foundational to ensuring production efficiency, system safety, and reliable operation. These capabilities are crucial for optimizing collaborative tasks, such as material handling, autonomous transport, and system-wide synchronization. However, in high-interference,

dynamically changing environments, wireless communication links often suffer from signal attenuation, occlusion, and multipath effects. These environmental factors lead to communication delays, data loss, and information staleness, which severely degrade control accuracy, coordination efficiency, and the overall stability of collaborative operations [1,2]. As the number of devices increases and tasks grow more complex, these challenges become more pronounced, requiring more robust solutions to maintain system performance.

To enhance communication reliability and mitigate channel interference, handshake-based frequency hopping communication has become a widely adopted technique in modern wireless systems [3]. It disperses interference and reduces the probability of signal collision by dynamically switching between multiple preset communication channels, thereby maintaining a relatively stable low-latency communication connection in high-noise or signal-fading environments. However, this mechanism faces multiple challenges in practical deployment: frequent channel switching itself introduces additional communication overhead; the randomness or simple strategy of channel selection increases the uncertainty and complexity of the system state; especially in the scenarios with dense equipment and high communication load, its interference avoidance capability decreases, and it is difficult to optimize the end-to-end latency and the cooperative performance, so that the validity of the information faces a severe test [4–6].

How to improve the efficiency of handshake frequency hopping and guarantee the timeliness of information transmission is the key to break through the above problems and realize efficient collaboration. Traditional Age of Information (AoI)-based optimization methods reflect the timeliness of information in a system by measuring the time interval between its generation and its acceptance, but it is difficult to accurately portray the nonlinear negative impact of delay on system performance. To this end, this paper proposes a control and communication co-optimization framework based on nonlinear AoI optimization, taking the multi-autonomous guided vehicle (AGV) co-optimization scenario as an example. Specifically, this paper establishes a joint state-space model for control and communication coupling by coupling the differential dynamics model, the state estimation and communication hysteresis model, and the control error model, and designs the adaptive handshake period and the multichannel frequency hopping communication mechanism to enhance the system's anti-interference ability and communication efficiency in the complex environment. And in order to accurately portray the nonlinear negative impact of delay on system performance, a nonlinear AoI metric mechanism is introduced to more accurately portray the nonlinear impact of communication delay on system performance. Therefore, this paper establishes a communication performance optimization problem based on perceptual control. Finally, to solve the problem, this paper proposes a cooperative optimization algorithm for perception control and communication based on nonlinear AoI optimization (PPO-CCBNA), which is used to solve the above problem and to achieve precise trajectory control perception and communication optimization of industrial control systems in dynamic environments.

The main contributions of this work are summarized below:

- (1) Integration of system models and communication mechanism: this work integrates the differential dynamics model, state estimation and communication hysteresis model, and control error model of AGVs to accurately capture the coupling between system motion control and communication performance. Based on this integration, an adaptive handshake period and multi-channel frequency hopping mechanism are developed to optimize communication resource allocation, enhance anti-jamming capability, and improve communication efficiency while being aware of control stability requirements.

- (2) Nonlinear AoI optimization for real-time control: to improve the real-time responsiveness and control accuracy of the system, this paper introduces a nonlinear AoI optimization mechanism. By incorporating nonlinear AoI into the optimization objective and designing a corresponding nonlinear AoI penalty function, the paper effectively models the nonlinear impact of delay on system performance. This approach encourages the agent to prioritize information updates in high-delay environments, thus mitigating the negative effects of information lag on control accuracy and overall collaboration.
- (3) POMDP-based algorithm for perceptual control and communication optimization: to address the non-convex minimization problem, this paper formulates it as a partially observable Markov decision process (POMDP) and proposes a cooperative optimization algorithm for perception control and communication based on nonlinear AoI optimization (PPO-CCBNA). Comparative experiments with a benchmark algorithm validate the superiority of the proposed approach in terms of control stability, communication efficiency, and energy efficiency, particularly in multi-agent collaborative tasks, where it demonstrates significant advantages.

The work that follows is organized as follows. Section 2 briefly describes the related work. Section 3 gives the system model and presents the joint optimization problem for control and communication performance. Section 4 gives the transformation procedure for the MDP problem. Section 5 presents the proposed PPO-CCBNA algorithm. Section 6 verifies the accuracy and good performance of the proposed algorithm through extensive comparative experiments. Finally, Section 7 summarizes the whole work.

2. Related Work

In this section, reviews and critically analyzes existing research related to task and resource scheduling in MEC for IIoT, emphasizing three key areas: co-design methods in IIoT, task and resource scheduling research in MEC for IIoT, and AoI in time-sensitive IIoT. The text highlights current achievements and identifies unresolved challenges, including complex task dependencies, delay propagation effects, and the accurate portrayal of information freshness. It also summarizes research gaps to motivate proposed solutions.

2.1. Wireless Network Transmission Technology

Frequency hopping communication technology is widely used in complex radio environments to avoid channel interference and improve the stability and reliability of communication. Frequency hopping communications are capable of dynamically switching between multiple channels, maintaining low latency and high efficiency by reducing interference and signal collisions. However, frequent channel switching may bring additional communication overheads and computational burdens, especially in multi-agent collaborative tasks, where uncertainty in channel selection may increase the complexity of the system and affect the effectiveness of the collaboration. Yang et al. [7] proposed an efficient cooperative-free channel rendezvous algorithm for heterogeneous cognitive radio networks. Lou et al. [8] systematically modeled and optimized AoI for the first time in a multichannel, multihop wireless network. Wang et al. [9] proposed a hopping time estimation method for the case of missing observations. Han et al. [10] proposed a novel SIC-enhanced multihop state updating mechanism. Wang et al. [11] proposed a very resource efficient OTA aggregation mechanism.

2.2. Age of Information Optimization

AoI is a key metric to measure the timeliness of information update in wireless communication, which is widely used in wireless networks, industrial control systems, and multi-agent collaborative tasks. Many existing studies use linear AoI optimization methods

to reduce the impact of information lag on system performance. Zhang et al. [12] addressed the problem of timeliness analysis and optimization in multi-source state update system for IoT, and proposed a method to assess the system timeliness through the default probability of AoI and PAoI. Long et al. [13] proposed AoI-STO algorithm by optimizing the trajectory, sensing, and forwarding strategies of the UAVs. Zhang et al. [14] proposed an algorithm to minimize the energy consumption of AoI and UAVs through joint optimization of the data collection time, UAV trajectory, and time slot duration to minimize the energy consumption of AoI and UAV. Huang et al. [15] proposed a wireless information acquisition scheduling framework for AoI guarantees in response to the difficulty of satisfying strict AoI constraints with unknown channel reliability. Shi et al. [16] introduced AoI into ADS task scheduling for the first time and demonstrated that optimizing AoI can optimize RT and throughput at the same time. Lu et al. [17] proposed a joint design architecture of control and transmission based on full loop AoI. Liao et al. [18] proposed a joint sensing-communication-control optimization framework based on ultra-low AoI.

2.3. Co-Optimization of Wireless Network and Control Systems for Robot Clusters

With the development of industrial intelligence, collaborative optimization is increasingly used in wireless network systems for robot clusters. Most of the existing collaborative optimization methods focus on task allocation, path planning, and collaborative control. Wang et al. [19] designed a QoS-based co-design method for sensing, communication and control for UAV positioning system. Qiao et al. [20] designed a wireless communication and control co-optimization method based on resource allocation, communication scheduling, and control strategies in cloud-controlled AGV systems. Ma et al. [21] proposed a co-design method based on beamforming technique and control optimization for the communication interference and energy consumption problems in multi-loop wireless control systems. Girgis et al. [22] proposed a co-design method based on beamforming technique and control optimization for the cooperative control problems of multiple systems in wireless hybrid logic dynamic system species. Zhao et al. [23] proposed a communication and control co-design methodology to address the need for real-time control in future wireless networks. Han et al. [24] proposed a wireless communication and control co-design framework based on finite time wireless system identification. Xu et al. [25] developed a nonsingular predefined-time adaptive dynamic surface control method for quantized nonlinear systems. Sui et al. [26] proposed a finite-time adaptive fuzzy event-triggered consensus control scheme for high-order MIMO nonlinear multi-agent systems.

In summary, although some research progress has been made on the existing frequency hopping communication techniques, information timeliness optimization and cooperative optimization methods for wireless networks, there are still obvious deficiencies in dealing with communication delays, information lags and efficient cooperation. Existing frequency hopping mechanisms have the problem of computational burden caused by frequent channel switching in practical applications, the AoI optimization method fails to fully consider the nonlinear effect of communication delay, and the co-design scheme is difficult to balance the control accuracy and communication efficiency in the dynamic interference environment. In order to solve these problems, this paper designs a control and communication synergy method by combining nonlinear AoI optimization and adaptive handshaking mechanism so as to guarantee the real-time and stability requirements of the system.

3. System Modeling and Problem Formulation

3.1. System Model

In this paper, we consider an intelligent industrial system consisting of m AGVs as shown in Figure 1. All AGVs are distributed in a two-dimensional plane and are responsible for tasks such as cooperative handling and path tracking. In the process of task execution, AGVs interact with each other through frequency-hopping device-to-device (D2D) communication for status and task information to ensure cooperative operation and information synchronization among AGVs in the system. The communication system adopts a multi-channel frequency hopping mechanism, which can effectively improve the overall anti-interference capability of the system and enhance the utilization efficiency of resources, thus ensuring the efficient and stable operation of AGV in complex environments [27,28].

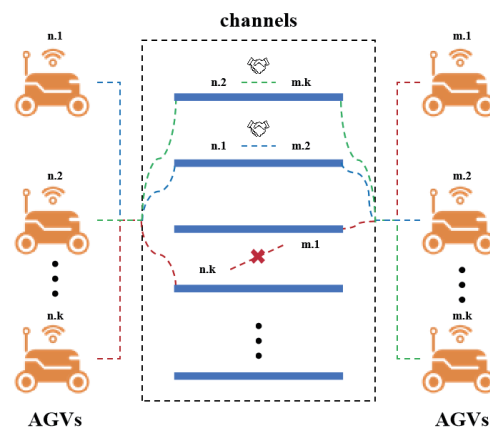


Figure 1. Multi-AGVs Intelligent Industrial System.

In time slot t , let $p_i(t) = (x_i(t), y_i(t))$ denote the two-dimensional position of the i -AGV in time slot t , its position in the two-dimensional plane. Let $\theta_i(t)$ denote the facing angle of the AGV, $v_i(t)$ and $\omega_i(t)$ are denoted as the linear and angular velocities of the AGV, respectively. On this basis, the evolution of the dynamics equation of the AGV can be expressed as follows:

$$\begin{cases} x_i(t+1) = x_i(t) + v_i(t) \cos \theta_i(t) \iota \\ y_i(t+1) = y_i(t) + v_i(t) \sin \theta_i(t) \iota \\ \theta_i(t+1) = \theta_i(t) + \omega_i(t) \iota, \end{cases} \quad (1)$$

where ι is the time step.

3.2. Communication Model

Let the set of available channels of the system be $\mathcal{F} = \{1, 2, \dots, F\}$. The channels selected by the i -AGV and the j -AGV at the moment of time are denoted as $f_i(t)$ and $f_j(t)$, respectively, and only when then $f_i(t) = f_j(t)$, the i -AGV and the j -AGV can interact with each other in terms of information. Accordingly, the direction angle of joint AGV propagation can be established as a basis for judging the frequency hopping of inter-AGV handshake as [29,30] follows:

$$\delta_{ij}^{(f,t)} = 1 \begin{bmatrix} f_i(t) = f_j(t) \\ \cos(\theta_i(t) - \phi_{i \rightarrow j}(t)) \geq \cos(\Omega \theta_{\max}) \end{bmatrix}, \quad (2)$$

where, $\phi_{i \rightarrow j}(t)$ denotes the direction angle from i -AGV to j -AGV, and $\Omega\theta_{\max}$ is the directional judgment threshold. Only when $f_i(t) = f_j(t)$ and $\cos(\theta_i(t) - \phi_{i \rightarrow j}(t)) \geq \cos(\Omega\theta_{\max})$, do we have $\delta_{i,j}^{(f,t)} = 1$, indicating that i -AGV and j -AGV successfully establish a handshake at time t .

Due to the small amount of data per communication packet, strict transmission delay and BER requirements in industrial control tasks, the traditional Shannon channel capacity, which assumes an infinitely long packet and an infinitesimal BER, is not applicable to the actual environment of this system. In order to truly reflect the communication performance under finite packet length and actual BER, this paper adopts the finite block length transmission model [31] in order to calculate the actual transmission rate of the i -AGV and the j -AGV at time slot t and channel f :

$$R_{i,j}^{(f,t)} = \delta_{i,j}^{(f,t)} \cdot \left[B \log_2 \left(1 + \frac{p_{i,j}^{(f,t)} g_{i,j}^{(f,t)}}{\sum_{k \neq i, f' \in \mathcal{F}} p_{k,j}^{(f',t)} g_{k,j}^{(f',t)} + \sigma^2} \right) + \sqrt{\frac{V_{i,j}^{(f,t)}}{n_f}} Q^{-1}(\epsilon) \right], \quad (3)$$

where B is the channel bandwidth, $p_{i,j}^{(f,t)}$ is the transmit power, $g_{i,j}^{(f,t)}$ is the channel gain, σ^2 is the noise power, n_f is the block length, ϵ is the block BER, Q is the inverse of the Gaussian function and $V_{i,j}^{(f,t)}$ is the channel dispersion can be expressed as

$$V_{i,j}^{(f,t)} = 1 - \frac{1}{\left(1 + \frac{p_{i,j}^{(f,t)} g_{i,j}^{(f,t)}}{\sum_{k \neq i, f' \in \mathcal{F}} p_{k,j}^{(f',t)} g_{k,j}^{(f',t)} + \sigma^2} \right)^2}. \quad (4)$$

For the i -AGV and the j -AGV, the information transmission delay is calculated as

$$T_{i,j}^{(f,t)} = \delta_{i,j}^{(f,t)} \frac{L_{i,j}^{(f,t)}}{R_{i,j}^{(f,t)}} + \varsigma, \quad (5)$$

where $L_{i,j}^{(f,t)}$ is the packet size and $\varsigma > 0$ is a very small positive number used to ensure that the denominator is always non-zero, and is introduced to enhance the robustness of the numerical calculations and maintain dimensional consistency.

The corresponding communication energy consumption can be calculated as

$$E_{i,j}^{(f,t)} = p_{i,j}^{(f,t)} T_{i,j}^{(f,t)}. \quad (6)$$

However, in high-density communication scenarios where the number of AGVs is much larger than the number of channels, it can lead to communication packet loss and message aging problems due to channel conflicts. For this reason, this paper introduces a time-division multiplexing mechanism as a complementary strategy in the system communication framework. When the system detects that channel resources are tight or conflicts are frequent, it switches to the time-division multiplexing mode and allocates communication time slots to different AGV communication pairs on the time axis, thus avoiding conflicts in the time domain and guaranteeing the timeliness and reliable transmission of mission-critical information.

3.3. Age of Information

Age of Information (AoI), as an important indicator of information timeliness in a system, has been widely used in various scenarios that require real-time data interaction,

and is especially suitable for wireless communication and control systems that have high requirements for information timeliness [32,33]. It is used to characterize the time interval that elapses between the time a message is generated and the time it is received, thus reflecting the freshness of the information currently in use. Specifically, for a task, assuming its data sampling time is τ_t , the Age of Information at the current moment t can be defined as

$$\Delta_t = t - \tau_t. \quad (7)$$

The timeliness of the information is crucial for the co-optimization of trajectory control and information communication. It directly affects the coordination and cooperation among AGV clusters, and the effect of information staleness on control accuracy and system stability will show a trend of nonlinear exacerbation as the communication delay increases. Specifically, when communication delays are long, inaccurate information may lead to erroneous decisions, which may trigger path deviations, improper task execution, or even system failures. As a result, traditional AoI, while reflecting the staleness of the information, is unable to accurately portray the negative impact of latency on system performance in the presence of high latency.

In order to more accurately reflect the negative effects of delay, a nonlinear AoI is introduced. As shown in Figure 2, in high-latency scenarios, nonlinear AoI can more effectively direct the system to prioritize states that are affected by greater latency by imposing greater penalties on older information. This nonlinear penalty mechanism enables the system to identify and prioritize the correction of situations with severe information lag in a timely manner, avoiding AGV path deviation or communication failure due to inaccurate or outdated information. Specifically, nonlinear AoI improves the sensitivity of the system to communication delays by more strongly penalizing those states with higher information ages, thus effectively improving control accuracy and system stability. In systems with stringent requirements for high-precision control and efficient collaboration, nonlinear AoI is better adapted to these needs, ensuring that the system can make more accurate and coordinated decisions in the face of delays [34].

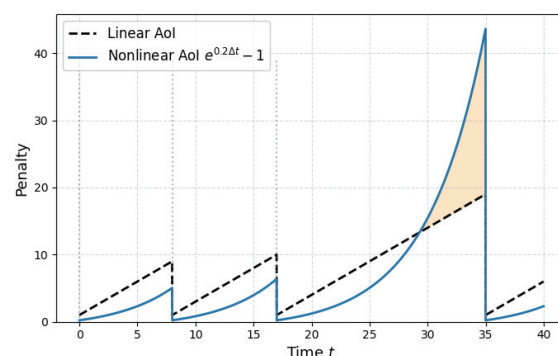


Figure 2. Comparison between Nonlinear AoI Function and Linear AoI Function.

In addition, the traditional definition of AoI only measures the time interval between when a message is generated from sampling and when it is received, and does not take into account the time overhead required for communication connection establishment (i.e., networking). However, in the system studied in this paper, the communication pairs between AGVs are not predefined, but the communication targets are dynamically decided after the generation of information packets, and the networking pairing is accomplished through the frequency-hopping handshake mechanism. Therefore, the whole process from information generation, communication pair determination, networking establishment, to successful data transmission is covered in the AoI considered in this paper.

3.4. Control Model

Since there is a communication delay for information synchronization between AGVs, we perform the computation of state estimation by introducing AoI, and the state estimation delay can be expressed as

$$\tau = \Delta_t \cdot l, \quad (8)$$

where Δ_t is the age of the information.

The current state estimate of the i -AGV is denoted as

$$\begin{cases} \hat{x}_i(t) = x_i(t - \tau) + v_i(t - \tau) \cos \theta_i(t - \tau) \tau, \\ \hat{y}_i(t) = y_i(t - \tau) + v_i(t - \tau) \sin \theta_i(t - \tau) \tau. \end{cases} \quad (9)$$

Define the trajectory error as

$$\tilde{\varepsilon}_i(t) = \begin{bmatrix} x_r^i(t) - \hat{x}_i(t) \\ y_r^i(t) - \hat{y}_i(t) \end{bmatrix}, \quad (10)$$

where $x_r^i(t)$ and $y_r^i(t)$ are reference trajectories.

Therefore, this design control law is

$$\begin{cases} v_i(t) = k_v \cdot \|\tilde{\varepsilon}_i(t)\| \\ \omega_i(t) = k_\omega \cdot (\text{atan2}(\tilde{y}, \tilde{x}) - \theta_i(t)), \end{cases} \quad (11)$$

where k_v and k_ω are proportional gains.

To evaluate the control performance, the control stability function is introduced as

$$\kappa_i(t) = \frac{1}{\alpha \cdot \|\tilde{\varepsilon}_i(t)\|^2 + \beta(v_i^2(t) + \omega_i^2(t)) + \chi}, \quad (12)$$

where α and β are weighting factors and $\chi > 0$ is a very small positive number to ensure that the denominator is bounded.

In order to improve the overall performance of the system, an adaptive handshake frequency adjustment function is introduced:

$$\eta_i(t) = \left(\omega_1 \|v_i(t)\| + \omega_2 \|x_i(t) - x_i(t-1)\| + \omega_3 p(\Delta_t) + \omega_4 \left(\frac{1}{\kappa_i(t)} \right) \right)^{-1}, \quad (13)$$

where $\omega_1, \omega_2, \omega_3, \omega_4$ are weighting factors.

Therefore, we can modify the handshake frequency hopping criterion to

$$\delta_{i,j}^{(f,t)} = 1 \begin{bmatrix} f_i(t) = f_j(t) \\ c_i(t) \geq \eta_i(t) \\ \cos(\theta_i(t) - \phi_{i \rightarrow j}(t)) \geq \cos(\Omega \theta_{\max}) \end{bmatrix}, \quad (14)$$

where $c_i(t)$ is the handshake frequency beat counter.

3.5. Description of the Problem

The focus of this paper is to minimize the system energy consumption, nonlinear AoI and maximize the system control stability under the constraints imposed on the system by considering the communication model and the control model. Therefore the following problem is constructed:

$$P: \min \sum_t \left[\lambda_1 \sum_{i,j} E_{i,j}^{(f,t)} + \lambda_2 \sum_{i,j} p(\Delta_{i,j}^t) - \lambda_3 \sum_i \kappa_i(t) \right] \quad (15)$$

$$\begin{aligned} \text{s.t. } C1: & 0 \leq p_{i,j}^{(f,t)} \leq p_{\max}, \\ C2: & \eta_i(t) \leq \eta_{\max}, \\ C3: & x_{\min} \leq x_i(t) \leq x_{\max}, \\ C4: & y_{\min} \leq y_i(t) \leq y_{\max}, \\ C5: & 0 \leq v_i(t) \leq v_{\max}, \\ C6: & |\omega_i(t)| \leq \omega_{\max}, \\ C7: & 0 \leq t \leq t_{\max}, \end{aligned} \quad (16)$$

where λ_1 , λ_2 , and λ_3 are the corresponding weight coefficients, all belonging to the range $[0, 1]$, and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. And constraint C1 is the maximum transmit power constraint of the AGV. C2 is the maximum adaptive handshake frequency regulation function constraint to ensure that it can be efficiently regulated within the constraint range. C3 and C4 are the positional boundary constraints of the AGV, and C5 and C6 are the velocity constraints of the AGV to ensure that it can perform its motions within feasible ranges and velocity requirements. C7 is the maximum handshake pairing period constraint.

4. Problem Formalization and Transformation

Since Problem $\mathcal{P}$ is a mixed integer nonlinear programming problem that is NP-hard. Therefore, it is necessary to transform the collaborative optimization into a POMDP problem and solve it using multi-agent deep reinforcement learning. Therefore, it is possible to construct an integrated decision model consisting of elements such as $\{S, A, Z, O, P, R\}$. Where $\mathcal{S}$ denotes the state space, $\mathcal{A}$ denotes the action space, $\mathcal{Z}$ is the state transfer probability matrix, $\mathcal{O}$ is the observation space, $\mathcal{P}$ is the observation probability, and $\mathcal{R}$ is the reward function.

4.1. State Space $\mathcal{S}$

At each decision moment t the state space consists of AGV position coordinates, AGV facing angle, AGV velocity, channel state, and AoI, which is defined as

$$\begin{aligned} \mathcal{S}(t) &= \{\mathcal{S}_i(t)\}_{i=1}^M \\ &= \{\{x_1(t), y_1(t), \theta_1(t), v_1(t), \omega_1(t), f_1(t), \Delta_1(t)\}, \dots \\ &\quad \{x_m(t), y_m(t), \theta_m(t), v_m(t), \omega_m(t), f_m(t), \Delta_m(t)\}\}, \end{aligned} \quad (17)$$

where $\mathcal{S}_i(t)$ is the state space of the i -AGV.

4.2. Action Space $\mathcal{A}$

At each decision moment t the state space consists of channel number selection and transmit power, defined as

$$\mathcal{A}(t) = \{\mathcal{A}_i(t)\}_{i=1}^M = \{\{f_1(t), p_1(t)\}, \dots, \{f_m(t), p_m(t)\}\}, \quad (18)$$

where $\mathcal{A}_i(t)$ is the action space of the i -AGV.

4.3. State Transition Probability $\mathcal{Z}$

At each decision period t , the state transfer probability $\mathcal{Z}_i(t)$ describes the probability that $\mathcal{S}_i(t)$ transfers to $\mathcal{S}_i(t+1)$ when action $\mathcal{A}_i(t)$ is performed, and is denoted by $\mathcal{Z}_i(t)(\mathcal{S}_i(t+1); \mathcal{S}_i(t), \mathcal{A}_i(t))$.

4.4. Observation Space $\mathcal{O}$

The observation environment is set as a partially observable environment, i.e., each agent can only access part of the state information, so the observation information is defined as

$$\mathcal{O}(t) = \{\mathcal{O}_i(t)\}_{i=1}^M = \{\{\tilde{x}_1(t), \tilde{y}_1(t), g_1(t), \Delta_1(t)\}, \dots, \{\tilde{x}_m(t), \tilde{y}_m(t), g_m(t), \Delta_m(t)\}\}, \quad (19)$$

where $\tilde{x}_m(t)$ and $\tilde{y}_m(t)$ are position information with noise interference and $g_m(t)$ is the channel gain of the current communication link.

4.5. Observation Probability $\mathcal{P}$

Observation probability is the probability that agent observes a particular observation given the system state and action, defined as follows:

$$\mathcal{P}_i(t) = \Pr(\mathcal{O}_i(t) | \mathcal{S}_i(t), \mathcal{A}_i(t-1)). \quad (20)$$

4.6. Reward $\mathcal{R}$

Based on the objective function given and considering matching failure, load balancing, and channel conflicts, the reward function is defined as

$$\mathcal{R}_i(t) = \lambda_3 \kappa_i(t) - \lambda_1 E_{i,j}^{(f,t)} - \lambda_2 p(\Delta_{i,j}^t) - P_{\text{fail},i}(t) - P_{\text{balance},i}(t) - P_{\text{collision},i}(t), \quad (21)$$

where $P_{\text{fail},i}(t)$ is the matching failure penalty, $P_{\text{balance},i}(t)$ is the load imbalance penalty, and $P_{\text{collision},i}(t)$ is the channel conflict penalty.

5. Collaborative Multi-Agent Control and Communication Algorithm Based on Nonlinear AoI Optimization

On the basis of the POMDP model for co-optimization of AGV network control and communication established above, in order to effectively solve the problems of state estimation delay, channel dynamic matching, and co-optimization between communication energy consumption and AoI for multi-AGV systems in industrial scenarios, this paper proposes a cooperative optimization algorithm for perception control and communication based on nonlinear AoI optimization (PPO-CCBNA), which is based on proximal policy optimization (PPO). The overall structure of the algorithm is shown in Figure 3. The algorithm adopts the framework of centralized training and distributed execution, fully combines the advantages of multi-intelligence body reinforcement learning and proximal policy optimization techniques, and utilizes the Actor-Critic architecture to achieve efficient optimization of policies. Specifically, each agent outputs its actions through a shared Actor network, which makes decisions on communication channel selection and transmit power control, respectively, while the Critic network is responsible for evaluating the value of the state in which the action is located, providing generalized advantage estimates (GAE) to enhance the stability of the training. In addition, the PPO-Clip technique is used to constrain the policy update amplitude, and gradient clipping is used to prevent gradient explosion, which guarantees the high efficiency and stability of the training process. At the same time, the algorithm introduces a strategy entropy term, which encourages diversification of

actions, effectively prevents the strategy from falling into local optimality, and significantly improves the overall control accuracy, communication efficiency and stability.

5.1. Algorithm Design

This algorithmic framework trains two networks for each agent: the Actor and the Critic. The Critic network learns to map the global state to its corresponding value function, providing evaluation for the current situation. Meanwhile, the Actor network is responsible for learning a probability distribution that maps local observations to possible actions, thereby enabling action selection through sampling. By adopting this design, the framework supports centralized training and distributed execution, allowing multiple agents to make both coordinated and autonomous decisions. The policy optimization is achieved by gathering trajectory data and utilizing it to iteratively update the network parameters. The detailed steps of the algorithm are illustrated in Algorithm 1.

Algorithm 1 PPO-CCBNA-Based Data Collection and Training

Input: Initialize $x_i(t), y_i(t), \theta_i(t), v_i(t), \omega_i(t), f_i(t)$, $\Delta_i(t)$ of the system parameters;
Output: $\pi_{\theta}^*, V_{\phi}^*, f_i(t), p_i(t)$;

- 1: Initialize the parameter θ for policy model π_{θ} and the parameter ϕ for value function V_{ϕ}
- 2: **for** episode = 1, 2, $\dots$, E **do**
- 3: set data buffer $D = \{\}$
- 4: **for** $w = 1, 2, \dots, W$ **do**
- 5: Set empty list $\tau = \{\}$
- 6: **for** $t = 1, 2, \dots, T$ **do**
- 7: **for** agent $i = 1, 2, \dots, M$ **do**
- 8: $P_t^{(i)} = \pi_{\theta}^{(i)}(a_t^{(i)} | o_t^{(i)})$
- 9: $a_t^{(i)} \sim P_t^{(i)}$
- 10: $V_t^{(i)} = V_{\phi}(s_t^{(i)})$
- 11: **end for**
- 12: Execute action $a(t) \rightarrow R(t), s(t+1), o(t+1)$
- 13: $\tau += \{s(t), o(t), a(t), R(t), s(t+1), o(t+1)\}$
- 14: **end for**
- 15: Compute reward $\mathcal{R}$ on τ
- 16: Compute advantage estimate $\hat{A}$ via GAE on τ
- 17: Compute $\tilde{V}_{\phi}(s_t^{(i)})$ with value-norm on τ
- 18: Split trajectory τ into chunks of length L
- 19: **for** $l = 1, 2, \dots, \frac{T}{L}$ **do**
- 20: $D = D \cup (\tau[l : l + T], \mathcal{R}[l : l + L], \hat{A}[l : l + L])$
- 21: **end for**
- 22: **end for**
- 23: **for** mini-batch $r = 1, 2, \dots, R$ **do**
- 24: $b \leftarrow$ random mini-batch from D with all agent data
- 25: **for** each data chunk c in the mini-batch b **do**
- 26: Update V
- 27: **end for**
- 28: **end for**
- 29: Adam update θ on $\mathcal{L}(\theta)$ with data b
- 30: Adam update ϕ on $\mathcal{L}(\phi)$ with data b
- 31: **end for**

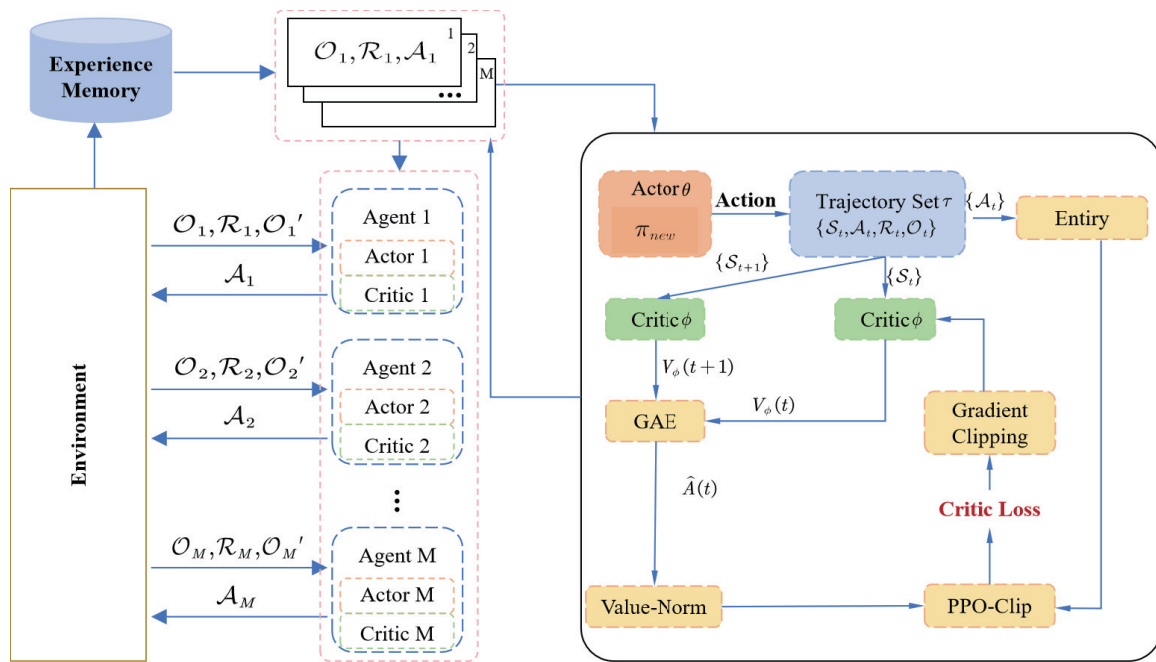


Figure 3. Structure of PPO-CCBNA.

Under the reinforcement learning framework, the agent continuously interacts with the environment to learn an optimal policy that maximizes cumulative rewards. To enhance the accuracy of Advantage function estimation and stabilize the policy optimization process, Generalized Advantage Estimation is adopted. The GAE is defined as follows:

$$\hat{A}_i(t) = \sum_{k=0}^{T-t-1} (\gamma\lambda)^k \zeta_i(t+k), \quad (22)$$

where λ is denoted as the attenuation factor of the GAE and $\zeta_i(t)$ is the time error denoted as

$$\delta_i(t) = R_i(t) + \gamma V_\phi(s_i(t+1)) - V_\phi(s_i(t)), \quad (23)$$

where $R_i(t)$ localizes the reward value, and $V_\phi(s_i(t))$ and $V_\phi(s_i(t+1))$ are the value functions of the current state $s_i(t)$ and the next moment state $s_i(t+1)$, respectively. Therefore, the objective value function can be estimated by generalized advantage estimation can be expressed as

$$\hat{V}_\phi(s_i(t)) = V_\phi(s_i(t)) + \hat{A}_i(t). \quad (24)$$

In order to further reduce the estimation error in a multi-agent setting, the dominance estimation is extended to the entire population of agent by introducing a value function normalization technique. The value function $\tilde{V}_\phi(s_i(t))$ after the normalization process is

$$\tilde{V}_\phi(s_i(t)) = \frac{\hat{V}_\phi(s_i(t)) - \hat{\mu}}{\hat{\sigma}}, \quad (25)$$

$$\hat{\mu} = \frac{1}{MT} \sum_{i=1}^M \sum_{t=1}^T \hat{V}_\phi(s_i(t)), \quad (26)$$

$$\hat{\sigma} = \frac{1}{MT} \sum_{i=1}^M \sum_{t=1}^T (\hat{V}_\phi(s_i(t)))^2, \quad (27)$$

where $\hat{\sigma} = \sqrt{\hat{\sigma} - \hat{\mu}^2}$ standard deviation, M denotes the number of agent, and T denotes the trajectory collection time. Accordingly, it can be obtained that the objective function of Actor network can be expressed as

$$\begin{aligned} \mathcal{L}(\theta) = & \frac{1}{WMT} \sum_{w=1}^W \sum_{i=1}^M \sum_{t=1}^T \\ & \min \left(r_i(t) \hat{A}_i(t), \text{clip}(r_i(t), 1 - \varepsilon, 1 + \varepsilon) \hat{A}_i(t) \right) \\ & + \frac{1}{WMT} \sum_{w=1}^W \sum_{i=1}^M \sum_{t=1}^T \varepsilon H(\pi_\theta), \end{aligned}$$

where W denotes the batch size, ε is the coefficient of the clipping function, ε is the coefficient of the strategy entropy regularization term, and $r_i(t)$ is the update strategy ratio, calculated as

$$r_{w,k}^{(i)} = \frac{\pi_\theta(a_i(t) | o_i(t))}{\pi_{\theta_{\text{old}}}(a_i(t) | o_i(t))}, \quad (28)$$

where $H(\pi_\theta)$ is the strategy entropy function, which can be expressed as

$$H(\pi_\theta) = \mathbb{E}_{a \sim \pi_\theta(\cdot | o_i(t))} [-\log \pi_\theta(a | o_i(t))]. \quad (29)$$

Similarly, the objective function of the Critic network can be expressed as

$$\mathcal{L}(\phi) = \frac{1}{WMT} \sum_{w=1}^W \sum_{i=1}^M \sum_{t=1}^T (V_\phi(s_i(t)) - \tilde{V}_\phi(s_i(t)))^2. \quad (30)$$

5.2. Algorithm Analysis

In the algorithm, the overall complexity of the algorithm needs to be calculated based on the complexity of each agent, since all agent are trained in parallel. Actor network consists of one input layer, one output layer and three hidden layers where the number of neurons in each layer is d_{ai} , d_{ao} and d_{ah} respectively. Therefore, the complexity of the Actor network can be calculated as $\mathcal{O}_{Actor} = (d_{ai}d_{ah} + 2d_{ah}d_{ah} + d_{ah}d_{ao})$. The same Critic network consists of one input layer, one output layer and three hidden layers, where the number of neurons in each layer is d_{ci} , d_{co} and d_{ch} . Therefore, its computational complexity is $\mathcal{O}_{Critic} = (d_{ci}d_{ch} + 2d_{ch}d_{ch} + d_{ch})$. Let the training E set, each set contains T step and $d_{ah} = d_{ch} = d_h$. Therefore, the overall complexity of the algorithm can be obtained as $\mathcal{O} = \mathcal{O}_{Actor} + \mathcal{O}_{Critic} = (ETd_h(d_{ai} + dn_h + d_{ao} + d_{ci} + 1))$. Therefore, it can be shown that the complexity of the algorithm is polynomial, indicating that it is executable.

6. Experimental Results

In this section, in order to evaluate the performance of the PPO-CCBNA algorithm, we build a simulation platform for a comprehensive evaluation.

6.1. Experimental Setup

(1) Experimental environment and parameter settings: For the PPO-CCBNA algorithm proposed in this paper, the specific parameters are set as follows: the number of neurons in the hidden layer of the Actor network and the Critic network are both set to 256, and the discount factor is 0.95. The learning rate of the two networks is uniformly set to 0.0005, and the cropping ratio in the PPO algorithm is set to 0.2 in order to improve the stability of the training and to prevent the strategy from updating too large. And in order to simulate the dynamic interference in industrial and logistic environments, a Rayleigh fading model is used to simulate the multipath propagation characteristics in

such environments. Additionally, a random interference factor is introduced to simulate unpredictable disturbances in the environment, adding variability to the communication channel. This factor models interference that could arise from external sources or changing environmental conditions, which impacts the overall system performance. The hardware platform used for the experiments is Intel i9-14900ks processor with NVIDIA RTX 4080 SUPER 16 GB graphics card (manufactured by ASUS in Taiwan, China), and the software environment is Python 3.8.20 and PyTorch 2.1.2 + CUDA 12.1. The weight coefficients for the optimization problem are set as follows: $\lambda_1 = 0.3$ (weight for energy consumption), $\lambda_2 = 0.2$ (weight for nonlinear AoI penalty), and $\lambda_3 = 0.5$ (weight for system stability). These coefficients are chosen to balance the trade-off between energy efficiency, information freshness, and system stability, with a higher emphasis placed on stability in the system. Other key parameters are set in Table 1 [35,36].

Table 1. System Parameters.

Parameter	Value
Number of AGVs	10
Number of Channels	9
$P_{\max}$	200 mW
σ^2	10^{-11} mW
B	20 MHz
Path Loss Exponent	3
L	800 Byte
n_f	128
ϵ	10^{-4}

(2) Algorithms for comparison: In order to consider the fairness of the experiment, the following benchmark algorithm was set up for the control test.

- PPO-CCBNA: Joint optimization algorithm for handshake communication based on MAPPO algorithm with nonlinear AoI perception.
- TD3-CCBNA: Joint optimization algorithm for handshake communication based on MATD3 algorithm with nonlinear AoI sensing.
- DDPG-CCBNA: Joint optimization algorithm for handshake communication based on MADDPG algorithm with nonlinear AoI perception.
- SAC-CCBNA: Joint optimization algorithm for handshake communication based on ISAC algorithm with nonlinear AoI perception.
- PPO-CCBA: Joint optimization algorithm for handshake communication based on MAPPO algorithm with AoI awareness.

6.2. Algorithmic Performance Analysis

Figure 4 demonstrates the training reward variation curve, and the proposed PPO-CCBNA algorithm significantly outperforms the other algorithms in terms of reward values. Its reward curve is smooth and less fluctuating, showing high stability during strategy updates. Compared with the TD3-CCBNA, SAC-CCBNA and DDPG-CCBNA algorithms, the PPO-CCBNA algorithm is able to achieve higher convergence reward values in a shorter time, which suggests that the PPO-CCBNA algorithm possesses a more efficient ability to explore and exploit the optimal control and communication performance. By introducing a nonlinear AoI-aware mechanism, PPO-CCBNA is able to optimize the communication strategy between intelligences more efficiently, further enhancing the overall performance. Therefore, PPO-CCBNA demonstrates significant superiority over other algorithms.

Table 2 compares the algorithm training, loading, and execution delays of PPO-CCBNA, TD3-CCBNA, DDPG-CCBNA, and SAC-CCBNA under different agent counts.

While PPO-CCBNA does not have the lowest training and loading delays compared to the other algorithms, these higher offline delays do not affect its online execution performance. The execution delay of PPO-CCBNA remains low and competitive, making it suitable for real-time applications. As the number of agents increases, the delays for all algorithms grow, but PPO-CCBNA continues to deliver efficient performance. These results suggest that PPO-CCBNA offers a good balance between training efficiency and low-latency execution, particularly when the number of agents scales up.

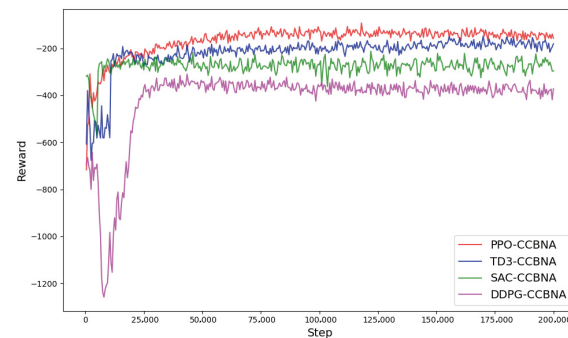


Figure 4. Reward curves based on different strategies and algorithms.

Table 2. Comparisons for Algorithmic Delay.

Agent	Algorithm	Training Delay (s)	Loading Delay (s)	Execution Delay (s)
10	PPO-CCBNA	1069.14	0.0852	0.0021
	TD3-CCBNA	6844.50	0.1309	0.0030
	DDPG-CCBNA	736.03	0.1052	0.0025
	SAC-CCBNA	1249.37	0.0209	0.0025
20	PPO-CCBNA	1947.82	0.0981	0.0024
	TD3-CCBNA	10,844.50	0.1505	0.0034
	DDPG-CCBNA	1252.52	0.1217	0.0028
	SAC-CCBNA	2471.03	0.0234	0.0028

Given that the industrial environment in which the AGV system is located has obvious dynamic change characteristics, and the state of each intelligent body, the mission objectives and the communication requirements will change in each round of mission execution, this paper adopts the information statistics results in terms of each round as the performance evaluation index. Based on the variation of energy consumption of different algorithms as shown in Figure 5, the energy consumption level of PPO-CCBNA algorithm is significantly lower than other algorithms, which is mainly due to the fact that the PPO-CCBNA algorithm adopts a more efficient mechanism of policy updating and exploring, which is combined with the clipping method to stabilize the optimization policy. Through this mechanism, PPO-CCBNA is able to effectively avoid redundant actions and unnecessary wastage of communication power, thus significantly reducing the overall energy consumption while ensuring performance.

Figure 6 demonstrates the control stability variation curves based on different algorithms, and the PPO-CCBNA algorithm exhibits higher stability. This is mainly due to the efficient exploration of the state space and the effective estimation of the global value function of the PPO-CCBNA algorithm. Through this approach, PPO-CCBNA is able to predict the system behavior more accurately and control it effectively, thus avoiding drastic fluctuations in the strategy and ensuring the smoothness and stability of the control process.

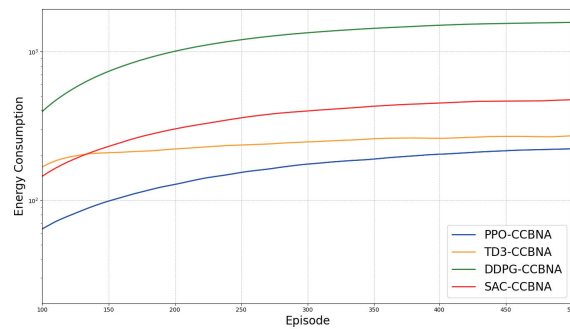


Figure 5. Energy consumption under different strategies.

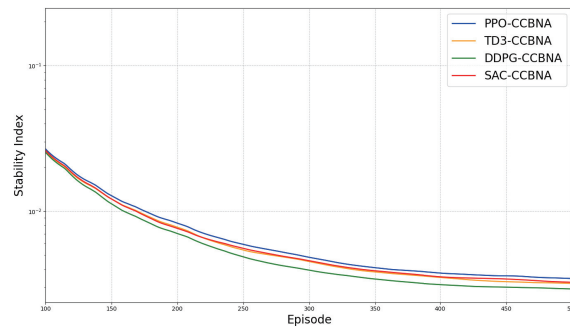


Figure 6. Control stability under different strategies.

Figure 7a and 7b show the nonlinear AoI penalty and AoI change curves based on different algorithms, respectively, and the PPO-CCBNA algorithm performs well in reducing AoI. Its excellent performance is mainly due to the introduction of the nonlinear AoI-aware mechanism, which imposes stronger penalties on higher AoI values, thus motivating the intelligences to be more active in data communication tasks and update information in a timely manner. Through this mechanism, PPO-CCBNA is able to effectively reduce the overall AoI value, ensuring the freshness of information and communication efficiency.

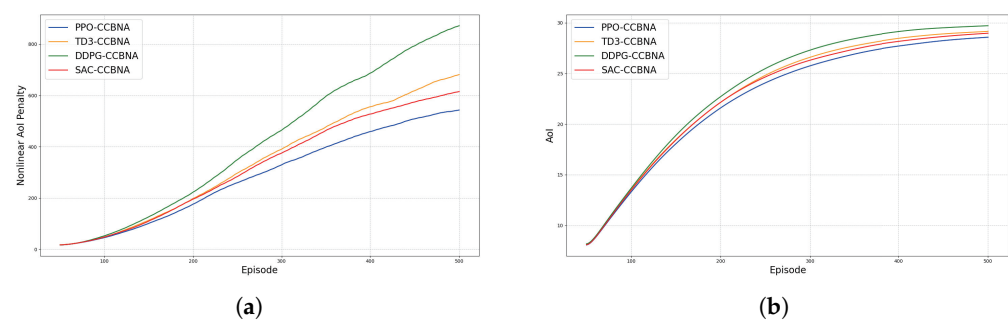


Figure 7. Nonlinear AoI Penalty and AoI Change. (a) Nonlinear AoI penalty as a function of system variables. (b) AoI dynamics across different scheduling strategies.

Figure 8a presents the handshake success index across various algorithms, while Figure 8b displays the channel conflict index. The handshake success index measures the proportion of successful handshake matches to the total number of attempted matches, whereas the channel conflict index quantifies how often agents select the same communication channel. The PPO-CCBNA algorithm achieves a notably higher handshake success rate compared to other methods. This improvement is mainly due to PPO-CCBNA's enhanced ability to learn effective communication pairing strategies, particularly in making more precise handshake-related decisions. By integrating a reward signal for successful handshakes during policy optimization, PPO-CCBNA incentivizes agents to optimize their communication behaviors, leading to much greater pairing efficiency. Additionally,

PPO-CCBNA demonstrates the lowest channel collision rate among all tested algorithms. This advantage arises from its superior capability to manage spectrum resource allocation among multiple agents through efficient action mapping and policy gradient updates. By fostering collaborative communication strategies and more accurate action selection, PPO-CCBNA significantly reduces communication interference and channel conflicts, resulting in enhanced overall communication efficiency.

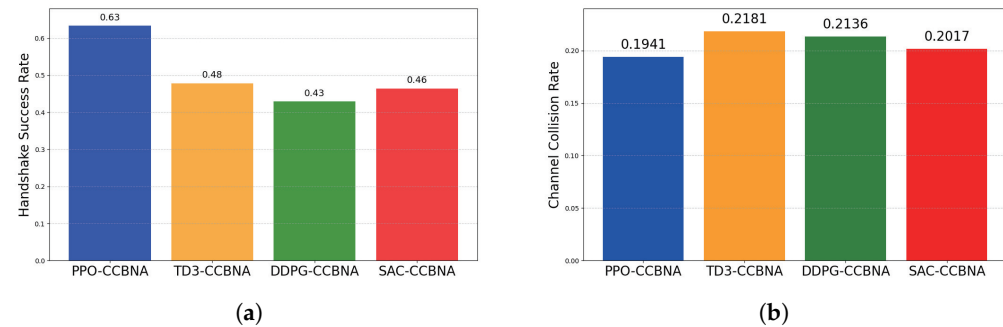


Figure 8. Comparison of Handshake Success Rate and Conflict Rate. (a) The handshake success rate under different strategies. (b) Conflict rate dynamics reflecting scheduling congestion and coordination.

Figure 9 illustrates the impact of the number of AGVs on the average reward. As the number of AGVs increases, the average reward gradually decreases, indicating that higher AGV density leads to increased communication interference and scheduling complexity. Among all algorithms, the PPO-CCBNA model achieves the highest average reward and shows better stability compared with TD3-CCBNA, SAC-CCBNA, and DDPG-CCBNA.

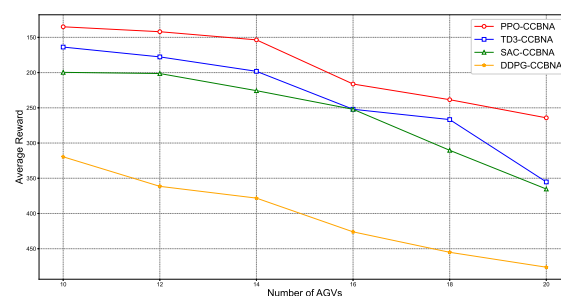


Figure 9. Average rewards for different numbers of AGVs.

Figure 10a–d illustrate the differences in performance between nonlinear AoI-based optimization and linear AoI-based optimization. The results show that the PPO-CCBNA algorithm, which adopts a nonlinear AoI optimization approach, delivers superior outcomes compared to the linear AoI-based PPO-CCBA algorithm across several core indicators, including handshake success rate, channel collision rate, message freshness, and nonlinear AoI penalty. This performance boost is mainly attributed to the nonlinear AoI mechanism's ability to impose stronger penalties on high AoI values, which in turn encourages agents to make more proactive and effective communication decisions. As a result, the PPO-CCBNA approach not only achieves higher handshake success rates and reduces channel collisions, but also significantly enhances the freshness of transmitted information throughout the system. Overall, these improvements mean that the proposed method offers clear advantages in terms of information transmission timeliness, stability, and the joint optimization of control and communication processes. In summary, the approach presented in this paper stands out in several key areas—including AoI management, handshake efficiency, channel resource allocation, and control stability—enabling rapid establishment of com-

munication links, reducing overall communication delay, and enhancing the system's real-time control capabilities.

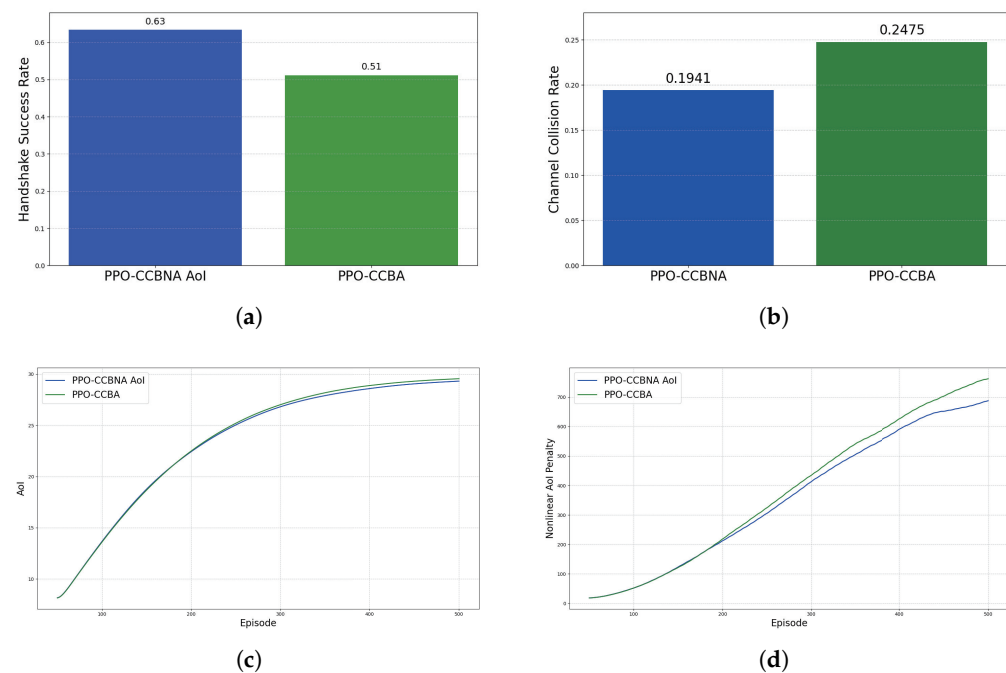


Figure 10. Comparison across four performance indicators. (a) Handshake success rate under various strategies. (b) Conflict rate variation reflecting message collisions. (c) AoI dynamics across time. (d) Nonlinear AoI penalty considering delay impact. All metrics provide insight into control efficiency and information freshness.

Figure 11 demonstrates the performance comparison of three PPO-based algorithms under different strategies: the PPO-CCBNA (time-division multiplexing + adaptive frequency hopping), the TDM-PPO-CCBNA (time-division multiplexing only) and the FF-PPO-CCBNA (fixed frequency). The results show that PPO-CCBNA achieves the highest handshake success rate and the lowest channel collision rate, proving the effectiveness of combining time-division multiplexing with adaptive frequency hopping. In contrast, TDM-PPO-CCBNA and FF-PPO-CCBNA present higher collision rates with lower success rates, with FF-PPO-CCBNA performing the worst due to the lack of both time-division multiplexing and frequency hopping mechanisms. This highlights the significant advantages of adaptive frequency hopping technology in improving communication efficiency and reducing interference.

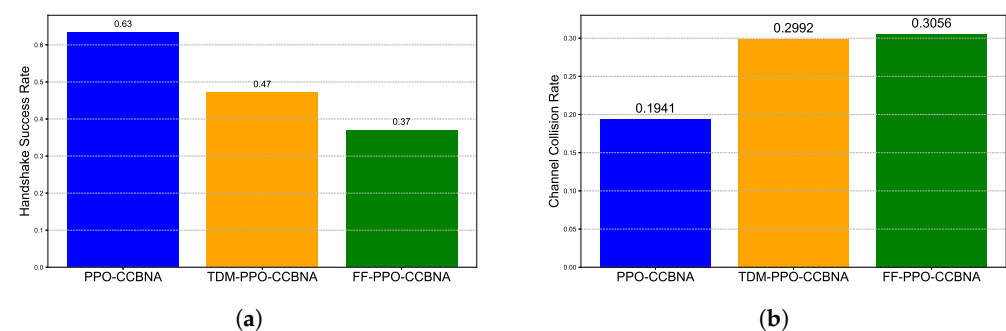


Figure 11. Performance evaluation of algorithms based on different complementary strategies. (a) Handshake Success Rate. (b) Conflict Rate.

Figure 12 compares the performance of the proposed PPO-CCBNA algorithm with two heuristic approaches: Greedy and RR+AoI-Trigger (Round Robin with AoI Trigger).

The results demonstrate the effectiveness of the proposed learning-based approach in optimizing handshake success and reducing channel collisions compared to simpler, non-learning baselines.

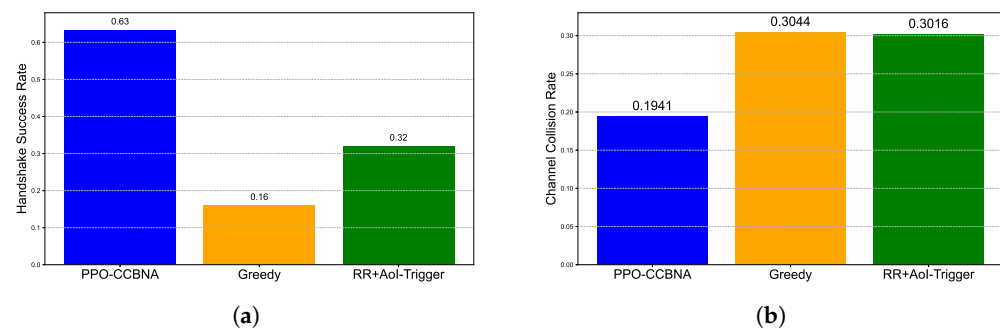


Figure 12. Performance evaluation with different algorithms. (a) Handshake Success Rate. (b) Conflict Rate.

7. Conclusions

To tackle the challenges of communication delays, information lags, and efficient collaboration in industrial intelligent systems, this paper proposes a cooperative optimization algorithm for perception control and communication based on nonlinear Aol optimization (PPO-CCBNA). This algorithm effectively mitigates the issue of information lag by introducing a nonlinear Aol penalty mechanism, which reduces the negative impact of communication delays on both control accuracy and system stability. Additionally, by integrating a differential dynamics model, state estimation and communication lag models, and a control error model, the algorithm incorporates an adaptive handshake period and a multi-channel frequency hopping communication strategy. These innovations optimize communication efficiency, enhance the system's resistance to interference, and improve overall communication performance. Experimental results show that the PPO-CCBNA algorithm significantly outperforms benchmark algorithms in terms of control stability, communication efficiency, and energy efficiency. Future research will focus on further improving the algorithm's robustness and adaptability to better handle complex real-world scenarios, ultimately enhancing the cooperative efficiency and control precision of multi-agent systems.

Author Contributions: Conceptualization, J.Y. and C.X. (Changqing Xia); methodology, J.Y.; software, J.Y.; validation, J.Y.; formal analysis, Y.X.; investigation, J.Y., C.X. (Changqing Xia), Y.L., C.X. (Chi Xu) and X.J.; resources, C.X. (Changqing Xia), Y.X., C.X. (Chi Xu) and X.J.; data curation, J.Y. and Y.L.; writing—original draft, J.Y. and C.X. (Changqing Xia); writing—review editing, C.X. (Changqing Xia), Y.X., Y.L. and X.J.; visualization, Y.X. and Y.L.; supervision, C.X. (Changqing Xia), Y.X., Y.L., C.X. (Chi Xu) and X.J.; project administration, C.X. (Changqing Xia), Y.X. and C.X. (Chi Xu); funding acquisition, C.X. (Changqing Xia), C.X. (Chi Xu) and X.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China under Grant 92367301; the Independent Subject of the State Key Laboratory of Robotics under Grant 2024-Z12; the Technology Program of Liaoning Province under Grant 2024-MSBA-84, 2024-MSBA-85, 2024JHI/11700050; and the Liaoning Revitalization Talents Program under Grant XLYC2202048.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Wang, Y.; Wu, S.; Lei, C.; Jiao, J.; Zhang, Q. A Review on Wireless Networked Control System: The Communication Perspective. *IEEE Internet Things J.* **2024**, *11*, 7499–7524. [\[CrossRef\]](#)
2. Jia, L.; Qi, N.; Su, Z.; Chu, F.; Fang, S.; Wong, K.-K.; Chae, C.-B. Game Theory and Reinforcement Learning for Anti-Jamming Defense in Wireless Communications: Current Research, Challenges, and Solutions. *IEEE Commun. Surv. Tutor.* **2025**, *27*, 1798–1838. [\[CrossRef\]](#)
3. Feng, D.; Jiang, C.; Lim, G.; Cimini, L.J.; Feng, G.; Li, G.Y. A Survey of Energy-Efficient Wireless Communications. *IEEE Commun. Surv. Tutor.* **2013**, *15*, 167–178. [\[CrossRef\]](#)
4. Xing, C.; Zhao, X.; Wang, S.; Xu, W.; Ng, S.X.; Chen, S. Hybrid Transceiver Optimization for Multi-Hop Communications. *IEEE J. Sel. Areas Commun.* **2020**, *38*, 1880–1895. [\[CrossRef\]](#)
5. Abedi, M. R.; Mokari, N.; Javan, M. R.; Saeedi, H.; Jorswieck, E. A.; Yanikomeroglu, H. Safety-Aware Age of Information (S-AoI) for Collision Risk Minimization in Cell-Free mMIMO Platooning Networks. *IEEE Trans. Netw. Serv. Manag.* **2024**, *21*, 3035–3053. [\[CrossRef\]](#)
6. Hu, W.; Wang, H.; Wang, J. Joint Optimizing High-Speed Cruise Control and Multi-Hop Communication in Platoons: A Reinforcement Learning Approach. *IEEE Trans. Intell. Transp. Syst.* **2024**, *25*, 18396–18407. [\[CrossRef\]](#)
7. Yang, B.; Zheng, M.; Liang, W. A Time-Efficient Rendezvous Algorithm with a Full Rendezvous Degree for Heterogeneous Cognitive Radio Networks. In Proceedings of the IEEE INFOCOM 2016—The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, 10–14 April 2016; pp. 1–9.
8. Lou, J.; Yuan, X.; Sigdel, P.; Qin, X.; Kompella, S.; Tzeng, N.-F. Age of Information Optimization in Multi-Channel Based Multi-Hop Wireless Networks. *IEEE Trans. Mob. Comput.* **2023**, *22*, 5719–5732. [\[CrossRef\]](#)
9. Wang, H.; Zhang, B.; Wang, H.; Wu, B.; Guo, D. Hopping Time Estimation of Frequency-Hopping Signals Based on HMM-Enhanced Bayesian Compressive Sensing with Missing Observations. *IEEE Commun. Lett.* **2022**, *26*, 2180–2184. [\[CrossRef\]](#)
10. Han, X.; Pan, H.; Wang, Z.; Li, J. Successive Interference Cancellation-Enabled Timely Status Update in Linear Multi-Hop Wireless Networks. *IEEE Trans. Mob. Comput.* **2025**, *24*, 5298–5311. [\[CrossRef\]](#)
11. Wang, F.; Lau, V.K.N. Multi-Level Over-the-Air Aggregation of Mobile Edge Computing over D2D Wireless Networks. *IEEE Trans. Wirel. Commun.* **2022**, *21*, 8337–8353. [\[CrossRef\]](#)
12. Zhang, T.; Chen, S.; Chen, Z.; Tian, Z.; Jia, Y.; Wang, M.; Wu, D.O. AoI and PAoI in the IoT-Based Multisource Status Update System: Violation Probabilities and Optimal Arrival Rate Allocation. *IEEE Internet Things J.* **2023**, *10*, 20617–20632. [\[CrossRef\]](#)
13. Long, Y.; Zhao, S.; Gong, S.; Gu, B.; Niyato, D.; Shen, X. AoI-Aware Sensing Scheduling and Trajectory Optimization for Multi-UAV-Assisted Wireless Backscatter Networks. *IEEE Trans. Veh. Technol.* **2024**, *73*, 15440–15455. [\[CrossRef\]](#)
14. Zhang, X.; Chang, Z.; Hämmäläinen, T.; Min, G. AoI-Energy Tradeoff for Data Collection in UAV-Assisted Wireless Networks. *IEEE Trans. Commun.* **2024**, *72*, 1849–1861. [\[CrossRef\]](#)
15. Huang, Z.; Wu, W.; Fu, C.; Chau, V.; Liu, X.; Wang, J.; Luo, J. AoI-Guaranteed Bandit: Information Gathering over Unreliable Channels. *IEEE Trans. Mob. Comput.* **2024**, *23*, 9469–9486. [\[CrossRef\]](#)
16. Shi, T.; Xu, Q.; Wang, J.; Xu, C.; Wu, K.; Lu, K.; Qiao, C. Enhancing the Safety of Autonomous Driving Systems via AoI-Optimized Task Scheduling. *IEEE Trans. Veh. Technol.* **2025**, *74*, 3804–3819. [\[CrossRef\]](#)
17. Lu, X.; Xu, Q.; Wang, X.; Lin, M.; Chen, C.; Shi, Z.; Guan, X. Full-Loop AoI-Based Joint Design of Control and Deterministic Transmission for Industrial CPS. *IEEE Trans. Ind. Inform.* **2023**, *19*, 10727–10738. [\[CrossRef\]](#)
18. Liao, H.; Zhou, Z.; Jia, Z.; Shu, Y.; Tariq, M.; Rodriguez, J.; Frasca, V. Ultra-Low AoI Digital Twin-Assisted Resource Allocation for Multi-Mode Power IoT in Distribution Grid Energy Management. *IEEE J. Sel. Areas Commun.* **2023**, *41*, 3122–3132. [\[CrossRef\]](#)
19. Wang, S.; Zhu, S.; Chen, C.; Shen, F.; Zhang, W.; Guan, X. Communication and Control Co-Design for Heterogeneous Industrial IoT: A Logic-Based Stochastic Switched System Approach. *IEEE J. Sel. Areas Commun.* **2023**, *41*, in press. [\[CrossRef\]](#)
20. Qiao, Y.; Fu, Y.; Yuan, M. Communication–Control Co-Design in Wireless Networks: A Cloud Control AGV Example. *IEEE Internet Things J.* **2023**, *10*, 2346–2359. [\[CrossRef\]](#)
21. Ma, B.; Lin, M.; Zhao, B.; Ouyang, H.; Wang, J. Robust Beamforming for Communication and Control Co-Design in Multi-Loop Wireless Control System. *IEEE Commun. Lett.* **2025**, *29*, 423–427. [\[CrossRef\]](#)
22. Zhao, G.; Imran, M.A.; Pang, Z.; Chen, Z.; Li, L. Toward Real-Time Control in Future Wireless Networks: Communication-Control Co-Design. *IEEE Commun. Mag.* **2019**, *57*, 138–144. [\[CrossRef\]](#)
23. Girgis, A.M.; Seo, H.; Park, J.; Bennis, M. Semantic and Logical Communication-Control Codesign for Correlated Dynamical Systems. *IEEE Internet Things J.* **2024**, *11*, 12631–12648. [\[CrossRef\]](#)
24. Han, Z.; Li, X.; Zhou, Z.; Huang, K.; Gong, Y.; Zhang, Q. Wireless Communication and Control Co-Design for System Identification. *IEEE Trans. Wirel. Commun.* **2024**, *23*, 4114–4126. [\[CrossRef\]](#)
25. Xu, H.; Yu, D.; Wang, Z.; Cheong, K.H.; Chen, C.L.P. Nonsingular Predefined Time Adaptive Dynamic Surface Control for Quantized Nonlinear Systems. *IEEE Trans. Syst. Man Cybern. Syst.* **2024**, *54*, 5567–5579. [\[CrossRef\]](#)

26. Sui, S.; Bai, Z.; Tong, S.; Chen, C.L.P. Finite-Time Adaptive Fuzzy Event-Triggered Consensus Control for High-Order MIMO Nonlinear MASS. *IEEE Trans. Fuzzy Syst.* **2024**, *32*, 671–682. [[CrossRef](#)]
27. Huang, T.; Liu, Y.; Liu, X.; Wang, M. A New Improved Multi-Sequence Frequency-Hopping Communication Anti-Jamming System. *Electronics* **2025**, *14*, 523. [[CrossRef](#)]
28. Li, X.; Chen, G.; Wu, G.; Sun, Z.; Chen, G. Research on Multi-Agent D2D Communication Resource Allocation Algorithm Based on A2C. *Electronics* **2023**, *12*, 360. [[CrossRef](#)]
29. Wang, Y.; Zhang, B.; Qin, S.; Peng, J. A Channel Rendezvous Algorithm for Multi-Unmanned Aerial Vehicle Networks Based on Average Consensus. *Sensors* **2023**, *23*, 8076. [[CrossRef](#)]
30. Liang, S.; Zhao, H.; Zhang, J.; Wang, H.; Wei, J.; Wang, J. A Multichannel MAC Protocol without Coordination or Prior Information for Directional Flying Ad hoc Networks. *Drones* **2023**, *7*, 691. [[CrossRef](#)]
31. Polyanskiy, Y.; Poor, H.V.; Verdú, S. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Trans. Inf. Theory* **2010**, *56*, 2307–2359. [[CrossRef](#)]
32. Kosta, A.; Pappas, N.; Angelakis, V. *Age of Information: A New Concept, Metric, and Tool*; Now Publisher: Delft, The Netherlands, 2017; Volume 12, pp. 162–259.
33. Sun, Y.; Uysal-Biyikoglu, E.; Yates, R.D.; Koksall, C.E.; Shroff, N.B. Update or Wait: How to Keep Your Data Fresh. *IEEE Trans. Inf. Theory* **2017**, *63*, 7492–7508. [[CrossRef](#)]
34. Sun, Y.; Cyr, B. Sampling for Data Freshness Optimization: Non-Linear Age Functions. *J. Commun. Netw.* **2019**, *21*, 204–219. [[CrossRef](#)]
35. Zou, Y.; Liu, Y.; Mu, X.; Zhang, X.; Liu, Y.; Yuen, C. Machine Learning in RIS-Assisted NOMA IoT Networks. *IEEE Internet Things J.* **2023**, *10*, 19427–19440. [[CrossRef](#)]
36. Liu, Y.; Zeng, P.; Cui, J.; Xia, C. Co-Design of Control, Computation, and Network Scheduling Based on Reinforcement Learning. *IEEE Internet Things J.* **2024**, *11*, 5249–5258. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.