SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

April 2026, Vol. 69, Iss. 4, 142301:1–142301:22
https://doi.org/10.1007/s11432-025-4701-7

Integrated sensing, communication, and control for multi-agent networked formation control

Ying ZHOU[1], Zhiyong FENG[1*], Zhiqing WEI[1], Dingyou MA[1], Danlan HUANG[1], Zeyang MENG[1], Yinglong FAN[1], Jie XU[2] & Ping ZHANG[1]

[1]*Key Laboratory of Universal Wireless Communications, Ministry of Education, School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*School of Science and Engineering (SSE), the Shenzhen Future Network of Intelligence Institute (FNii-Shenzhen), and the Guangdong Provincial Key Laboratory of Future Networks of Intelligence, The Chinese University of Hong Kong, Shenzhen, Shenzhen 518172, China*

Received 27 June 2025/Revised 15 October 2025/Accepted 3 December 2025/Published online 28 January 2026

Abstract This paper studies the multi-agent networked formation control in dynamic industrial environment, in which multiple automated guided vehicle (AGV) agents are coordinated by a base station (BS) to collectively perform cooperative transportation tasks. In this system, the limited local sensing capability at each agent and their frequent interactions may cause large synchronization errors and high closed-loop latency, degrading the networked formation control performance. To address these challenges, we propose a new communication-sensing enhanced multi-agent formation control strategy based on the idea of integrated sensing, communication, and control (ISCC). First, we establish an ISCC system design to accurately capture the interdependencies among sensing, communication, and control in networked formation control. Then, we design a dynamic obstacle avoidance risk map using the conditional value at risk, which quantifies the collision risks under communication latency, thus helping to reserve sufficient time for smooth obstacle avoidance and reduce material extrusion risks during emergency braking. Next, we formulate the multi-agent formation control problem as a partially observable Markov decision process, which is solved via the multi-agent proximal policy optimization (MAPPO) by exploiting the global ISCC states. Furthermore, to decrease the extra overhead for ISCC states interaction, we design a dynamic communication cycle allocation mechanism via global states, which effectively balances the synchronization precision and communication overhead. In addition, we employ a heterogeneous framework to mitigate gradient conflicts and boost control efficiency for heterogeneous agents. Simulation results reveal that our strategy improves the synchronous performance via reducing the error by at least 59.9% compared to baselines, significantly reducing closed-loop latency and traveling time.

Keywords integrated sensing communication and control (ISCC), networked formation control, industrial wireless networks, closed-loop control, multi-agent system

Citation Zhou Y, Feng Z Y, Wei Z Q, et al. Integrated sensing, communication, and control for multi-agent networked formation control. Sci China Inf Sci, 2026, 69(4): 142301, https://doi.org/10.1007/s11432-025-4701-7

1 Introduction

With advancements in artificial intelligence and communication technologies, multi-agent systems such as collaborative automated guided vehicles (AGVs) have gained widespread adoption in industrial automation and intelligent logistics [1–5]. Conventionally, multi-agent systems rely heavily on prearranged indoor simultaneous localization and mapping (SLAM) maps [4,6], radio-frequency identification (RFID) tags [7], or other established infrastructures to facilitate their localization and navigation for supporting various commercial applications. However, such systems typically adhere to fixed static trajectories, lacking robust sensing capabilities to adapt to dynamic, unforeseen environments. The limitation is particularly evident in highly dynamic automotive flexible assembly workshops. Furthermore, existing control systems may have limited communication capabilities as well. When they face stringent geometric formation constraints, they may struggle to achieve high-precision formation synchronization and fail to ensure smooth, safe transportation of multi-agent fleets [8,9]. Consequently, the development of an efficient, scalable, and adaptive networked formation control framework has emerged as one of the most pressing challenges in multi-agent systems for flexible manufacturing.

To achieve collision-free movement and synchronized control, there have been various prior studies in the literature developing precise sensing and positioning methods, alongside low-latency, scalable communication networks, to

* Corresponding author (email: fengzy@bupt.edu.cn)

enhance obstacle avoidance and cooperative control performance. Generally speaking, existing formation control methods primarily comprise two categories: sensing-based and communication-based approaches.

(i) Sensing-based formation control methods: Existing sensing-based formation control methods can be categorized into displacement-based, distance-based, position-based, and bearing-based approaches, depending on sensing performance and control law constraints [10–12]. Among them, bearing-based methods benefit from advancements in computer vision, enabling millimeter or even sub-millimeter level accuracy with low cost visual sensors [13]. These methods facilitate real-time trajectory adjustments by observing neighboring agents, supporting static obstacle avoidance and simple path planning [14], such as agent formation in circular motion [8] or linear motion between points [15]. However, due to the limited field-of-view, visual sensors suffer from localization offsets, which sacrifice speed to mitigate collision risks under sensing depth and visibility constraints in practical formation control applications [16, 17]. Furthermore, the lack of an efficient communication-sharing strategy in these methods results in under-utilized local sensing data, hindering multi-agent coordination and limiting control performance. Consequently, such low-speed, simplistic formation control strategies are ill-suited for high-dynamic environments, such as narrow corridors and dense AGV material zones, where the limited sensing range and delay exacerbate collision risks and congestion amid intersecting dynamic and static obstacles.

(ii) Communication-based formation control methods: These methods employ inter-agent information sharing to enable collaborative decision-making, primarily categorized into behavior-based [18], virtual structure [19], and leader-follower [15] approaches. First, inspired by the stable collective behaviors observed in bird flocks and fish schools [20], behavior-based methods design multi-agent kinematic models describing distributed formation control using local interaction information processing. These methods activate distinct decision-making to environmental conditions to ensure collision-free movement. Next, the virtual structure approach treats all agents as a unified rigid body controlled via centralized or distributed communication [21]. Furthermore, compared to the above methods, leader-follower control [15, 22] has emerged as the popular strategy, combining centralized and distributed communication benefits. The leader maintains low-latency communication with the base station (BS) for path navigation, and the followers coordinate formation maintenance through peer-to-peer communication, substantially reducing collaborative control complexity. In practice, errors from local sensors accumulate over time, requiring periodic BS interactions [23, 24] to correct pose and trajectory. This process enables the correction of position and trajectory while mitigating sensor and control-decision errors [25]. However, distributed formation control struggles to access global environmental information, resulting in poor adaptability to dynamic and complex environments, increased synchronization errors, and failing to maintain strict geometric formations. Moreover, centralized formation control is constrained by the BS communication resources and computational capacity, hindering its ability to support large-scale multi-agent collaboration. As the number of nodes grows and task complexity increases, frequent information exchange and calibration substantially elevate communication overhead, increase closed-loop latency and synchronization errors, and ultimately fail to meet the demands of complex formation tasks requiring ultra-low latency and high-efficiency cooperation.

Furthermore, the above kinematic model-based methods rely on predefined rules for decision-making, limiting their adaptability in complex and unknown environments [26–28]. With advances in machine learning [29,30], recent research has shifted away from kinematic model-based methods to learning-based formation control [31]. Deep reinforcement learning (DRL) techniques, such as multi-agent deep deterministic policy gradient (MADDPG) [32] and multi-agent proximal policy optimization (MAPPO) [33] have been actively applied to formation control in dynamic environment. Unlike traditional methods, these learning-based approaches enable multi-agent systems to autonomously learn optimal strategies from unknown environments without relying on pre-programmed rules, demonstrating superior performance in dynamic, uncertain, and highly nonlinear scenarios.

Despite these advancements, the above existing methods still have not overcome the limitations of independent redundant designs in sensing, communication, and control. First, due to limited communication resources and under-utilization of sensor data, these multi-agent systems frequently compute the optimal control strategy using low-dimensional local sensing information, such as the relative position and velocity [15, 34], leading to suboptimal cooperative control performance. Next, the independent design paradigm hinders the optimal multi-dimensional resource allocation for sensing, control, and communication, making it difficult to properly balance between the global sensing error accumulation versus the communication overhead. Consequently, systems often compromise trajectory smoothness to enhance maneuverability, resulting in incoherent emergency accelerations or decelerations during navigation that amplify synchronization errors [34, 35]. These issues will lead to material squeezing, causing material deformation or dropping damage [36]. Moreover, existing research lacks systematic investigation into the coupling relationships among sensing, communication, and control modules in multi-agent high-precision formation control. As the optimization constraints and the number of controllable agents increase, it is difficult for the system to effectively coordinate their multi-module interactions and dynamically adapt to the needs of complex

tasks. Meanwhile, the complexity of formation control may grow sharply, rendering the discovery of viable solutions exceptionally challenging. In this case, it is a difficult task to satisfy basic requirements like ensuring safe obstacle avoidance and formation synchronization in static obstacle environments, and it is becoming even more challenging to achieve high mobility and strict geometric formation synchronization in dynamic and complex environments. As such, prior studies on formation control in a structured known static obstacle environment were not applicable for unforeseen complex and dynamic environments for flexible manufacturing workshop.

To address the above problems, we propose a novel ISCC multi-agent formation control strategy that enables agents to autonomously learn optimal control and communication policies in complex, dynamic environments. Distinct from existing learning-based formation control approaches that treat communication and control optimization as independent or weakly coupled processes, our model establishes a unified closed-loop optimization architecture that explicitly models the bidirectional coupling among sensing, communication, and control. This unified design enhances decision relevance and adaptivity under partial observability, allowing agents to jointly optimize communication cycles and control actions based on global ISCC state information. Simulation results demonstrate that the proposed strategy achieves strict geometric synchronization and smooth trajectories while balancing communication overhead, sensing precision, and formation stability, thereby providing a scalable and efficient paradigm for intelligent industrial collaboration control. The main contributions of this paper are summarized as follows.

• **ISCC system-based multi-agent formation control process:** To accurately characterize the interactions among sensing, communication, and control in a highly dynamic and complex environment, we propose a novel ISCC multi-agent formation control system model for cooperative transportation in the automotive assembly workshop. By analyzing the relationships among communication closed-loop interaction cycles, synchronization errors, and control traveling time in the closed-loop formation control process, the proposed system effectively reduces redundant resource consumption from conventional decoupled designs and achieves balanced trajectory smoothness and maneuverability in multi-agent formation control.

• **Smooth formation synchronization with conditional value at risk (CVaR) map:** To evaluate the impact of dynamic environmental uncertainties on formation control, we develop a dynamic obstacle risk map for communication-sensing enhanced control using the CVaR model. By integrating multi-source sensing data, control pose information, and communication cycle time, we incorporate stricter tail expectations to constrain extreme risks such as congestion and collisions. During closed-loop formation control, the potential severity of congestion-induced collisions between AGVs and dynamic obstacles is quantitatively evaluated. This allows the formation to adjust its velocity smoothly, reducing the risk of material extrusion caused by emergency braking, and accordingly maintaining precise and synchronized motion without sacrificing speed.

• **Efficient MAPPO formation control via adaptive interaction cycle adjustment:** We formulate a multi-agent formation control problem, which aims to achieve efficient and safe movement of agent formations in complex, dynamic environments by optimizing coordination under partial observability. To handle this challenging problem, we propose a multi-agent formation control algorithm, named ISCC-based MAPPO (ISCC-MAPPO). In ISCC-MAPPO, leaders and followers as heterogeneous agents can be trained to handle optimal control and communication decisions in continuous and discrete action spaces simultaneously. During training, to reduce the extra communication overhead brought by ISCC global information, we propose a dynamic communication cycle allocation mechanism based on networked control states, guaranteeing high-precision formation synchronization while reducing communication overhead and latency. Additionally, to promote efficient cooperation among heterogeneous agents, we propose a heterogeneous actor-critic framework for leader and follower agents, improving the performance and applicability of the ISCC-MAPPO algorithm. Simulations show that our algorithm has better performance than baselines in terms of closed-loop latency and controllable agent scale.

The remainder of this paper is organized as follows. Section 2 introduces the system model and problem formulation. In Section 3, we introduce the proposed optimization method. In Section 4, the simulation results are given and discussed. Finally, in Section 5, we discuss the significance of research results and summarize the study.

2 System model

In this work, we consider a multi-agent formation control system based on ISCC, by particularly focusing on the cooperative transportation scenario. Firstly, we present an ISCC system model for multi-agent closed-loop formation control as shown in Figure 1, which consists of N AGVs, sensors, a BS, and an edge server. Each AGV formation consists of a leader agent and three follower agents, which collectively perform the cooperative transportation task of automobile bodies. The set of AGVs is defined as $\mathcal{N} = \{AGV_1, AGV_2, ..., AGV_n, ..., AGV_N\}$. The formation adopts a hierarchical communication structure, comprising a leader layer and a follower layer. The leader layer, consisting of

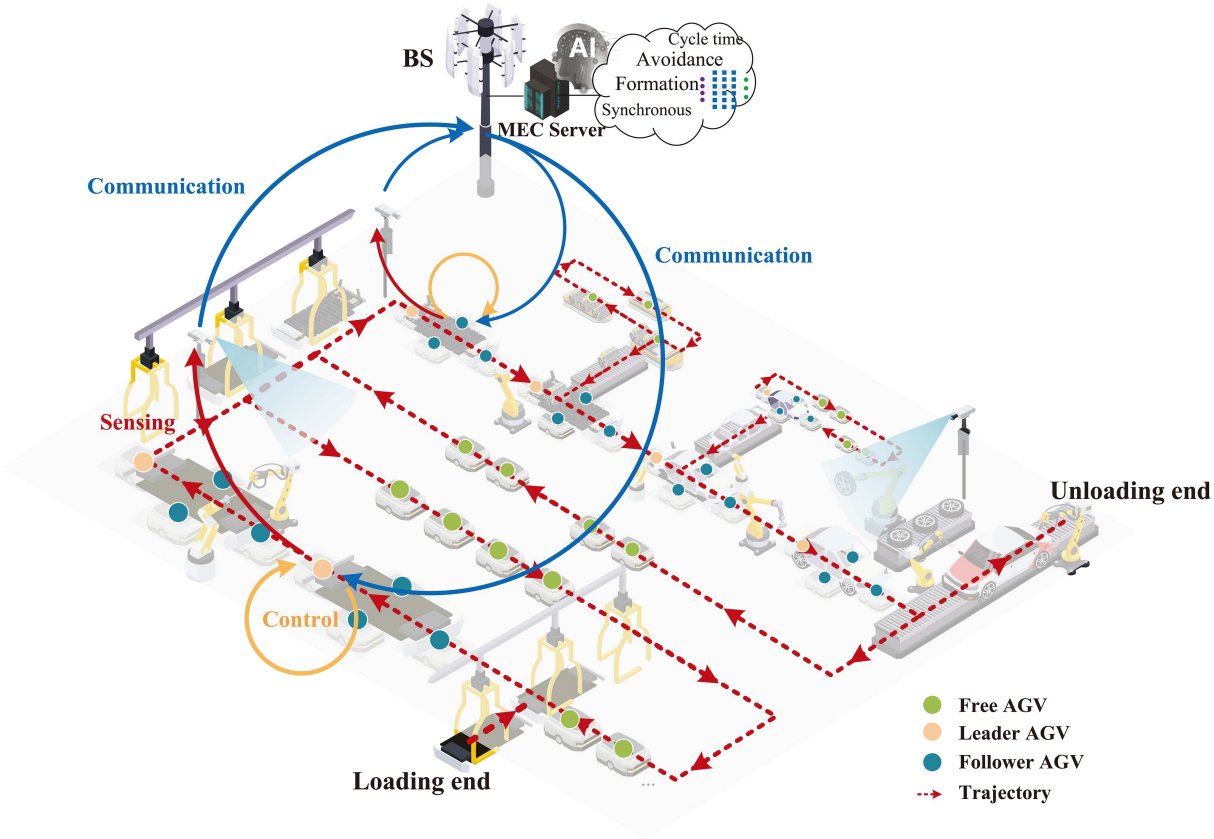

Figure 1 (Color online) ISCC system model for multi-agent closed-loop formation control in the automotive manufacturing assembly workshops.

leader entities, uploads their locally sensed motion states and control requests to the BS, receives trajectory control instructions and global state information fused from workshop sensors, and performs path navigation and obstacle avoidance. Then, the follower layer, comprising follower entities, uploads their locally sensed motion states and control requests, receives global state information and leader pose information, and performs trajectory correction and formation synchronization. To facilitate global state information fusion, depth camera sensors are deployed in the workshop, transmitting environment SLAM data to the BS via optical fiber. The BS utilizes 5G ultra-reliable low-latency communication (URLLC) protocol to communicate with AGVs and other intelligent machines, receiving sensing and task data from AGVs through the uplink and transmitting control instructions and global state information through the downlink. An edge server, deployed on the BS side, features a built-in programmable logic controller (PLC) and accesses the 5G user plane function (UPF), enabling routing and forwarding of 5G core network user plane data packets. The PLC exchanges key formation control information with AGVs, including real-time sensing motion states, control instructions, communication cycle, and other ISCC global information. Additionally, the edge server trains deep reinforcement learning models, calculates and updates the reference path of AGVs in real time, and makes control decisions.

To accurately characterize the interactions among sensing, communication, and control in a highly dynamic and complex environment, the closed-loop formation control process for multi-agent operates as follows. During formation movement, at slot t, the leader agent AGV_n receives control decisions and ISCC global state information $o_t^{(n)}$ from the BS via the downlink, which consists of the AGVs pose $\mathbf{p}^{(n)}$, velocity $v^{(n)}$, formation parameters $\mathbf{H}_n^f$, the upper bound of the closed-loop interaction cycle $\hat{T}_{Ct}^{(n)}$, and global risk map data $CVaR_\alpha$. Subsequently, it performs trajectory calibration and local obstacle avoidance, executing cooperative transportation tasks along the planned path. In the subsequent slot $T_{Ct}^{(n)}$, the leader agent utilizes local observation information from its onboard sensor and the actor network calculates control decisions, optimizes its trajectory locally, ensuring safe and stable advancement. At slot $t+T_{Ct}^{(n)}$, the leader agent transmits its current local observation data to the BS via the uplink. The edge server, located on the BS side, fuses and processes multi-source sensor data and communication network information to generate updated ISCC global information, which is then transmitted to the leader agent via the downlink. The leader agent then initiates a new round of trajectory calibration and collaborative transportation

tasks, thus forming a closed-loop control process. The follower agent in the formation is analogous to that of the leader agent, involving the receipt of ISCC global information via the downlink and the performance of trajectory calibration. Additionally, each follower agent also considers local trajectory optimization to maintain a precise formation shape. In the following, we introduce the sensing, communication, and formation control model.

2.1 Sensing model

Firstly, in multi-agent formation control, the BS receives sensor data from multiple sensors and local observation data from each agent at regular intervals of $T_{Ct}^{(n)}$. Multi-source sensing data are then provided to the edge server and AGVs as part of the global state information. The global high-precision sensing information enables the reduction of sensing errors accumulated by AGVs based on local observation data. It is assumed that during each closed-loop interaction, M visual SLAM sensors observe the current motion state $\mathbf{X}_n$ of AGV$_n$, with the original dataset perceived by the sensors denoted as $\{S_1, S_2, ..., S_M\}$. After processing by the edge server, the state observation values are represented as $\left\{\mathbf{X}_n^1, \mathbf{X}_n^2, ..., \mathbf{X}_n^M\right\}$, where $\mathbf{X}_n^i$ can be modeled by $\mathbf{X}_n^i = \mathbf{X}_n + \mathbf{N}_n^i$ $(1 \leqslant i \leqslant M)$, representing the observe data of the n-th agent sensed by the i-th sensor. $\mathbf{N}_n$ is the Gaussian white noise of observation value, $\mathbf{N}_n \sim \mathcal{N}\left(0, \sigma^2\right)$. The state observation value sensed by the AGV itself is represented as $\mathbf{X}_n^0$. The estimated value of $\mathbf{X}_n$ is expressed as

$$\hat{\mathbf{X}}_n = \frac{1}{M+1} \sum_{i=0}^{M} \mathbf{X}_n^i, \tag{1}$$

where the mean square error of $\hat{\mathbf{X}}_n$ is given by

$$\Delta_n = \mathbb{E}\left\{\left(\hat{\mathbf{X}}_n - \mathbf{X}_n\right)^2\right\} = \frac{\sigma^2}{M+1}. \tag{2}$$

It is evident from (2) that increasing the number of sensor samples M can significantly enhance the estimation accuracy of the target value $\mathbf{X}_n$.

Then, during the periodic interval when the agent exchanges sensor data with the BS, the agent typically estimates its own motion state using its onboard pose sensor and accordingly makes decisions. Take the two-wheeled differential AGV model illustrated in Figure 2 as an example. The pose of the AGV in the two-dimensional global coordinate system is represented by $\mathbf{p} = (x, y, \phi)$, where ϕ denotes the direction of movement. Furthermore, the velocity vector of the agent is defined as $\mathbf{u} = [v, \omega]^{\mathrm{T}}$, comprising linear velocity v and angular velocity ω. The kinematic model of the AGV is expressed as

$$\dot{\mathbf{p}}^{(n)} = \begin{bmatrix} \dot{x}^{(n)} \\ \dot{y}^{(n)} \\ \dot{\phi}^{(n)} \end{bmatrix} = \mathbf{J}^{(n)} \mathbf{u}^{(n)} = \begin{bmatrix} \cos\phi^{(n)} & 0 \\ \sin\phi^{(n)} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v^{(n)} \\ \omega^{(n)} \end{bmatrix}, \tag{3}$$

where the matrix $\mathbf{J}^{(n)}$ denotes the transposed matrix. Additionally, the odometer pose sensor embedded in AGV$_n$ measures the cumulative number of pulses from the driving motor via the encoder, enabling the estimation of $v^{(n)}$ and $\omega^{(n)}$ for AGV$_n$. The relationship is expressed as

$$\begin{bmatrix} v^{(n)} \\ \omega^{(n)} \end{bmatrix} = \begin{bmatrix} \frac{r^{(n)}}{2} & \frac{r^{(n)}}{2} \\ \frac{r^{(n)}}{L_w^{(n)}} & -\frac{r^{(n)}}{L_w^{(n)}} \end{bmatrix} \begin{bmatrix} \omega_L^{(n)} \\ \omega_R^{(n)} \end{bmatrix}, \tag{4}$$

$$\begin{cases} \Delta x_L^{(n)} = \frac{2\pi r^{(n)} N_L^{(n)}}{Z^{(n)}}, \\ \Delta x_R^{(n)} = \frac{2\pi r^{(n)} N_R^{(n)}}{Z^{(n)}}, \end{cases} \tag{5}$$

where $r^{(n)}$ denotes the radius of the tire of AGV$_n$, and $L_w^{(n)}$ represents the wheelbase of the AGV$_n$ chassis. The angular velocities of the left and right tires of AGV$_n$ are defined as $\omega_L^{(n)}$ and $\omega_R^{(n)}$, respectively. Furthermore, $\Delta x_L^{(n)}$ and $\Delta x_R^{(n)}$ represent the distances traveled by the left and right wheels of AGV$_n$ within the sampling time step $T_s^{(n)}$. The number of pulse signals output by the left and right wheels of AGV$_n$ within $T_s^{(n)}$ are defined as $N_L^{(n)}$ and $N_R^{(n)}$, respectively, where $N_L^{(n)} = \hat{N}_L^{(n)} + N_n^{(n)}$ and $N_R^{(n)} = \hat{N}_R^{(n)} + N_n^{(n)}$. Here, $\hat{N}_L^{(n)}$ and $\hat{N}_R^{(n)}$ are the estimated values

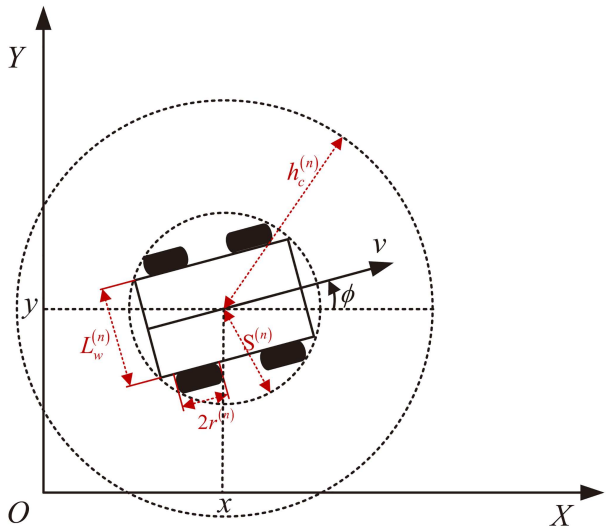

Figure 2 (Color online) Two-wheel differential AGV motion model.

of the pulse count, and $N_n^{(n)}$ denotes the Gaussian white noise, $N_n^{(n)} \sim \mathcal{N}(0, \sigma_s^2)$. Additionally, $Z^{(n)}$ represents the number of lines of the encoder in the pose sensor.

From (5), we derive the expressions for $\omega_L^{(n)}$ and $\omega_R^{(n)}$ as $\omega_L^{(n)} = \left(2\pi N_L^{(n)}\right) / \left(T_s^{(n)} Z^{(n)}\right)$ and $\omega_R^{(n)} = \left(2\pi N_R^{(n)}\right) / \left(T_s^{(n)} Z^{(n)}\right)$, respectively. Substituting them into (4), the estimated angular velocity is expressed as

$$\omega^{(n)} = \frac{2\pi r^{(n)} \left(N_L^{(n)} - N_R^{(n)}\right)}{L_w^{(n)} Z^{(n)} T_s^{(n)}}, \tag{6}$$

where $\Delta\phi^{(n)} = \omega^{(n)} T_s^{(n)}$. By substituting (4) and (6) into (3), the motion trajectory of AGV$_n$ based on its local sensor observation is estimated as

$$\mathbf{p}_s^{(n)}(t+1) = \mathbf{p}_s^{(n)}(t) + T_s^{(n)} \cdot \dot{\mathbf{p}}_s^{(n)}(t), \tag{7}$$

$$\begin{bmatrix} x^{(n)}(t+1) \\ y^{(n)}(t+1) \\ \phi^{(n)}(t+1) \end{bmatrix} = \begin{bmatrix} x^{(n)}(t) \\ y^{(n)}(t) \\ \phi^{(n)}(t) \end{bmatrix} + T_s^{(n)} \cdot \begin{bmatrix} \frac{\pi r^{(n)} \left(\cos\phi^{(n)}(t)\right) \left(N_L^{(n)} + N_R^{(n)}\right)}{T_s^{(n)} Z^{(n)}} \\ \frac{2\pi r^{(n)} \left(\sin\phi^{(n)}(t)\right) \left(N_L^{(n)} - N_R^{(n)}\right)}{T_s^{(n)} Z^{(n)} L_w^{(n)}} \\ \frac{2\pi r^{(n)} \left(N_L^{(n)} - N_R^{(n)}\right)}{T_s^{(n)} Z^{(n)} L_w^{(n)}} \end{bmatrix}. \tag{8}$$

2.2 Communication model

In our study, we assume that the current AGVs are ready to transmit their control request packets and that the user arrival rate follows a Poisson distribution. In the initial transmission process, each control decision packet transmission occupies B Hz of bandwidth in the frequency domain. The probability of a single intelligent machine successfully transmitting a data packet in L repeated transmissions is given by

$$P_n(n_{Ret} = L) = \sum_{l=1}^{L} \binom{L}{l} p_{ret}^l (1-p_{ret})^{L-l} = 1 - (1-p_{ret})^L, \tag{9}$$

where p_{ret} is the probability of successful transmission after uplink unauthorized access without collision, which is expressed as $p_{ret} = (1 - 1/N_0)^{(N^{ar}-1)}$. Here, N_0 denotes the number of available preamble sequence codes at the BS, and N^{ar} represents the number of business requests from all intelligent machines to the BS per unit time. Additionally, the average number of retransmissions for a single user is given by $\mathbb{E}(n_{Ret}) = \sum_{n_{Ret}=1}^{L} n_{Ret} \cdot P_n(n_{Ret})$.

In networked formation control, the closed-loop interaction process between an AGV and the BS involves three components: the uplink sensing data transmission delay $T_U^{(n)}(t)$, the BS internal data processing delay $T_{BS}^{(n)}(t)$, and

the downlink feedback control delay $T_D^{(n)}(t)$. The closed-loop communication latency is thus expressed as

$$\begin{aligned} T_C^{(n)}(t) &= T_U^{(n)}(t) + T_{BS}^{(n)}(t) + T_D^{(n)}(t) \\ &= \mathbb{E}(n_{Ret}) \cdot (T_{US}^{(n)}(t) + T_{UB}^{(n)}(t)) + (T_Q^{(n)}(t) + T_O^{(n)}(t)) + (T_{DS}^{(n)}(t) + T_{DB}^{(n)}(t)) \\ &= \mathbb{E}(n_{Ret}) \cdot \left(\frac{D_U^{(n)}(t)}{R^{(n)}(t)} + \frac{d^{(n)}}{c} \right) + \left(\frac{N^{ar}}{\mu(\mu - N^{ar})} + \frac{D_U^{(n)}(t)\delta_{MEC}}{f^r} \right) + \left(\frac{D_D^{(n)}(t)}{R^{(n)}(t)} + \frac{d^{(n)}}{c} \right), \end{aligned} \tag{10}$$

where $T_{US}^{(n)}(t)$ represents the transmission delay for AGV$_n$ to send data to the BS at slot t, which is equivalent to the transmission time of each bit stream data. Additionally, $T_{UB}^{(n)}(t)$ denotes the broadcast delay of the channel between AGV$_n$ and the BS at slot t, and $D_U^{(n)}(t)$ and $D_D^{(n)}(t)$ denote the uplink and downlink business data sent by AGV$_n$ to BS and by BS to AGV$_n$, respectively. Furthermore, $T_Q^{(n)}(t)$ is the queue time when the uplink business data of AGV$_n$ arrives at BS, while $T_O^{(n)}(t)$ is the processing time of AGV$_n$ business data by BS. The parameters μ, δ_{MEC}, and f^r represent the number of AGVs supported by the BS per unit time, the number of CPU cycles required for the edge server to process 1 bit of data, and the local computing resource of the edge server, respectively. Moreover, $R^{(n)}(t)$ is the data transmission rate between the BS and AGV$_n$, $d^{(n)}$ is the Euclidean distance between AGV$_n$ and the BS, and c is the speed of electromagnetic wave propagation.

According to the finite blocklength theory, the short packet transmission rate $R^{(n)}$ (in bits per second) in an intelligent machine network can be approximated as

$$R^{(n)} = B \left(\log_2(1 + SNR^{(n)}) - \sqrt{\frac{V^{(n)}}{{t_d}^{(n)} B}} Q^{-1}(\varepsilon) \right), \tag{11}$$

where $SNR^{(n)}$ denotes the signal-to-noise ratio of AGV$_n$, and ${t_d}^{(n)}$ represents the transmission duration of short packets. Additionally, B is the channel bandwidth, ε is the decoding error rate of short packets, $V^{(n)}$ is the channel dispersion, and $Q^{-1}(\cdot)$ is the inverse function of the Gaussian Q-function, where $Q(x) = \int_x^\infty (1/\sqrt{2\pi}) \exp(-t^2/2) dt$.

Building upon this, we further account for the control decision lag caused by closed-loop communication latency. Accordingly, the discrete-time motion state of an agent considering communication latency is expressed as

$$\mathbf{p}^{(n)}(t+1) = \mathbf{p}^{(n)}(t) + T_C^{(n)}(t) \cdot \mathbf{J}(t) \cdot \mathbf{u}^{(n)}(t), \tag{12}$$

$$\mathbf{J}(t) = \begin{bmatrix} \cos\theta_n(t) & -\sin\theta_n(t) & 0 \\ \sin\theta_n(t) & \cos\theta_n(t) & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{13}$$

where $\mathbf{p}^{(n)}$ denotes the pose information of AGV$_n$, $\mathbf{J}(t)$ is the rotation matrix that maps the AGVs' local velocity commands to the global coordinate system. The control decision is defined as $\mathbf{u}^{(n)}(t) = \begin{bmatrix} \dot{x}^{(n)}(t) & \dot{y}^{(n)}(t) & \dot{\phi}^{(n)}(t) \end{bmatrix}^{\mathrm{T}}$.

2.3 Formation control model

As illustrated in Figure 3, networked formation control typically employs the leader-follower control method to achieve collaboration among multiple agents. Initially, an arbitrary agent in each formation is designated as the leader, which guides the remaining followers to safely reach their destination. During the journey, the follower agents obtain the leaders' pose information from the BS or their own sensors and calculate the ideal relative distance and relative angle based on the initial formation information. It ensures that all followers form a strict geometric pattern with the leader, which is determined by parameters such as relative distance and relative angle. Geometric formations vary widely, encompassing distinct patterns such as columnar and diamond arrangements. In this study, we consider the diamond formation as an example, and the expected formation parameter matrix at slot t is defined as

$$\hat{\mathbf{H}}^f(t) = \left[\hat{\mathbf{H}}_1^f, \hat{\mathbf{H}}_2^f, ..., \hat{\mathbf{H}}_n^f \right]_{4\times n}. \tag{14}$$

In (14), the expected queue parameter of AGV$_n$ in the formation is defined as $\hat{\mathbf{H}}_n^f$, which is expressed as $\hat{\mathbf{H}}_n^f = \left[\hat{h}_1^{(n)} \ \hat{h}_2^{(n)} \ \hat{h}_3^{(n)} \ \hat{h}_4^{(n)} \right]$. Specifically, $\hat{h}_1^{(n)}$ represents the ID number of the agent AGV$_n$ in the formation, $\hat{h}_2^{(n)}$ represents

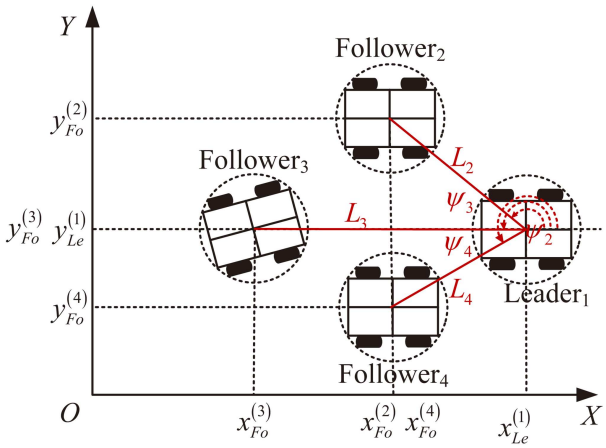

Figure 3 (Color online) Networked multi-agent formation control model.

the ID number of the leader agent in the formation where AGV_n is located, $\hat{h}_3^{(n)}$ represents the ideal geometric center distance $\hat{L}^{(n)}$ between AGV_n and the leader, and $\hat{h}_4^{(n)}$ represents the ideal angle $\hat{\psi}^{(n)}$ from the positive x-axis to the geometric center connecting AGV_n and the leader. The expected pose information $\hat{\mathbf{p}}^{(n)}(t)$ of the agent AGV_n at slot t is given by

$$\hat{\mathbf{p}}^{(n)}(t)=\begin{bmatrix}\hat{x}_{Fo}^{(n)}(t)\\ \hat{y}_{Fo}^{(n)}(t)\\ \hat{\psi}^{(n)}(t)\end{bmatrix}=\begin{bmatrix}x_{Le}(t)+\hat{L}^{(n)}\cos\hat{\psi}^{(n)}(t)\\ y_{Le}(t)+\hat{L}^{(n)}\sin\hat{\psi}^{(n)}(t)\\ \hat{\psi}^{(n)}(t)\end{bmatrix},\tag{15}$$

where $x_{Le}(t)$ and $y_{Le}(t)$ denote the pose information of the leader at slot t, respectively.

2.4 Problem formulation

Due to the limited communication resources and sensing range, in order to ensure sufficient braking reaction time for obstacle avoidance and formation maintenance, the AGV formation control system may need to compromise the movement performance, via reducing the driving speed and increasing the braking acceleration. Although such approaches mitigate the communication delay effects, they cannot meet the high maneuverability requirements of large-scale AGV systems in complex dynamic environments and may cause serious pose deviation issues due to sensor errors and autonomous navigation. Therefore, it is necessary to further analyze the formation control constraints due to error accumulation.

Considering the accumulation of sensing errors in local pose information over time, the cumulative position error during periods of autonomous driving without BS control information is calculated using (8) as

$$e_p^{(n)}(t)=\begin{bmatrix}x_e^{(n)}(t)\\ y_e^{(n)}(t)\\ \phi_e^{(n)}(t)\end{bmatrix}=\left\|\mathbf{p}_s^{(n)}(t)-\mathbf{p}^{(n)}(t)\right\|.\tag{16}$$

Additionally, based on the definition in (8), the formation parameter matrix $\mathbf{H}^f(t)$ of the actual formation is calculated as $\mathbf{H}^f(t)=\left[\mathbf{H}_1^f,\mathbf{H}_2^f,...,\mathbf{H}_n^f\right]_{4\times n}$, where $\mathbf{H}_n^f$ represents the queue parameter of AGV_n in the formation, and can be represented as $\mathbf{H}_n^f=\left[h_1^{(n)}\ h_2^{(n)}\ h_3^{(n)}\ h_4^{(n)}\right]$.

The formation synchronization error between two relative formation patterns $\hat{\mathbf{H}}^f(t)$ and $\mathbf{H}^f(t)$ is defined as the square of the Euclidean norm of the actual formation and the expected formation, given by

$$e_s(t)=\left\|\hat{\mathbf{H}}^f(t)-\mathbf{H}^f(t)\right\|.\tag{17}$$

In summary, in order to achieve stable formation control, both the position error $e_p^{(n)}(t)$ and the synchronization error $e_s(t)$ must converge within a safe threshold. Furthermore, considering that the agent moves in a 2D plane

with obstacles, other intelligent bodies can be treated as dynamic obstacles with uncertainty. To ensure formation safety and avoidance driving, the following control constraint conditions must be satisfied:

$$\begin{cases} \left\| \mathbf{p}^{(n)}(t) - \mathbf{p}^{(n')}(t) \right\| \geqslant \frac{1}{2} h_s^{(n)}, \\ \left\| \mathbf{p}^{(n)}(t) - \mathbf{p}_o^{(n)}(t) \right\| \geqslant h_s^{(n)}, \end{cases} \tag{18}$$

where $\mathbf{p}^{(n')}(t)$ denotes the pose information of the agent $\mathrm{AGV}_{n'}$, which belongs to the neighbor set of AGV_n. $\mathbf{p}_o^{(n)}(t)$ represents the position information of obstacles surrounding AGV_n. The safe braking distance of AGV_n is given by $h_s^{(n)} = \left(v_{\max}^2 + 2a_{\max} \cdot S^{(n)}\right) / \left(\omega_0 \cdot 2a_{\max}\right)$, where $S^{(n)}$ is the inner radius of the ground projection contour of AGV_n, ω_0 is the warning distance factor based on the collision probability threshold, and $v_{\max}$ and $a_{\max}$ are the maximum driving speed and acceleration of AGV_n, respectively.

In the complex environment of industrial multi-agent formation control, it is imperative not only to meet the control requirements mentioned above but also to ensure rapid and safe driving while maintaining the integrity and consistency of the formation. Towards this end, our objective is to minimize the formation's travel time, subject to constraints on sensing error, communication quality, and control synchronization error, by optimizing the cooperative control strategy to safely and efficiently accomplish the transportation task. The optimization problem is defined as

$$\begin{aligned} \min_{A_t} \quad & \frac{1}{K} \sum_{k=1}^{K} \left(T_{cur}^{(k)} + s^{(k)}(\pi_\theta) \cdot \Delta T \right) \\ \text{s.t. } & C_1: \ \mathbb{E}(n_{Ret}) \leqslant n_{Ret}^{th}, \\ & C_2: \ \left\| \mathbf{p}_s^{(n)}(t) - \mathbf{p}^{(n)}(t) \right\| \leqslant \varepsilon_{ep}^{th}, \forall n \in \mathcal{N}, \\ & C_3: \ \Delta^{(n)} \leqslant \Delta_{th}, \forall n \in \mathcal{N}, \\ & C_4: \ \left\| \mathbf{p}^{(n)}(t) - \mathbf{p}^{(n')}(t) \right\| \geqslant \frac{1}{2} h_s^{(n)}, \\ & \qquad \left\| \mathbf{p}^{(n)}(t) - \mathbf{p}_o^{(n)}(t) \right\| \geqslant h_s^{(n)}, \forall n \in \mathcal{N}, \\ & C_5: \ \left\| \hat{\mathbf{H}}^f(t) - \mathbf{H}^f(t) \right\| \leqslant \varepsilon_{es}^{th}, \forall n \in \mathcal{N}, \end{aligned} \tag{19}$$

where $T_{cur}^{(k)}$ denotes the formations travel time, π_θ represents the control strategy, and $s^{(k)}(\cdot)$ is the number of time steps required for the k-th formation to reach its destination from current position under π_θ. The joint action space A_t comprises the agents' control decisions, including linear and angular velocities, along with communication cycle decisions. $s^{(k)}(\pi_\theta) = d_{rem}^{(k)} / \left(v^{(k)} \Delta T\right)$, where $v^{(k)}$ is the current speed of the leader agent in formation, and ΔT is the controllers sampling step length. The constraints C_1–C_5 are imposed to ensure the safe and efficient operation of the formation. Specifically, C_1 limits the average number of retransmissions per agent to prevent indefinite retransmission attempts, which could occupy excessive resources and cause congestion, thereby preventing other users from communicating normally. Constraint C_2 restricts the pose error of agents to ensure that they can quickly respond to the BS's control requests and correct their trajectories to avoid obstacles. Constraint C_3 limits the sensing error of the agent's motion state to achieve high-precision positioning and control decisions, ensuring the stability and safety for the multi-agent formation control system. Constraint C_4 ensures a safe distance between agents and dynamic obstacles, allowing agents to emergency brake within a safe distance to avoid collisions. Finally, constraint C_5 restricts the synchronization error of agents to ensure that they move in strict accordance with the geometric formation.

3 Multi-agent formation control using MAPPO with ISCC-based global information sharing

In this section, we model the formation control problem as a POMDP for communication-sensing enhanced control, addressing formation control under various sensing, communication, and control constraints. Specifically, we propose the dynamic obstacle avoidance risk map using the CVaR model, and the dynamic communication cycle allocation mechanism based on networked control states. Additionally, we present a multi-agent formation control algorithm, named ISCC-based MAPPO (ISCC-MAPPO).

3.1 Dynamic obstacle avoidance risk map based on CVaR

In complex and dynamic environments with dense obstacles, AGVs are prone to low-probability but high-loss congestion collisions with dynamic and static obstacles, resulting in material damage and drops. Traditional methods

often sacrifice movement performance by using low-speed driving to replace a more sufficient reaction distance. However, they still make decisions based on the safe distance from obstacles within the local sensing range, inevitably leading to emergency braking decisions, which may result in severe speed fluctuations, increase synchronization errors, and cause material damage. To address this issue, it is essential to consider both a comprehensive sensing range to warn and avoid collisions and the smoothness of variable speed to ensure strict synchronization and material safety of the formation. Therefore, we propose a quantitative evaluation of the impact of dynamic environmental uncertainty on multi-agent obstacle avoidance and formation synchronization, using a network delay impact model and a dynamic obstacle avoidance risk map based on the CVaR model.

Specifically, the CVaR model [37] provides a new perspective for dynamic obstacle avoidance by quantifying extreme risks through tail risk measurement and forward-looking decision-making. By evaluating the potential collision loss intensity of obstacle interaction in the worst-case scenario, CVaR predicts and quantifies the motion conflict loss degree of the AGV formation in the high-confidence interval of the risk map. It upgrades traditional obstacle avoidance based on relative distance probability avoidance to risk value control under the collision probability distribution. The dynamic obstacle avoidance risk map based on CVaR not only depicts environmental uncertainty in probability risk assessment but also strengthens local path planning for low-probability high-risk events, through the α-quantile tail risk focusing mechanism. This enables the output of smooth variable speed strategies within the safe threshold of the action space while maintaining the precision of formation synchronization.

In actual industrial environments, multiple AGVs travel in sequence along ring-shaped production lines, and neighboring AGVs can be regarded as dynamic obstacles with uncertainty. Since AGVs and other dynamic obstacles start braking and slowing down only after receiving braking instructions or global state information, it is necessary to calculate the movement distance during the communication delay. Assuming the worst-case scenario, the dynamic obstacle and AGV decelerate simultaneously at the maximum speed and just collide when they stop braking. The relative speed and safety braking distance are derived as

$$v^{(nn')} = \sqrt{\left(v^{(n)}\right)^2 + \left(v^{(n')}\right)^2 - 2v^{(n)}v^{(n')}\cos\left(\phi^{(n)} - \phi^{(n')}\right)}, \tag{20}$$

$$d_{rb}^{(nn')} = v^{(nn')}\bar{T}_C + \left\| \frac{\left(v^{(n)}\right)^2}{2a_{\max}^{(n)}} \begin{bmatrix} \cos\phi^{(n)} \\ \sin\phi^{(n)} \end{bmatrix} - \frac{\left(v^{(n')}\right)^2}{2a_{\max}^{(n')}} \begin{bmatrix} \cos\phi^{(n')} \\ \sin\phi^{(n')} \end{bmatrix} \right\|, \tag{21}$$

where $\bar{T}_C$ denotes the average network closed-loop latency of the agents in the formation. Since static obstacles are typically fixed in position, it is assumed that the AGV has planned a collision-free path using its own sensors. When the current relative distance $d_{cur}^{(nn')}$ is no greater than $d_{rb}^{(nn')}$, i.e., $d_{cur}^{(nn')} \leqslant d_{rb}^{(nn')}$, the congestion and collision will occur. To model the randomness of relative velocity, a random variable $Z = d_{rb}^{(nn')} - d_{cur}^{(nn')}$ is defined. The congestion and collision probability is then equivalent to the probability that $Z \leqslant 0$, i.e.,

$$\begin{aligned} P_{coll}^{(n)} &= \mathbb{P}\left(d_{cur}^{(nn')} \leqslant d_{rb}^{(nn')}\right) \\ &= \mathbb{P}\left(Z \geqslant 0\right). \end{aligned} \tag{22}$$

To ensure that the tail risk of the collision probability does not exceed the safety threshold ϵ_{cv}, the target function CVaR based on the congestion and collision probability risk value is designed, $CVaR_\alpha\left(P_{coll}^{(n)}\right) \leqslant \epsilon_{cv}$, where $\alpha \in (0,1)$ is the confidence level. For consistency across experiments, the confidence level was empirically fixed at $\alpha = 0.9$, enabling fair performance comparison under the same risk threshold and achieving a balanced trade-off between safety (collision avoidance) and control efficiency. Using the CVaR optimization framework, the potential congestion collision risk of the AGV formation and dynamic obstacles under the influence of network delay is quantified as a risk value $CVaR_\alpha$.

$$\begin{aligned} CVaR_\alpha &= \mathbb{E}[P_{coll}^{(n)} | P_{coll}^{(n)} \geqslant VaR_\alpha] \\ &= \frac{1}{1-\alpha}\int_{P_{coll}^{(n)} \geqslant VaR_\alpha} P_{coll}^{(n)} f\left(P_{coll}^{(n)}\right) dP, \end{aligned} \tag{23}$$

where $f\left(P_{coll}^{(n)}\right)$ is the probability density function of the collision probability, and VaR_α represents the maximum probability of collision between the AGV and the obstacle at a certain confidence level.

Finally, a two-dimensional dynamic risk map of the environment is constructed using real-time sensor and communication network feature information, where each grid area is assigned a risk value. The grid size is adaptively adjusted based on the actual response time and speed smoothing requirements. By integrating sensing data with dynamic pose information, network latency, and other parameters, the AGV receives real-time global dynamic map feedback, enabling it to respond promptly to environmental changes. In addition, our future work will explore an adaptive confidence level α that dynamically varies with obstacle density and environmental uncertainty to enhance real-time risk perception and control flexibility in complex formation scenarios.

3.2 Dynamic communication cycle allocation mechanism based on networked control status

The AGV can sense environmental data through local sensors, which contains sensing and synchronization errors that accumulate temporally. To mitigate this error accumulation, the AGV must interact with the BS at a fixed interaction cycle to calibrate the information and eliminate errors. However, decreasing the cycle of AGV interaction with the BS global state information, while reducing the accumulation of sensing and synchronization errors, also introduces additional communication overhead, and increases the closed-loop communication delay. To balance the global information communication interaction cycle versus the accumulation of sensing and synchronization errors, a dynamic communication cycle allocation mechanism for networked control status is proposed. The action space dynamically outputs a suitable communication interaction cycle based on ISCC state observation values, such as motion state and collision risk value of the driving area. The communication interaction cycle of multi-agent formation can be analyzed in two primary cases.

Case 1: When the formation operates in an area with a low CVaR risk value, such as a non-assembly area or a low AGV density area, the control system is relatively stable and less prone to interference. Under traditional fixed interaction cycle mechanisms, the BS receives redundant perception data from stable state AGVs and transmits similar control instructions at a high frequency, resulting in a waste of network resources and excessive communication overhead. To address the above problem, the upper bound of the closed-loop interaction cycle is used as state space data to jointly train control and communication strategies in the critic network. In particular, the model learns to compute the optimal interaction cycle that ensures efficient and stable formation control. This enables adaptive extension of the interaction cycle under stability constraints, thereby reducing the frequency of each agent's access to the BS and minimizing communication overhead. As a result, the number of AGVs controllable by the BS is increased, ultimately enhancing overall production efficiency.

Specifically, the motion process of multi-agent driving in sequence along a ring-shaped production line is approximated using a leader-follower tracking trajectory. Based on the stability analysis of the multi-agent control system in our previous work [38], each agent can derive the upper bound of the closed-loop interaction cycle $\hat{T}_{Ct}^{(n)}(o_t^{(n)})$ under control system stability conditions by utilizing the received state observation date $o_t^{(n)}$. According to [35], the upper bound of the closed-loop interaction cycle is given by

$$\hat{T}_{Ct}^{(n)}\left(o_t^{(n)}\right) \leqslant \frac{1+(T_{cyc})^2[V'(h_s^{(n)})]^2-2V'(h_s^{(n)})^2T_{cyc}+2V'(h_s^{(n)})T_{cyc}}{2V'(h_s^{(n)})^2}, \tag{24}$$

$$T_{Ct}^{(n)}\left(o_t^{(n)}\right) = \hat{T}_{Ct}^{(n)}\left(o_t^{(n)}\right) - \mathrm{mod}\left(\hat{T}_{Ct}^{(n)}\left(o_t^{(n)}\right), T_{cyc}\right), \tag{25}$$

where $V'(\cdot)$ is the optimal velocity function and $V(\Delta x) = (v_{\max}/2)\cdot[\tanh(\Delta x - h_s) + \tanh(h_s)]$. The modulo function $\mathrm{mod}(\cdot)$ is used to calculate the remainder between $\hat{T}_{Ct}^{(n)}\left(o_t^{(n)}\right)$ and T_{cyc}, ensuring that the actual AGV communication cycle $T_{Ct}^{(n)}$ is an integer multiple of the typical business cycle, thereby maintaining the communication clock synchronization. Consequently, the AGV dynamically adjusts its closed-loop interaction cycle, which reduces unnecessary interactions between AGVs in empty areas and BSs, optimizes network resource utilization, and minimizes average closed-loop latency.

When the closed-loop communication latency is significantly smaller than the closed-loop interaction cycle, the control commands are not updated within the current cycle. Consequently, the agent continues to execute the previous control decision, and its motion state is updated as

$$\mathbf{p}^{(n)}(t+1) = \mathbf{p}^{(n)}(t) + T_{Ct}^{(n)}(t)\cdot\mathbf{J}(t)\cdot\mathbf{u}^{(n)}(t). \tag{26}$$

Case 2: When multiple agents are formed in areas with high CVaR risk values, such as assembly areas and high-density AGV areas, the control system trajectory is highly susceptible to environmental changes. As the AGV relies

on its local physical movement information, it is prone to rapid accumulation of pose errors and synchronization errors. Then, AGVs heavily depend on the BSs transmitted dynamic obstacle avoidance risk map, high-precision pose information, and low-latency control decisions. The global information enables AGVs to mitigate potential congestion collision risks, calibrate trajectory and pose information, and maintain formation shape synchronization. Consequently, the AGV training outcome is that the closed-loop control cycle approximates the control business cycle time, prioritizing strict formation shape and safe formation control, $T_{Ct}^{(n)}\left(o_t^{(n)}\right) \approx T_{cyc}$.

When the control delay is approximately equal to the closed-loop interaction cycle, the control command can be updated within each cycle, and the agents' motion state follows the same form as in (12). Additionally, since AGVs cannot receive global information from the BS within the closed-loop interaction cycle, the interaction cycle reward function in the local actor-critic network of the AGV remains unchanged. Consequently, the interaction cycle calculated locally within the cycle interval is approximately a constant value.

In summary, the proposed communication mechanism leverages global state information from ISCC to dynamically adjust AGV communication cycles. We filter redundant or low-value control instructions, reduce AGV access frequency to the BS, and mitigate network congestion and closed-loop latency caused by excessive access. Additionally, stable low-latency communication minimizes suboptimal decisions from delayed global state data, improves estimation accuracy, and enhances critic networks' value function evaluation, leading to more efficient policy convergence. The following simulation results show that this mechanism achieves stronger generalization capabilities. The dynamic communication cycle transmission scheme also supports training and execution across diverse intelligent machine scenarios, making the policy more adaptable to complex and dynamic environments.

3.3 ISCC-based MAPPO formation control strategy

Solving problem (19) poses a challenge due to the involvement of discrete and continuous decision variables. To address the non-convex mixed-integer nonlinear optimization problem, we formulate the multi-agent formation control problem as a partially observable Markov decision process under sensing, communication, and control constraints. We propose an MAPPO formation control strategy based on ISCC global information, which leverages communication and sensing system information, such as risk maps and interaction cycles, as ISCC state observation values to enhance the training efficiency of multi-agent formation control. Next, by learning optimal communication policies in complex dynamic environments, the agents dynamically adjust their communication cycles to ensure synchronization precision while minimizing communication overhead due to additional ISCC global state. In ISCC-MAPPO, we exploit high-dimensional state information from the ISCC system and differential collaborative learning among heterogeneous leader agents and follower agents. Specially, to achieve cooperation between heterogeneous agents, we introduce the centralized training and decentralized execution (CTDE) framework for training. Meanwhile, unlike common shared networks, in the CTDE framework, we decouple the actor-critic network into two independent agent networks to separately compute decisions for leader and follower agents, avoiding gradient conflicts and improving training efficiency. Several rewards such as obstacle avoidance, formation control, navigation and interaction overhead are adopted to handle continuous and discrete action spaces, and improve the performance of the ISCC-MAPPO algorithm as well.

POMDP can be represented as $\langle N, S, A, P, r, O, Z\rangle$, where N denotes the number of agents, S represents the global state, A is the action chosen by each agent, P is the state transition function, r is the global joint reward function, O is the local observation of the environment obtained by each agent, and Z is the observation function. Each AGV agent has a limited view of its surroundings and must sense enough environmental information to update owner credibility about the current state, thereby informing its decision-making process for subsequent action selection. As illustrated in Figure 4, cooperative transportation in industrial environments is primarily achieved through the collaboration of two types of agents: leader agents and follower agents. Although these agents share the same state and action spaces, their reward mechanisms differ, necessitating the design of distinct actor networks to learn their respective policies. This approach avoids conflicts between gradient directions, accelerates convergence, and enhances the algorithms' dynamic adaptability and generalization capabilities. Notably, the critic network can be shared among agents, enabling each of them to learn the most suitable strategy for its role.

(1) State space: The state space is a fundamental component of multi-agent formation control systems, as it accurately describes the local state of AGVs and reflects their real-time environmental and self-state information in complex environments. The state space of an agent should encompass various key local information, including the AGVs pose $\mathbf{p}^{(n)}$, velocity $v^{(n)}$, formation parameters $\mathbf{H}_n^f$, the upper bound of the closed-loop interaction cycle $\hat{T}_{Ct}^{(n)}$, and global risk map data $CVaR_\alpha$. The ISCC observation state $o_t^{(n)}$ of agent n is expressed as

$$o_t^{(n)} = \left\{\mathbf{p}^{(n)}, v^{(n)}, \mathbf{H}_n^f, \hat{T}_{Ct}^{(n)}, CVaR_\alpha\right\}. \tag{27}$$

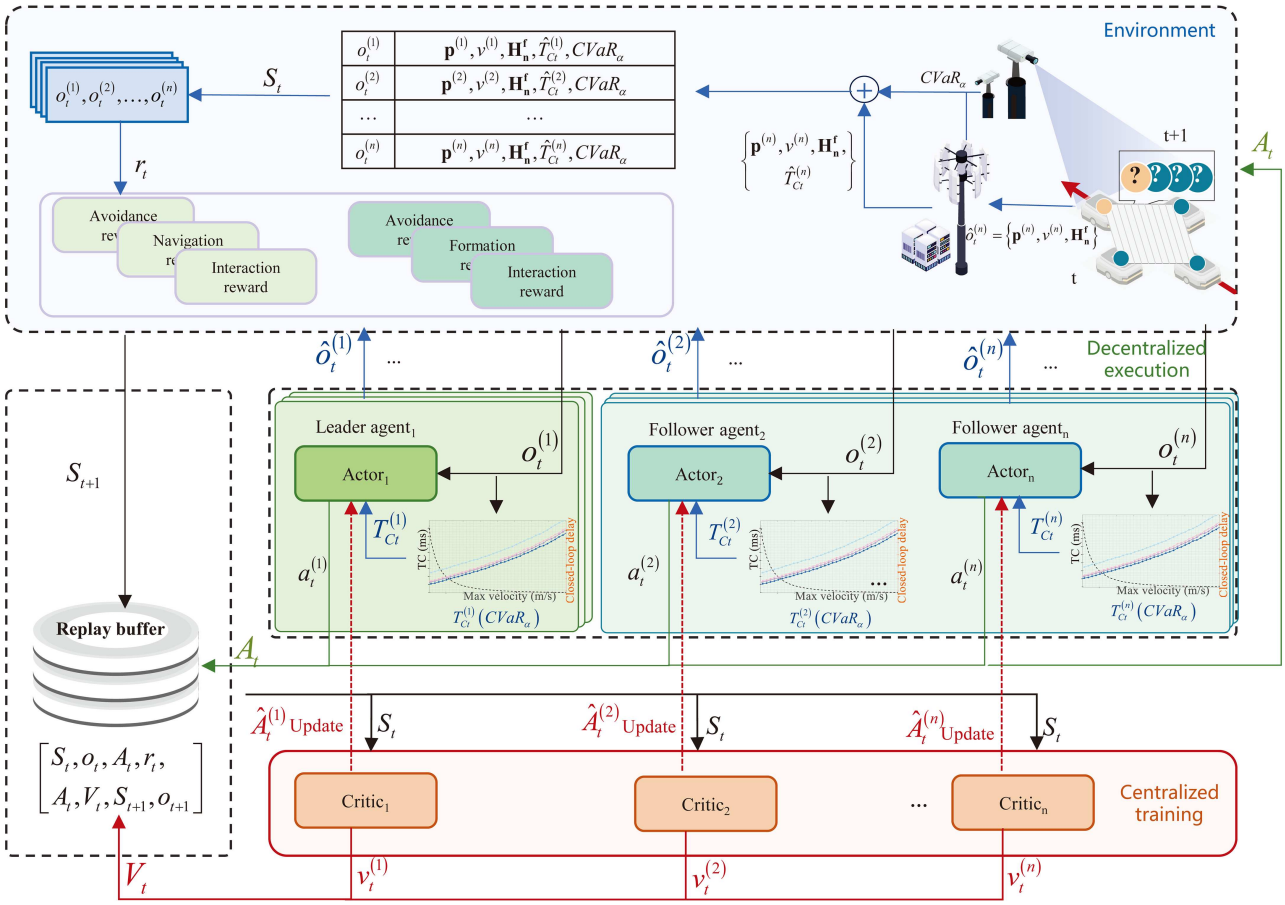

Figure 4 (Color online) MAPPO control strategy based on ISCC global information in flexible manufacturing workshop.

Additionally, the global information S_t is defined as $S_t = \left\{o_t^{(1)}, o_t^{(2)}, \ldots, o_t^{(n)}\right\}$.

(2) Action space: The design of the action space should enable AGVs to learn and optimize their behavior strategies in dynamic network topologies and industrial environments. In complex environments with narrow corridors, dense material areas, and dynamic and static obstacles, AGVs must constantly adjust their motion state to minimize travel time while adhering to sensing, communication, and control constraints. Therefore, the action space should be designed to account for various situations that AGVs may encounter during task execution, allowing the agent to flexibly respond to environmental changes. The action space a_t of the multi-AGV system comprises continuous values, velocity vectors, and discrete values, as well as closed-loop interaction cycles, which are expressed as

$$a_t^{(n)} = \left\{\mathbf{u}^{(n)}, T_{Ct}^{(n)}\right\}. \tag{28}$$

By adjusting these parameters, multi-agent can achieve various behavior choices, including acceleration, deceleration, steering, and adjusting the communication interaction cycle, to adapt to dynamic network topology and industrial environments. The joint action space is defined as $A_t = \left\{a_t^{(1)}, a_t^{(2)}, \ldots, a_t^{(n)}\right\}$.

(3) Reward function: To mitigate gradient conflicts among heterogeneous agents during ISCC-MAPPO training, we decouple the actor-critic framework into distinct agent networks, each learning collaborative transportation policies. Meanwhile, as AGV systems in flexible manufacturing workshops should balance requirements including safe obstacle avoidance, smooth movement, and precise synchronization, we implement differentiated reward functions. Specifically, the leaders reward consists of obstacle avoidance reward, navigation reward, and interaction reward, aimed at guiding it to the target point while avoiding obstacles and optimizing interaction costs. The follower reward comprises obstacle avoidance reward, formation reward, and interaction reward, with the goal of maintaining formation with the leader while avoiding obstacles and optimizing interaction costs.

(a) CVaR-based obstacle avoidance: In complex environments with narrow corridors, messy material areas, and dynamic and static obstacles, high-delay closed-loop control commands and untimely obstacle avoidance braking

can increase the risk of collision. To mitigate this, an obstacle avoidance penalty is set based on the risk value $CVaR_\alpha\left(P_{coll}^{(n)}\right)$. Specifically, when the risk value falls below the safety threshold, a larger negative reward is imposed to incentivize the AGV to maintain a safe distance from dynamic obstacles, select low-risk areas for path planning, and prevent reduced transportation efficiency resulting from collisions or congestion. This reward function design improves the overall average driving speed and reduces driving time while maintaining driving safety. The obstacle avoidance reward function is

$$r_{obs}=\begin{cases}\lambda_{\text{cvar}}\cdot e^{\frac{\varepsilon_{cv}}{CVaR_\alpha\left(P_{coll}^{(n)}\right)+\varepsilon_{cv}}}, & CVaR_\alpha\left(P_{coll}^{(n)}\right)\leqslant\varepsilon_{cv},\\ -\lambda_{\text{cvar}}\cdot e^{1-\frac{\varepsilon_{cv}}{CVaR_\alpha\left(P_{coll}^{(n)}\right)}}, & \text{otherwise},\end{cases}\tag{29}$$

where λ_{cvar} is the obstacle avoidance sensitivity coefficient, and $CVaR_\alpha\left(P_{coll}^{(n)}\right)$ represents the conditional risk value under the confidence level α, reflecting the expected probability of collision in the worst-case scenario.

(b) Smooth formation synchronization: Unlike other multi-agent formation control systems with dynamic changing formations, the formation of AGV teams in flexible workshops requires a static and strict geometric formation, where AGVs must maintain high-precision synchronized motion with other AGVs in the team to avoid material damage. Therefore, the formation control reward prioritizes minimizing the synchronization error. When the synchronization error is lower than a threshold, the positive incentive of the reward function increases. If the synchronization error exceeds the maximum position deviation threshold allowed by the team, a penalty is imposed. The formation control reward is

$$r_{for}=\begin{cases}\omega_{for}\left(1-\frac{\left\|\hat{\mathbf{H}}^f(t)-\mathbf{H}^f(t)\right\|}{\omega_{for}\cdot\varepsilon_{es}^{th}}\right)^2, & \left\|\hat{\mathbf{H}}^f(t)-\mathbf{H}^f(t)\right\|\leqslant\varepsilon_{es}^{th},\\ -\omega_{for}\left(1-\frac{\left\|\hat{\mathbf{H}}^f(t)-\mathbf{H}^f(t)\right\|-\varepsilon_{es}^{th}}{\omega_{for}}\right), & \text{otherwise},\end{cases}\tag{30}$$

where ω_{for} is the formation control sensitivity coefficient, representing the degree of influence of the formation control reward on the immediate reward, and ε_{es}^{th} is the synchronization error threshold. The formation control reward is only applied to follower agents.

(c) Efficient navigation: Due to high communication delays and limited local observation, the current speed of AGVs in the workshop is greatly restricted, increasing travel time. Therefore, when designing the reward function, under the condition of ensuring synchronization error and obstacle avoidance risk, AGVs are encouraged to increase their speed to reduce travel time and improve control efficiency. The navigation reward function is expressed as

$$r_{nav}=\omega_{nav}v^{(n)}+3\omega_{nav}\left(d_t^{(n)}-d_{t-1}^{(n)}\right),\tag{31}$$

where ω_{nav} is the navigation sensitivity coefficient, representing the degree of influence of navigation reward on immediate reward. $d_t^{(n)}$ and $d_{t-1}^{(n)}$ denote the total distance traveled by the formation toward the destination at the current and previous moments, respectively. Furthermore, when the AGV is the leader, it needs to consider the formation error to adjust its own speed.

(d) ISCC-based communication cycle adjustment: While ensuring that the pose error and synchronization error are lower than the constrained threshold, AGVs traveling in low-risk areas are encouraged to adjust the closed-loop interaction cycle. Therefore, AGV will reduce the number of interactions with the BS, decrease the probability of data retransmission and communication latency, and reduce communication overhead. Meanwhile, it balances and improves the communication quality and environmental response velocity of AGVs traveling in high-risk areas. The interaction reward function is expressed as

$$r_{com}=\begin{cases}r_{for}\left(\omega_{com}\frac{T_{Ct}^{(n)}}{\hat{T}_{Ct}^{(n)}}\right)^{\left(1-CVaR_\alpha\left(P_{coll}^{(n)}\right)\right)}, & e_p^{(n)}\leqslant\varepsilon_{ep}^{th},e_s\leqslant\varepsilon_{es}^{th},\\ -r_{for}\left(\omega_{com}\frac{T_{Ct}^{(n)}}{\hat{T}_{Ct}^{(n)}}\right)^{\left(1-CVaR_\alpha\left(P_{coll}^{(n)}\right)\right)}, & \text{otherwise},\end{cases}\tag{32}$$

where ω_{com} represents the interaction coefficient, which quantifies the impact of the reward of each closed-loop interaction cycle on the immediate reward.

Then, the combined reward function for multiple agents can be defined as

$$r_t = \begin{cases} r_{obs} + r_{for} + r_{com}, \forall \mathrm{AGV}_{follower}, \\ r_{obs} + r_{nav} + r_{com}, \forall \mathrm{AGV}_{leader}. \end{cases} \tag{33}$$

Building on the reward function, we propose an ISCC-MAPPO formation control using ISCC global information, comprising two phases: training and execution in Figure 4.

During the training phase, each agent uploads its local observation $\hat{o}_t^{(n)}$ to the BS via the uplink and receives the ISCC global information observation $o_t^{(n)}$ from the BS via the downlink. Utilizing this information, the internal actor network generates the action $a_t^{(n)}$ within the continuous action space. The edge server constructs a global shared observation S_t based on the received ISCC global information observations from all agents and feeds it back to the critic network, which estimates the value function of the agent under the current state, $v_t^{(n)}$. The agent then stores the experience $[S_t, o_t, A_t, r_t, V_t, S_{t+1}, o_{t+1}]$ in its experience buffer and updates the loss function parameters of the actor and critic networks through random sampling.

In the execution phase, each agent's actor network calculates the motion control strategy and new communication interaction cycle based on the received current ISCC global information observation $o_t^{(n)}$ from the BS. Furthermore, within the communication interaction cycle, each agent generates the motion control strategy $\mathbf{u}^{(n)}$ and communication interaction cycle $T_{Ct}^{(n)}$ based on its local observation and previous ISCC global observation data.

Additionally, during the training phase, the actor network updates the joint advantage function $\hat{A}_t$ based on the global information of the agent in the critic network, enabling effective action decisions. Concurrently, the actor network parameters $\theta^{(n)}$ are updated using the clipped loss function $L(\theta)$ and gradient ascent optimization.

$$L\left(\theta^{(n)}\right) = \mathbb{E}_t\left[\min\left(r_{\theta,t}^{(n)}\hat{A}_t^{(n)}, clip\left(r_{\theta,t}^{(n)}, 1-\varepsilon, 1+\varepsilon\right)\hat{A}_t^{(n)}\right)\right], \tag{34}$$

where $r_{\theta,t}^{(n)} = \pi_\theta\left(a_t^{(n)}\middle|o_t^{(n)}\right)/\pi_{\theta\ old}\left(a_t^{(n)}\middle|o_t^{(n)}\right)$ represents the importance sampling proportion coefficient, which is the ratio of the probability of the new policy to the old policy of the actor. The clipping function $clip()$ is used to limit the range of $r_{\theta,i}^{(n)}$ and prevent excessively large update steps, with ε denoting the clipping range. Additionally, $\hat{A}_t^{(n)}$ represents the generalized advantage estimation (GAE) calculated for each time step. The advantage function is utilized to evaluate the feasibility of selecting a specific action $a_t^{(n)}$ under a certain state $o_t^{(n)}$, given by

$$\hat{A}_t^{(n)} = r_{\theta,t}^{(n)} + \gamma V\left(o_{t+1}^{(n)}\right) - V\left(o_t^{(n)}\right), \tag{35}$$

where $V\left(o_t^{(n)}\right)$ is the value function, and γ is the discount coefficient. Furthermore, the actor network parameters of the leader and follower agents are dynamically updated based on the type of AGV agent. Considering the critic network is typically shared, take the global observation as input to estimate the value function, thereby eliminating the need for distinction.

In the critic network, the primary objective is to minimize the loss function $L(\varphi)$, and the gradient descent optimization is performed to update the network parameters $\left[\varphi^{(n)}\right]$, as expressed in

$$L\left(\varphi^{(n)}\right) = \mathbb{E}_t\left[\left(V_{t,\varphi}\left(o_t^{(n)}\right) - R_t\right)^2\right], \tag{36}$$

where $V_{t,\varphi}$ represents the estimated value function of the critic network for the state $o_t^{(n)}$, and R_t denotes the accumulated return, given by $R_t = r_{\theta,t}^{(n)} + \gamma V\left(o_{t+1}^{(n)}\right)$. The MAPPO multi-agent formation control algorithm based on ISCC global information is outlined in Algorithm 1.

Finally, to illustrate the interplay between sensing, communication, and control in the closed-loop control process of multi-AGV formation, we introduce high-dimensional feature state values, such as the upper bound of communication closed-loop interaction cycle and the sensed CVaR risk value, into the proposed MAPPO formation control strategy. By substituting the original low-order observation data with higher-order information in the local observation, including relative distance, velocity, and angle of obstacles, as well as quality of communication channels, the state space dimension is reduced, thereby increasing the decision-making information available to the actor and critic networks and decreasing computational complexity. Specifically, assuming D fully connected layers in the actor network, with ω_θ^d neural units in the d-th layer, where the input and output layers have dimensions equal to the ISCC

Algorithm 1 ISCC-based MAPPO multi-agent formation control algorithm.

Initialize the actor network parameters $\theta^{(n)}$ and critic network parameters $\varphi^{(n)}$, learning rate α_{ISCC}.
while *episodes* $\leqslant M_{eps}$ **do**
  for *step* $= 1$ **to** L **do**
    Initialize trajectory list τ;
    Initialize RNN states $h_{0,\pi}^{(1)}, ..., h_{0,\pi}^{(n)}$ and $h_{0,V}^{(1)}, ..., h_{0,V}^{(n)}$;
    for $t = 1$ **to** T **do**
      for each agent $n \in \{1, ..., N\}$ **do**
        Generate ISCC state $o_t^{(n)}$ from $\hat{o}_t^{(n)}, CVaR_\alpha, \hat{T}_{Ct}^{(n)}$;
        Execute action $a_t^{(n)} \leftarrow \pi(o_t^{(n)})$;
        Compute value $v_t^{(n)} \leftarrow V(s_t^{(n)})$;
      end for
      Observe reward r_t and new state s_{t+1};
      Store $\tau \leftarrow [S_t, o_t, A_t, r_t, V_t, S_{t+1}, o_{t+1}]$;
    end for
    Calculate $\hat{A}, \hat{R}$ via GAE;
    Store τ in D;
  end for
  Update RNN states for π and V;
  Update θ using $\theta \leftarrow \theta - \alpha_{ISCC} \nabla L(\theta)$;
  Update φ using $\varphi \leftarrow \varphi - \alpha_{ISCC} \nabla L(\varphi)$;
end while
for each agent $n \in \{1, ..., N\}$ **do**
  Obtain $\hat{o}_t^{(n)}$ and execute $a_t^{(n)}$ offline during $T_{Ct}^{(n)}$;
  Generate new $o_t^{(n)}$ after $T_{Ct}^{(n)}$ delay;
  Update $T_{Ct}^{(n)}$ and restart closed-loop control;
end for

global observation state space and action space, respectively, and ω_θ^r and ω_θ^t units in the ReLU and Tanh layers, respectively. The Critic network has a similar structure, with ω_θ^d neural units in the d-th layer, and ω_θ^r and ω_θ^t units in the ReLU and Tanh layers, respectively. When predicting the action of intelligent body $n \in N$, the complexity is caused by the forward propagation of action decision calculation, expressed as $\mathcal{O}\left(\sum_{d=1}^{D} \omega_\theta^d \omega_\theta^{d+1} + \sum_{d=1}^{D} \omega_\varphi^d \omega_\varphi^{d+1}\right)$.

During network training, the complexity is caused by the forward propagation of value function calculation and the backward propagation of network parameter gradient descent, which is $\mathcal{O}\left(\sum_{d=1}^{D} \omega_\theta^d \omega_\theta^{d+1}\right) + \mathcal{O}\left(\sum_{d=1}^{D} \omega_\varphi^d \omega_\varphi^{d+1}\right)$ and $\mathcal{O}\left(\omega_\theta^r + \sum_{d=1}^{D} \omega_\theta^d \omega_\theta^{d+1} + 6\omega_\theta^t + \omega_\varphi^r + \sum_{d=1}^{D} \omega_\varphi^d \omega_\varphi^{d+1} + 6\omega_\varphi^t\right)$. Additionally, the complexity of GAE is $\mathcal{O}\left(N^2 S_f\right)$, where N is the number of intelligent bodies and S_f is the vector dimension of the global state space. In summary, the complexity of the proposed MAPPO formation control strategy is $\mathcal{O}\left(\sum_{n=1}^{N} 2T\left(N_D + 1\right)\left(\sum_{d=1}^{D} \omega_\theta^d \omega_\theta^{d+1} + \sum_{d=1}^{D} \omega_\varphi^d \omega_\varphi^{d+1}\right)\right) + \mathcal{O}\left(TN^2 S_f\right)$.

4 Simulation and discussion

In this section, we use the Python and MATLAB platforms to simulate the 100 m × 100 m environment of a flexible assembly workshop for automobiles. In the simulated assembly workshop, a diamond formation of 4 AGVs as a baseline case, we generate 7 distinct multi-agent formations and deploy them randomly across multiple ring-shaped production lines. Then, we simulate the multi-agent formation performing cooperative transportation tasks. The formations' traveling trajectory and motion state data are recorded and analyzed, to validate the safety and scalability of the proposed algorithm. Furthermore, we conduct comparative analysis against both learning-based traditional MAPPO and MADDPG strategies, as well as kinematic model-based virtual structure approach and leader-follower approach, evaluating formation synchronization accuracy, closed-loop latency, and travel time in dynamic environments. Simulation parameters are detailed in Table 1.

Simulation results comparing the convergence of learning effects in formation control systems under varying AGV scales within the ISCC system. As shown in Figure 5, the x-axis denotes the number of training episodes, while the y-axis represents the total reward. The figure comprises five curves, corresponding to deployments of 8, 16, 24, 28, and 32 agents per thousand square meters. The solid lines indicate the average reward function values derived from five parallel simulations conducted in distinct random environments, with the shaded regions delineating the upper and lower bounds of the positive and negative standard deviations. It is evident that as the number of controllable agents increases from 8 to 28, the convergence value of the reward curves rises progressively, accompanied by an accelerated convergence rate. Concurrently, the shaded areas narrow, signifying enhanced

Table 1 Simulation parameters.

Symbol	Simulation parameter	Value
ε_{es}^{th}	Formation synchronization error threshold	30 mm
T_{cyc}	Cycle time	4 ms
N	Number of AGVs	28 /km^2
v	Velocity of AGVs	1–5 m/s
h_s	Safe braking distance of AGVs	5 m
$D_U^{(n)}$	Average data value of uplink sensing signal	40–250 bytes
ε	Decoding error rate of short packets	1e−5
M_{eps}	Maximum episode times	400
L	Maximum learning step times	500

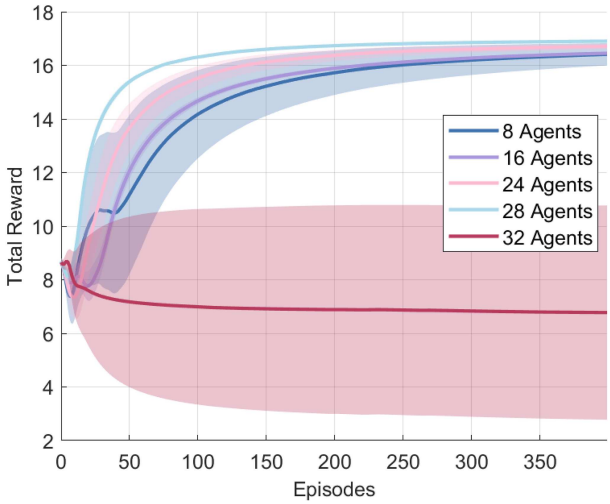

Figure 5 (Color online) Training results of the ISCC-based formation control strategy under different AGV scales in dynamic and complex environments; each curve (8, 16, 24, 28, 32 AGVs per 1000 m^2) shows the smoothed mean reward versus training episodes with shaded 90% confidence intervals over five random seeds.

algorithmic convergence and stability. The improvement is attributed to agent collaboration and latent information sharing, such as updates to the CVaR dynamic risk map, which bolsters the systems' capabilities in safe obstacle avoidance and navigation. Furthermore, a greater number of formations expands the environmental area accessible for learning by the actor-critic network, enabling the system to more comprehensively explore control decisions across diverse environmental feedback scenarios. This reduces the likelihood of decisions converging to local optima, thereby expediting the learning process and enhancing the convergence speed of rewards associated with obstacle avoidance, navigation, and formation control. Additionally, the different formation scales used in Figure 5 can be viewed as analogous to varying operational loads in real industrial workshops. Smaller-scale formations correspond to low-load conditions with sparse AGV interactions, whereas larger formations represent high-load scenarios characterized by dense coordination in confined spaces. The consistent improvement in convergence and stability across these varying scales demonstrates the robustness and adaptability of the proposed ISCC-based MAPPO algorithm to diverse operational environments, thereby strengthening the argument for its practical generalizability.

However, when the agent scale increases to 32, the rewards decline gradually, and the shaded areas widen. This degradation stems from the limited network resources that are not able to support the escalated communication overhead induced by the surge in agents. Delayed sensing and control decisions result in significant synchronization errors in pose estimation, elevating the risk of congestion and collisions, which in turn diminish rewards for navigation and formation tasks. When the number of agents per unit area exceeds the ideal capacity threshold of the networked control system, the multi-agent system struggles to meet the ultra-low latency and efficient coordination demands of complex formation tasks. It is evident that the proposed algorithm effectively characterizes the coupling among sensing, communication, and control, enabling an expansion of the controllable agent population, reducing communication overhead and latency, and further enhancing control performance in navigation and obstacle avoidance.

We simulate the collaborative material handling process of 28 AGVs in an unknown dynamic environment. Figure 6 depicts the movement trajectories and motion state trends of four AGVs within the diamond-shaped

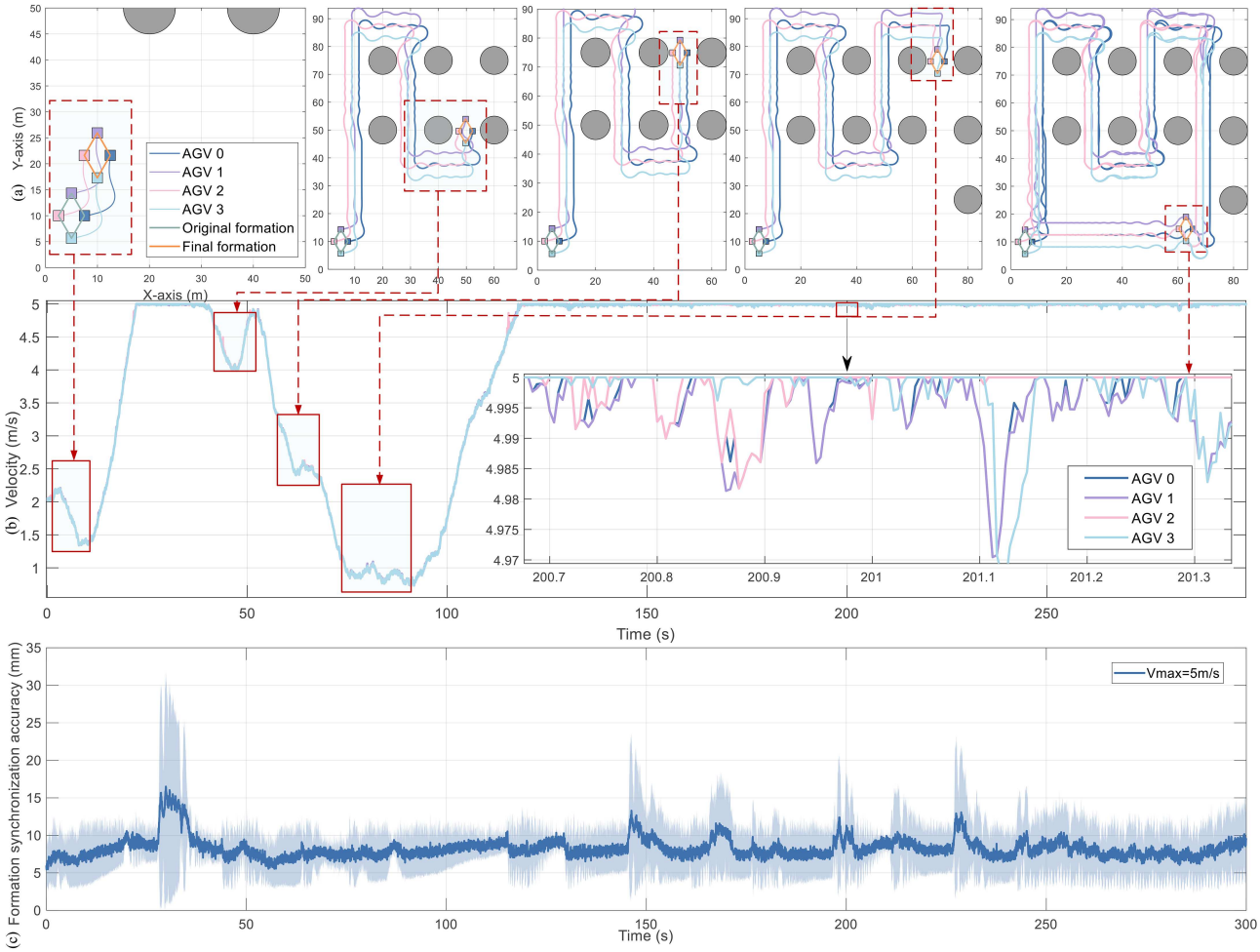

Figure 6 (Color online) The multi-agent formation dynamically adjusts its motion state to accommodate complex environmental changes involving both dynamic and static obstacles. Three key metrics illustrate the formation control process: (a) movement trajectories of agents, (b) linear velocity variations across agents, and (c) formation synchronization accuracy.

formation No.1. As observed from Figure 6(a), upon receiving a collaborative handling request at the initial moment, formation No.1 reduces its speed to maintain its configuration while seeking the shortest path to the destination. After 55 s, Figure 6(b) reveals a sharp decline in the formation's speed, as indicated by the velocity fluctuation curve. The formation anticipates that the next path will pass through point (50, 70) based on the obtained global CVaR risk map information. Specifically, unknown static obstacles near this point lie within the safety clearance threshold of the formation planned route. Meanwhile, other AGV formations in the vicinity are treated as dynamic obstacles, presenting a high risk of potential congestion and collision. Consequently, the leader AGV in formation No.1 cannot accurately determine the probability of an impending collision. To mitigate the risk of load displacement or damage caused by abrupt speed changes, it initiates a gradual and smooth deceleration. To preserve strict geometric synchronization, the follower AGVs in the formation also reduce their speed accordingly.

After 80 s, as the formation moves close to the exit of the obstacle zone, it progressively accelerates to its maximum speed. Subsequently, at 140 s, the formation passes through point (50, 70) again. Based on learned optimal control strategies from prior operations in similar environments, formation No.1 determines this area to be free of safety threats. The decision to proceed without deceleration markedly enhances its adaptability and travel efficiency in a complex, unknown dynamic environment.

Furthermore, Figure 6(c) highlights that when the formation first reaches a maximum speed of 5 m/s around 25 s, the synchronization accuracy of some formations approaches the predefined safety threshold of 30 mm. Here, the synchronization accuracy is computed from (17), where h_3 and h_4 denote relative positional drift between formations. The resulting formation synchronization error approximates the formation synchronization accuracy in millimeters. However, by 120 s, when the maximum speed of 5 m/s is achieved again, the synchronization accuracy across all formations stabilizes at approximately 10 mm. This improvement is twofold. First, the proposed algo-

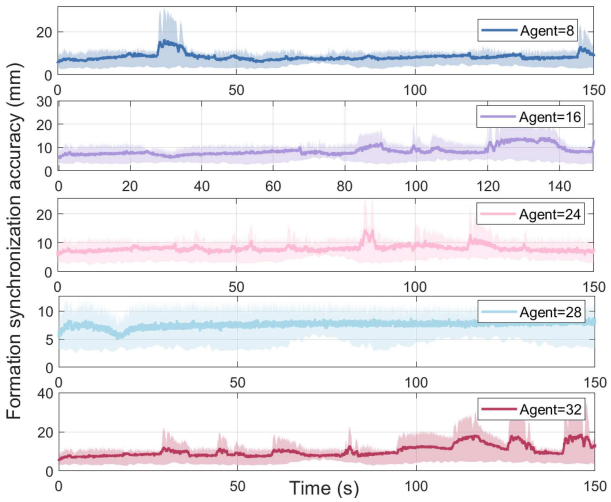

Figure 7 (Color online) Training results of formation synchronization accuracy under different numbers of agents in dynamic and complex environments; each curve (8, 16, 24, 28, 32 AGVs per 1000 m^2) shows the smoothed mean reward with shaded 90% confidence intervals over five random seeds.

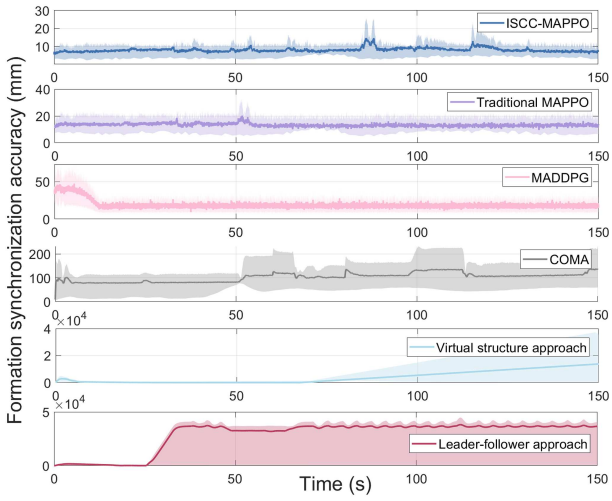

Figure 8 (Color online) Comparison of formation synchronization accuracy for different algorithms in dynamic and complex environments. The curves show the smoothed mean and 90% confidence interval over five runs for the proposed ISCC-MAPPO method (blue) and baseline algorithms, including traditional MAPPO (purple), MADDPG (pink), COMA (gray), virtual structure (light blue), and leader-follower (red).

rithm leverages global information from the ISCC system to construct a CVaR risk map, quantifying the severity of potential congestion and collisions between AGVs and dynamic obstacles. The foresight allows sufficient time for smooth speed adjustments and formation synchronization, enabling AGVs to maintain high speeds without risking material deformation due to emergency braking or startup. Concurrently, the proposed dynamic communication cycle allocation mechanism, based on networked control states, ensures high-precision synchronization while minimizing communication overhead, thus avoiding high pose errors arising from delayed control decisions or sensing data. Second, by exploiting high-dimensional state information from the ISCC system and differential collaborative learning among heterogeneous AGVs, the algorithm refines optimal control strategies from historical environmental interactions, substantially improving formation synchronization accuracy and traveling efficiency.

Figure 7 illustrates the average formation synchronization accuracy across five parallel random environments, with the solid line representing the mean and the shaded areas delineating the positive and negative standard deviations. The results demonstrate that the strategy proposed in this study achieves a stable synchronization accuracy of 8 mm when 28 agents are deployed per unit area, significantly outperforming the performance at other deployment scales. This superior performance arises because, in multi-agent systems, a higher number of agents can leverage the global information provided by the ISCC framework to more comprehensively explore control decisions across diverse environments. It reduces the risk of converging to local optima and accelerates the learning process for collaborative transportation tasks. Furthermore, by dynamically adjusting the communication cycles of agents based on their networked control states, the proposed approach substantially mitigates the communication overhead resulting from frequent data exchanges during learning and decision-making. Meanwhile, the system strikes an optimal balance between the communication overhead sustainable by the network and the efficiency of the learning algorithm. The 28 agents are identified as the ideal threshold for the unknown environmental capacity under consideration. Consequently, agents at this scale can more rapidly study optimal control strategies from historical environmental interactions, efficiently execute tasks, and achieve high rewards. It enables high mobility and stringent geometric formation synchronization in dynamic and complex environments.

Figure 8 reveals that the proposed strategy maintains a stable synchronization accuracy of 10 mm, markedly surpassing the performance of learning-based traditional MAPPO and MADDPG strategies. The advantage stems from the strategy-integrated design, which enhances formation control by synergizing sensing, communication, and control. By fully exploiting global sensing information to construct a CVaR map, the approach effectively mitigates potential collision risks, significantly reducing the likelihood of emergency braking. It provides sufficient time for smooth velocity adjustments and formation synchronization. Additionally, based on state observations from the ISCC framework, the multi-agent system dynamically adjusts the interaction cycles with the BS, enhancing communication resource utilization and minimizing latency. It allows agents to capture dynamic environmental

Table 2 The average simulation results for six test parameters of the proposed algorithm and the benchmark learning-based formation control strategy are presented. The improvement effect indicated in parentheses denotes the lower bound of the performance enhancement achieved when comparing the proposed algorithm to the other two algorithms.

Path length (m)	Traveling time (s)	Cycle time (ms)	Closed-loop latency (ms)	Formation sync. accuracy (mm)	AGV velocity (m/s)	Supported AGV scale
Proposed algorithm						
20	11.63(+26.7%)	4.62(+15.5%)	3.73(−52.7%)	1.78(−59.9%)	2.07(−22.2%)	
40	19.42(−5.2%)	4.69(+17.3%)	2.72(−67.6%)	1.72(−71.4%)	2.27(+0.9%)	
60	24.90(−11.5%)	4.68(+17.0%)	2.45(−72.8%)	1.51(−82.3%)	2.58(+13.7%)	28(+40%)
80	29.10(−14.4%)	4.67(+16.8%)	3.18(−61.3%)	2.01(−76.5%)	2.87(+17.1%)	
100	33.03(−16.8%)	4.64(+16.0%)	2.82(−65.9%)	2.64(−69.1%)	3.12(+17.7%)	
Traditional MAPPO						
20	11.34		7.89	15.01	2.12	
40	20.48		8.39	16.25	2.14	
60	28.13	4	9.46	16.50	2.27	20
80	33.99		8.22	16.60	2.45	
100	39.68		8.98	16.00	2.65	
MADDPG						
20	9.18		9.01	4.44	2.66	
40	23.14		8.63	6.02	2.29	
60	217.99	4	9.02	8.53	0.81	20
80	230.63		8.63	8.54	1.04	
100	234.98		8.27	8.55	1.11	
COMA						
20	9.71		9.66	226.17	1.50	
40	22.96		10.9	226.17	1.50	
60	37.06	4	10.56	231.93	1.50	16
80	50.40		10.30	235.92	1.56	
100	62.42		10.41	341.23	1.50	

changes with minimal delay and assess real-time variations in formation structure, leading to more precise synchronization decisions and improved synchronization accuracy and control efficiency. Meanwhile, as shown in Figure 8, COMA [39] achieves fast initial convergence but exhibits large fluctuations and higher steady-state error. This is mainly because COMA relies on a centralized critic with counterfactual baselines for credit assignment, optimizing policies purely from global rewards without modeling the coupling among communication, perception, and control. Such a decoupled design limits responsiveness and synchronization accuracy in dynamic industrial environments. In contrast, the proposed ISCC-MAPPO integrates communication-perception-control synergy and adaptive interaction-cycle optimization, achieving higher precision and stability in formation control. Moreover, as depicted in the figure, the leader-follower approach sustains high-precision formation synchronization during the initial 25 s of the environment. However, as time progresses, system struggles to maintain strict geometric formations. The limitation arises because kinematic model-based formation control methods rely on pre-defined rules for decision-making, which ensure system stability in rule-compliant environments. In an unpredictable or dynamically evolving environment, inaccuracies in the model and elevated communication latency undermine control effectiveness, failing to guarantee the validity of all decisions in complex, unknown environments. It is evident that the proposed algorithm effectively characterizes the interaction relationship among sensing, communication, and control, reducing communication overhead and latency. And it can further enable an expansion of the controllable agent population, and further improve control performance in navigation and obstacle avoidance.

As shown in Table 2, when multiple agents execute a collaborative transportation task along a 100-meter path, the average travel time is 33.03 s, representing a reduction of at least 16.8% compared to baseline algorithms. Particularly, although COMA employs a centralized critic to perform credit assignment among agents, it struggles to accurately model individual contributions in continuous action spaces, resulting in longer travel times and unstable convergence in complex dynamic environments. Unlike the traditional MAPPO, MADDPG, and COMA baselines, which rely primarily on low-dimensional motion states such as sensed positions and velocities, the proposed algorithm emphasizes the interdependencies among sensing, communication, and control processes. In contrast, the proposed algorithm leverages the rich global information provided by ISCC to jointly learn optimal strategies for continuous

control and discrete communication cycles across diverse conditions. Consequently, agents can maintain higher movement speeds without compromising safety, reducing the risk of material deformation during braking and acceleration, which significantly improves formation velocity and reduces overall travel time.

In Table 2, cycle time denotes the communication interval between agents and the BS, where a larger value corresponds to lower communication traffic per unit time, approximating reduced network overhead. The results demonstrate that the proposed algorithm, by dynamically adjusting cycle time, achieves at least 15.5% reduction in communication overhead compared to baseline methods, thereby alleviating communication congestion. During networked formation control, the closed-loop latency decreases by at least 52.7%, substantially improving communication performance. The enhancement arises because the proposed algorithm enables agents to learn both optimal control strategies, leveraging comprehensive ISCC global information, and optimal communication strategies tailored to current control demands. A dynamic communication cycle allocation mechanism, allows AGVs to adjust interaction cycles based on ISCC state observations within the action space. It minimizes redundant information exchanges, ensuring high-precision formation synchronization while significantly reducing communication overhead and closed-loop latency. In contrast, traditional MAPPO, MADDPG, and COMA algorithms overlook the varying communication needs of control performance in complex environments, employing fixed-cycle communication strategies ill-suited to dynamic, intricate formation tasks. For instance, unlike in assembly areas with complex or dense dynamic obstacles, agents traveling in open, non-assembly zones require less frequent BS interactions for strict obstacle avoidance and navigation, resulting in lower communication demands. However, the baseline algorithms allocate uniform communication resources to agents, leading to redundant high-frequency data exchanges, wasted resources, elevated delays, and diminished control efficiency and safety due to delayed decision-making and sensing.

Moreover, the proposed algorithm surpasses baseline methods in critical metrics such as formation synchronization accuracy and AGV speed, while also supporting larger-scale AGV formations, showcasing superior scalability. These benefits stem from the multi-dimensional optimization of high-precision formation control enabled by the integrated sensing, communication, and control design. The approach allows agents to efficiently acquire rich ISCC global information with minimal communication overhead, facilitating rapid and effective learning of optimal control and communication strategies in dynamic, complex environments. While the current experiments focus on an automobile assembly scenario, the proposed ISCC-based framework is expected to be adaptable to other formation control production settings, such as furniture assembly and warehousing logistics, where low-latency and high-synchronization coordination are equally important. By reparametrizing the kinematic and environmental variables, the approach could potentially maintain comparable control efficiency across different industrial contexts. Future work will further validate the proposed algorithm in these diverse collaborative manufacturing environments.

5 Conclusion

In this study, we propose a novel ISCC-based multi-agent formation control strategy tailored for flexible automotive assembly scenarios. By considering a new form of communication-sensing enhanced control, we develop an ISCC formation control model and construct a dynamic obstacle avoidance risk map based on CVaR to quantify collision risks, improve sensing accuracy and sensing range, and ensure smooth velocity transitions, and mitigate material extrusion deformation risks without compromising velocity. The formation control problem is formulated as a POMDP and solved using an ISCC-MAPPO algorithm. To reduce communication overhead and sensing errors, we introduce a dynamic communication cycle allocation mechanism that enhances efficiency and expands the number of controllable AGVs. A decoupled actor-critic framework further improves training stability and mitigates gradient conflicts in heterogeneous agents. Simulations show our work reduces synchronization error by at least 59.9% and communication overhead by 15.5%, while supporting larger AGV fleets and reducing travel time. In future work, we will extend the ISCC framework by incorporating computational resource constraints, aiming to develop an integrated sensing, communication, computation, and control paradigm for large-scale intelligent systems. This effort will provide advanced theoretical insights and technical solutions to support large-scale intelligent machine group control in flexible manufacturing workshops.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 92267202, 62321001, U25A20390), National Key Research and Development Program of China (Grant No. 2020YFA0711302), and BUPT Excellent Ph.D. Students Foundation (Grant No. CX2023146).

References

1 Chen Q, Wang Y, Jin Y, et al. A survey of an intelligent multi-agent formation control. Appl Sci, 2023, 13: 5934
2 Pan Z, Sun Z, Deng H, et al. A multilayer graph for multiagent formation and trajectory tracking control based on MPC algorithm. IEEE Trans Cybern, 2022, 52: 13586–13597

3 Mathew E. Swarm intelligence for intelligent transport systems: opportunities and challenges. In: Swarm Intelligence for Resource Management in Internet of Things. Pittsburgh: Academic Press, 2020. 131–145
4 Lajoie P Y, Beltrame G. Swarm-SLAM: sparse decentralized collaborative simultaneous localization and mapping framework for multi-robot systems. IEEE Robot Autom Lett, 2024, 9: 475–482
5 Dong C, Xiong X X, Xue Q L, et al. A survey on the network models applied in the industrial network optimization. Sci China Inf Sci, 2024, 67: 121301
6 Li Z X, Cui G H, Li C L, et al. Comparative study of slam algorithms for mobile robots in complex environment. In: Proceedings of the 2021 6th International Conference on Control, Robotics and Cybernetics (CRC), Shanghai, 2021. 74–79
7 Bernardini F, Buffi A, Motroni A, et al. Particle swarm optimization in SAR-based method enabling real-time 3D positioning of UHF-RFID tags. IEEE J Radio Freq Identif, 2020, 4: 300–313
8 Liu Z, Chen W, Lu J, et al. Formation control of mobile robots using distributed controller with sampled-data and communication delays. IEEE Trans Contr Syst Technol, 2016, 24: 2125–2132
9 Huang Z, Pan Y J, Bauer R. Edge-based communication-triggered formation tracking control with application to multiple mobile robots. IEEE Trans Contr Syst Technol, 2024, 32: 1015–1026
10 Boughellaba M, Tayebi A. Bearing-based distributed pose estimation for multi-agent networks. IEEE Control Syst Lett, 2023, 7: 2617–2622
11 Zhang Y, Li S, Wang S, et al. Distributed bearing-based formation maneuver control of fixed-wing UAVs by finite-time orientation estimation. Aerospace Sci Tech, 2023, 136: 108241
12 Chen W, Chen L, Mei J, et al. Displacement-based formation control with measurement noises. In: Proceedings of the 2023 62nd IEEE Conference on Decision and Control (CDC), Singapore, 2023. 5171–5176
13 Lian Y, Xie W, Yang Q, et al. Improved coding landmark-based visual sensor position measurement and planning strategy for multiwarehouse automated guided vehicle. IEEE Trans Instrum Meas, 2022, 71: 1–16
14 Song Z, Xie M, Huang H. Bearing-only formation tracking control for multi-agent systems with time-varying velocity leaders. IEEE Control Syst Lett, 2024, 8: 2027–2032
15 Bai C, Yan P, Pan W, et al. Learning-based multi-robot formation control with obstacle avoidance. IEEE Trans Intell Transp Syst, 2022, 23: 11811–11822
16 Miao Z, Zhong H, Lin J, et al. Vision-based formation control of mobile robots with FOV constraints and unknown feature depth. IEEE Trans Contr Syst Technol, 2021, 29: 2231–2238
17 Dai S L, Lu K, Fu J. Adaptive finite-time tracking control of nonholonomic multirobot formation systems with limited field-of-view sensors. IEEE Trans Cybernetics, 2022, 52: 10695–10708
18 Yu D, Li J, Wang Z, et al. An overview of swarm coordinated control. IEEE Trans Artif Int, 2024, 5: 1918–1938
19 Bekmez A, Aram K. Three dimensional formation control of unmanned aerial vehicles in obstacle environments. Balkan J Electrical Comput Eng, 2023, 11: 387–394
20 Chen F, Tang Y, Li N, et al. A study of collaborative trajectory planning method based on starling swarm bionic algorithm for multi-unmanned aerial vehicle. Appl Sci, 2023, 13: 6795
21 Wickramasinghe O E, Rajan N, Wanniarachchi C, et al. Evolution of formation control algorithms for unmanned aerial vehicles. In: Proceedings of the 2024 IEEE 33rd International Symposium on Industrial Electronics (ISIE), Ulsan, 2024. 1–7
22 Elamvazhuthi K, Kakish Z, Shirsat A, et al. Controllability and stabilization for herding a robotic swarm using a leader: a mean-field approach. IEEE Trans Robot, 2021, 37: 418–432
23 3GPP. Technical Specification Group Services and System Aspects; Study on Communication for Automation in Vertical Domains, TS 22.804 V16.3.0, 2020. https://www.3gpp.org/ftp/Specs/archive/22_series/22.804
24 AII. 5G/5G-A ultra-low latency communication industrial scene white paper. Whitepaper, 2022. https://www.aii-alliance.org/uploads/1/20230703/cd2c4894dc550d834e57eaa70fdb6b40.pdf
25 Yang H, Wang L, Feng Z, et al. Dynamic power allocation for integrated sensing and communication-enabled vehicular networks. IEEE Trans Wireless Commun, 2024, 23: 12313–12330
26 Zhou X, Wen X, Wang Z, et al. Swarm of micro flying robots in the wild. Sci Robot, 2022, 7: eabm5954
27 Jiang C, Chen Z, Guo Y. Multi-robot formation control: a comparison between model-based and learning-based methods. J Control Decision, 2020, 7: 90–108
28 Zhao W, Liu H, Lewis F L. Robust formation control for cooperative underactuated quadrotors via reinforcement learning. IEEE Trans Neur Net Lear, 2021, 32: 4577–4587
29 Wei Z, Du Y, Zhang Q, et al. Integrated sensing and communication driven digital twin for intelligent machine network. IEEE Internet Things M, 2024, 7: 60–67
30 Yang H, Feng Z, Wei Z, et al. Intelligent computation offloading for joint communication and sensing-based vehicular networks. IEEE Trans Wireless Commun, 2024, 23: 3600–3616
31 You X H, Huang Y M, Zhang C, et al. When AI meets sustainable 6G. Sci China Inf Sci, 2025, 68: 171301
32 Li S, Song Q. Cooperative control of multiple AGVs based on multi-agent reinforcement learning. In: Proceedings of the 2023 IEEE International Conference on Unmanned Systems (ICUS), Hefei, 2023. 512–517
33 Xu Y, Zhou Y, Yao Z. Formation and collision avoidance via multi-agent deep reinforcement learning. In: Proceedings of the 2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS), Kaifeng, 2024. 156–161
34 de Sant Ana P M, Marchenko N, Soret B, et al. Goal-oriented wireless communication for a remotely controlled autonomous guided vehicle. IEEE Wireless Commun Lett, 2023, 12: 605–609
35 Vakaruk S, Sierra-Garcia J E, Mozo A, et al. Forecasting automated guided vehicle malfunctioning with deep learning in a 5G-based industry 4.0 scenario. IEEE Commun Mag, 2021, 59: 102–108
36 Jaroonsorn P, Neranon P, Dechwayukul C, et al. Performance comparison of compliance control based on PI and FLC for safe human-robot cooperative object carrying. In: Proceedings of the 2019 First International Symposium on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP), Bangkok, 2019. 13–16
37 Hakobyan A, Kim G C, Yang I. Risk-aware motion planning and control using CVaR-constrained optimization. IEEE Robot Autom Lett, 2019, 4: 3924–3931
38 Zhou Y, Feng Z, Song Z, et al. Integrated sensing, communication, and control driven multi-AGV closed-loop control. IEEE Trans Veh Technol, 2025, 74: 10853–10868
39 Foerster J, Farquhar G, Afouras T, et al. Counterfactual multi-agent policy gradients. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, 2018. 1–9