

Co-evolutionary Mechanism of Resource-Information Flows in Complex Industrial Systems: An AI-Enabled Framework for Manufacturing

Shutian Huang*

School of Economics and Management, Wuhan Technical College of Communications, Wuhan, Hubei, 430000, China

Abstract

Complicated manufacturing systems need to coordinate physical resources and information flow in a changing production environment to ensure the stable operation of the production process. Industrial Internet of Things (IIoT) technology provides important support for equipment perception and data transmission, but existing research usually optimizes resource scheduling and information transmission separately, leaving the bidirectional interaction between manufacturing-state changes and information freshness insufficiently represented in the decision process. Thus, this paper proposes a resource-information co-evolution framework (STCF) for artificial intelligence-driven manufacturing systems, in which high-order manufacturing-state representation, information freshness, and resource-information decision-making are jointly modeled and optimized. Specifically, a dynamic heterogeneous hypergraph is constructed to represent many-to-many relationships among machines, products, operators, and sensors and to characterize their time-varying resource-information dependencies. Subsequently, a Proximal Policy Optimization (PPO)-based multi-agent deep reinforcement learning mechanism jointly adjusts production-resource allocation and information-acquisition strategies according to changes in resource states and information freshness. Experiments are conducted on two publicly available industrial datasets, including the AI4I 2020 Predictive Maintenance Dataset and the NASA C-MAPSS dataset. The results show that STCF reduces total energy consumption by 14.2% and decreases the average Age of Information (AoI) by 21.8% compared with the genetic algorithm baseline. Furthermore, on AI4I 2020, the proposed framework improves the Resilience Score by approximately 18.3% compared with the standard DQN baseline under dynamic equipment conditions. These results indicate that explicitly coordinating resource scheduling and information acquisition can improve the balance among resource efficiency, information freshness, and operational resilience in complex manufacturing systems.

Keywords: resource-information co-evolution, heterogeneous hypergraph, multi-agent deep reinforcement learning, intelligent manufacturing, resilience

Received on 28 July 2026, accepted on 10 September 2026, published on 27 October 2026

Copyright © 2026 Shutian Huang *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [CC BY-NC-SA 4.0](#), which permits copying, redistributing, remixing, transformation, and building upon the material in any medium so long as the original work is properly cited.

doi: 10.4108/eetsis.14212

1. Introduction

Modern manufacturing systems are increasingly integrating production equipment, human operators, industrial sensors,

and cyber-physical control platforms to achieve higher efficiency, adaptability, and resilience. In this context, the Industrial Internet of Things (IIoT) provides essential sensing, communication, and data transmission capabilities for

*Corresponding author. Email: susiehwang8946@163.com

intelligent manufacturing infrastructures [1,2]. However, the increasing complexity of industrial processes introduces challenges in coordinating physical production activities and cyber information services. In complex industrial systems, the stability of the production process largely depends on the synergy between physical resource flows (such as materials, energy, and equipment status) and network information flows (such as sensor data, control signals, and computing resources). Recent graph-, hypergraph-, and knowledge-based manufacturing studies have further demonstrated the value of structured relational representations for capturing heterogeneous manufacturing entities and their interactions [3–5]. Due to factors such as equipment aging, production process fluctuations, changes in communication environments, and human operations, the resource flows and information flows need to be dynamically adjusted according to actual circumstances.

Currently, many studies still treat resource scheduling and information transmission as two relatively independent problems for optimization. Recent intelligent scheduling studies have applied knowledge-enhanced multi-agent reinforcement learning, IIoT-based reinforcement learning, integrated manufacturing control, and graph-assisted deep reinforcement learning to improve distributed resource coordination and dynamic scheduling [6–9]. For example, knowledge-enhanced multi-agent reinforcement learning has been used to support adaptive manufacturing scheduling [6], while single- and multi-agent reinforcement learning has been investigated for increasingly complex IIoT-enabled scheduling environments [7]. Other studies have integrated multi-agent learning with manufacturing system control [8] or combined structured process representations with deep reinforcement learning for real-time scheduling [9]. These approaches mainly focus on production efficiency, task allocation, resource utilization, or process control, while information management methods pay more attention to communication quality, data transmission reliability, and network-resource utilization. At the broader system level, AI-assisted manufacturing and human–automation-oriented smart manufacturing frameworks have also strengthened the integration of intelligent decision support, security, and human-centered production processes [10,11]. Nevertheless, when untimely information affects resource allocation or when changes in production status require the system to obtain new information, existing methods may still have difficulty coordinating resource scheduling and information acquisition as mutually dependent decisions.

In recent years, the development of digital twins and cyber-physical systems (CPS) has provided new approaches for connecting physical production processes with digital models [12,13]. By describing and feeding back the actual production status through digital models, the system can more promptly perceive and analyze production changes. However, many methods still rely on pre-defined system structures and physical-digital correspondence relationships. When the manufacturing environment changes or new interactions occur between multiple manufacturing entities, the models may have difficulty reflecting these changes in a timely manner. Recent intelligent and automated manufacturing

studies have further emphasized the growing complexity of customized manufacturing interactions and dynamic resource scheduling in highly automated production systems [14,15]. In practical manufacturing environments, machines, products, operators, and sensing devices often form high-order and many-to-many relationships, where changes in one component can influence multiple interconnected processes. These limitations may result in information synchronization delays, increased energy consumption, and reduced system resilience under operational disturbances. Therefore, developing an AI-enabled mechanism for jointly modeling and optimizing resource and information flows remains a key challenge for complex manufacturing systems.

Related approaches can therefore be broadly distinguished by their primary focus: graph-, hypergraph-, and knowledge-based methods emphasize structural representation of manufacturing entities [3–5]; multi-agent and deep reinforcement learning methods mainly address distributed scheduling and resource coordination [6–9]; and digital-twin, CPS, and intelligent manufacturing frameworks focus on physical–digital integration and adaptive production support [10–15]. Although these research streams provide complementary capabilities, the bidirectional interaction between manufacturing resource states and information freshness is not usually represented as a unified decision process in which resource changes drive information acquisition while updated information simultaneously influences subsequent resource allocation. Accordingly, the methodological gap addressed here concerns the integration of high-order manufacturing representation with explicit AoI-aware information freshness and bidirectional resource–information decision-making within a common optimization process.

To address this gap, this paper proposes a Resource-Information Co-evolutionary Framework (STCF) for AI-enabled manufacturing environments. The methodological distinction of STCF lies in coupling high-order manufacturing-state representation, information freshness, and multi-agent decision-making within the same resource-information optimization process, rather than treating them as separate representation, communication, or scheduling functions. A heterogeneous hypergraph represents many-to-many relationships among machines, products, sensors, and operators, while a Multi-Agent Deep Reinforcement Learning (MADRL) mechanism based on Proximal Policy Optimization (PPO) jointly adjusts production-resource scheduling and information-acquisition actions. In this process, changes in resource states affect sensing and information-update decisions, and the resulting information freshness is fed back into subsequent resource decisions. Therefore, the distinction of STCF lies in the coordinated formulation of these components rather than in any individual use of hypergraph modeling or reinforcement learning. Table 1 provides a direct comparison between STCF and representative-related approaches used for manufacturing representation and scheduling.

Table 1. Methodological comparison between STCF and representative intelligent manufacturing approaches

Representative approach	Structural representation	Information freshness	Resource-information coordination	Optimization / resilience
Hypergraph-based manufacturing [3,4]	High-order relations	Not explicit	Not jointly optimized	Mainly representation-oriented
Knowledge-enhanced / IIoT MARL [6–8]	Knowledge- or state-based	Limited	Mainly resource scheduling	Multi-agent optimization
Structured DRL scheduling [9]	Process-structure representation	Not AoI-based	Mainly scheduling-oriented	DRL-based control
STCF (Ours)	Dynamic heterogeneous hypergraph	Explicit through AoI	Bidirectional joint decision	PPO-based MARL with resilience evaluation

As summarized in Table 1, previous methods provide important capabilities in structural representation, distributed scheduling, and reinforcement-learning-based control. The distinction of STCF is that resource-state evolution and information freshness are explicitly linked in both directions: resource changes influence information-acquisition actions, while the updated information state subsequently affects resource-allocation decisions. This coupling is further integrated with high-order heterogeneous manufacturing representation and resilience-aware evaluation within the same decision framework.

The specific contributions of this study are as follows. (1) Coupled high-order representation: a dynamic heterogeneous hypergraph is constructed to represent machine-product-sensor-operator interactions and integrate high-order manufacturing dependencies into the state representation used for resource-information decision-making. This representation is used specifically to support time-varying resource-information coupling within the subsequent decision process. (2) Co-evolutionary decision mechanism: a PPO-based multi-agent optimization mechanism jointly controls resource scheduling and information acquisition, allowing resource-state changes to influence sensing decisions while information freshness feeds back into subsequent resource allocation. This establishes a bidirectional resource-information decision process instead of optimizing the two objectives independently. Its methodological contribution therefore lies in integrating AoI-

aware information acquisition with production-resource actions within a shared multi-agent policy. (3) Resilience-aware optimization and evaluation: resilience is incorporated into the optimization and evaluation process to quantify system behavior under equipment failures and operational fluctuations, allowing the coordinated policy to be assessed not only by energy consumption and information freshness but also by its ability to maintain operational stability under disturbances.

Together, these components form a representation-decision-evaluation chain for resource-information co-evolution: heterogeneous hypergraph modeling provides the coupled manufacturing state, multi-agent optimization determines adaptive resource and information actions, and resilience evaluation characterizes system behavior under dynamic disturbances. Accordingly, STCF integrates high-order manufacturing representation, AoI-aware information acquisition, and multi-agent decision-making to support explicit resource-information co-evolution. The proposed framework therefore focuses on coordinated adaptation between physical resource operations and cyber information processes in complex manufacturing systems.

2. Methodology

2.1 Problem Formulation

In the industrial manufacturing environment, there is a close connection among equipment operation, production tasks, and information transmission. Any change in one aspect may affect the entire production process. For instance, changes in equipment status will impact production scheduling, and the timeliness of information acquisition will also influence the system's judgment of the current situation. Therefore, this paper defines the manufacturing process as a dynamic system, in which production resources and information conditions constantly change according to the actual operating environment. The decision-making process needs to combine the current production situation and the available information simultaneously.

To describe this change process, this paper characterizes the system state from two aspects: production operation and information acquisition. The resource flow state $S_R(t)$ is used to represent the changes in resources during the manufacturing process, including factors such as equipment operation status, production progress, energy consumption, etc., that vary with time. The information flow state $S_I(t)$ reflects the current information that the system can obtain, mainly involving sensor data, communication status, and information update speed. By combining the state changes of resources and information, the overall system state can be expressed as:

$$S(t) = \{S_R(t), S_I(t)\} \quad (1)$$

Among them, the correlation between $S_R(t)$ and $S_I(t)$ reflects the mutual influence between the manufacturing state changes and the information update process.

Based on this, this paper further develops a joint decision-making strategy for resource allocation and information acquisition:

$$\pi^* : S(t) \rightarrow A(t) \quad (2)$$

$A(t)$ represents the joint actions that the system can perform at the current moment, including resource adjustment and sensor sampling control, etc. The optimization objective comprehensively considers the system's energy consumption, information transmission delay, and stability performance in dynamic environments.

$$\min_{\pi} J = \lambda_1 E_{\text{total}} + \lambda_2 \Delta - \lambda_3 \Omega \quad (3)$$

E_{total} is used to measure the overall energy consumption, Δ is used to describe the information delay, Ω is used to reflect the system's resilience under dynamic disturbances, and λ_1 , λ_2 , and λ_3 are used to adjust the weights among different objectives.

To explicitly define resilience, let q_{ref} denote the normalized resource-state performance immediately before disturbance onset and $q(t)$ the performance at time t . The instantaneous resilience state is

$$\Omega(t) = 1 - \frac{\max[0, q_{\text{ref}} - q(t)]}{q_{\text{ref}} + \varepsilon}, \quad 0 \leq \Omega(t) \leq 1 \quad (4)$$

where $\varepsilon > 0$ prevents division by zero. A larger $\Omega(t)$ indicates greater retention of pre-disturbance performance. The resilience term Ω in Eq. (3) is obtained by averaging $\Omega(t)$ over the disturbance–recovery interval $[t_d, t_r]$:

$$\Omega = \text{RS} = \frac{1}{N} \sum_{t=t_d}^{t_r} \Omega(t), \quad N = t_r - t_d + 1 \quad (5)$$

where $\text{RS} \in [0,1]$ is the Resilience Score reported in the experiments. A higher value indicates smaller cumulative performance loss and faster recovery, directly linking the optimization objective to the experimental resilience metric.

This optimization process needs to take into account various constraints in actual production, such as production task requirements, equipment operation conditions, and the communication environment. Based on these constraints, the model needs to learn the coordination between resource scheduling and information acquisition so that the system can adjust its operation strategy according to environmental changes and improve manufacturing flexibility.

2.2 System Architecture

The proposed resource and information co-evolution framework is based on IIoT technology and integrates the physical manufacturing process, the network information environment, and the artificial intelligence optimization method to achieve collaborative management of the manufacturing system. As shown in Figure 1, this framework is mainly composed of five interrelated parts, including the physical resource flow layer, the network information flow layer, the resource-information co-evolution interface, the AI-driven collaborative optimization module, and the adaptive manufacturing decision module.

The physical resource flow layer mainly describes the actual operating objects in the manufacturing process, including machines, products, energy consumption, production status, and equipment operating conditions. By monitoring these factors, the system can obtain the time-varying resource state $S_R(t)$ and reflect the current manufacturing environment.

The network information flow layer is responsible for the collection, transmission, and processing of sensor data. Unlike traditional information processing approaches that mainly focus on communication efficiency, this layer dynamically adjusts information acquisition according to changes in resources, thereby enabling the system to obtain information that matches the current production status. The resulting information state $S_I(t)$ incorporates sensor observations, communication conditions, and information freshness characterized by Age of Information (AoI).

The resource-information co-evolution interface connects the production process and the information system, enabling changes at the production site to be reflected in the information state. During production, equipment conditions and task changes affect the information that needs to be collected, while sensor observations and communication conditions influence subsequent production arrangements. This bidirectional interaction provides the coupled state information used to update the dynamic heterogeneous hypergraph described in Section 2.3.

The AI-enabled co-evolution optimization module, combined with the multi-agent decision-making method, optimizes production arrangements and information collection. Using the coupled resource-information state, dynamic hypergraph representation, AoI, and resilience information $\Omega(t)$, the module jointly determines resource-allocation and information-acquisition actions. The system then adjusts equipment and energy use, sensor sampling, and other controllable operations according to the current production condition.

Through this process, STCF can adjust its operation mode according to changes in the production site, coordinating physical-resource decisions with cyber-information updates while maintaining production stability, efficiency, and adaptability under changing operating conditions.

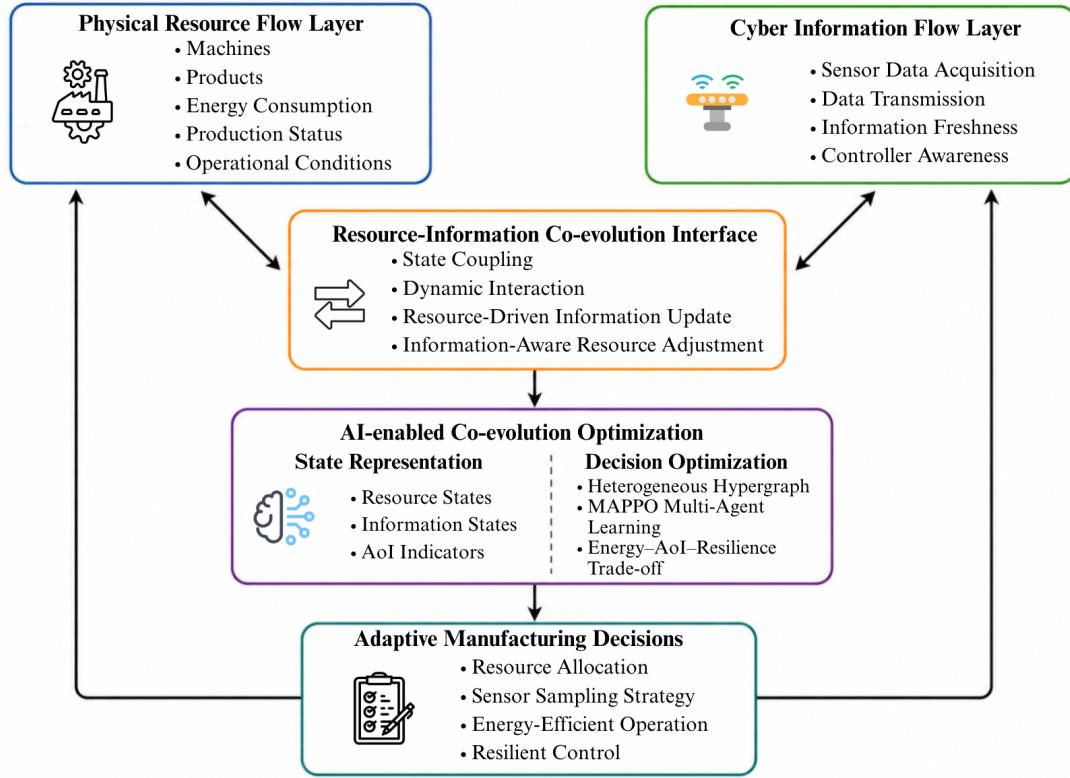


Figure 1. The overall architecture of the resource-information co-evolution framework in the intelligent manufacturing system

2.3 Dynamic Heterogeneous Hypergraph Modeling

In complex manufacturing systems, there are often intricate relationships between different types of objects. For instance, machines, products, operators, and sensors may all be involved in the same production process and interact with each other. Such connections are difficult to accurately describe using traditional graph structures.

In order to better illustrate the connection between manufacturing resource changes and information interaction, this paper constructs a resource-information heterogeneous hypergraph that evolves with the manufacturing state:

$$G(t) = (V(t), \mathcal{E}(t), W(t)) \quad (6)$$

where $V(t)$ contains the active manufacturing entities, $\mathcal{E}(t)$ represents hyperedges connecting entities involved in the same production task or information interaction, and $W(t) = w_e(t)$ contains their time-varying weights.

To capture changes in task assignment, sensing association, and equipment participation, the hyperedge set is updated as

$$\mathcal{E}(t+1) = \{e \subseteq V(t+1) \mid c_e(t+1) \geq \tau_e\} \quad (7)$$

where $c_e(t) \in [0,1]$ is the association strength of hyperedge e and τ_e is an activation threshold. A hyperedge is added or removed when the corresponding task-machine-sensor-operator relationship becomes active or inactive; otherwise, its topology is retained and its weight is updated.

The physical meaning of the hyperedge weight is defined using normalized task dependency $d_e(t)$, resource load $l_e(t)$, and information freshness $f_e(t)$:

$$w_e(t) = \eta_1 d_e(t) + \eta_2 l_e(t) + \eta_3 f_e(t), \quad \eta_i \geq 0, \sum_{i=1}^3 \eta_i = 1 \quad (8)$$

Here, $d_e(t)$ measures dependency among entities participating in the same task, $l_e(t)$ represents resource utilization, and $f_e(t)$ represents sensing-information freshness and decreases with AoI. Thus, $w_e(t)$ reflects the current operational importance of the high-order resource-information relationship. The association measure, activation threshold, and hyperedge-weight coefficients are fixed according to implementation before policy training and are not optimized by PPO.

The time-dependent incidence relation is defined as

$$h_t(v, e) = \begin{cases} 1, & v \in e, \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

The transmission probability matrix is accordingly expressed as

$$P(t) = D_v^{-1}(t)H(t)W(t)D_e^{-1}(t)H^T(t) \quad (10)$$

where $H(t) = [h_t(v, e)]$, and $D_v(t)$ and $D_e(t)$ are the corresponding vertex- and hyperedge-degree matrices.

Through this hypergraph model, the system can simultaneously describe the actual connections and information transmission relationships among different objects in the manufacturing process, while allowing the relationship structure and importance to vary with production tasks, resource loading, and information freshness.

2.4 Dynamic Coupling Mechanism

The operation changes of the system are mainly influenced by the interaction between the resource state $S_R(t)$ and the information state $S_I(t)$. Different from methods that treat communication quality as an independent constraint, this study considers information freshness as a variable coupled with manufacturing conditions.

The resource state $S_R(t)$ describes equipment conditions, production status, energy consumption, and operational variations, while the information state $S_I(t)$ characterizes sensor observations and communication freshness. Age of Information (AoI) is adopted to measure information timeliness at the controller, which has been widely used to characterize information freshness and update efficiency in industrial IoT systems [16].

Let t_k denote the generation time of the k -th update and t'_k its successful reception time. The instantaneous AoI is defined as:

$$\Delta(t) = t - t_k, \quad t'_k \leq t < t'_{k+1} \quad (11)$$

The bidirectional coupling between resource and information states is explicitly represented as

$$S_R(t+1) = F_R(S_R(t), a_R(t), S_I(t)) + \xi_R(t) \quad (12)$$

$$S_I(t+1) = F_I(S_I(t), a_I(t), S_R(t)) + \xi_I(t) \quad (13)$$

where $a_R(t)$ and $a_I(t)$ are resource-scheduling and information-acquisition actions, $F_R(\cdot)$ and $F_I(\cdot)$ are state-transition functions, and $\xi_R(t)$ and $\xi_I(t)$ denote operational and communication disturbances. Thus, resource decisions depend on the available information state, while information acquisition responds to manufacturing conditions.

For a discrete decision interval δ , AoI evolves as

$$\Delta(t+1) = \begin{cases} d(t+1), & u(t+1) = 1, \\ \Delta(t) + \delta, & u(t+1) = 0 \end{cases} \quad (14)$$

where $u(t+1) \in \{0, 1\}$ indicates successful reception of a new update and $d(t+1)$ is its transmission delay. Equations (12)–(14) therefore describe the closed loop in which

resource changes affect information acquisition and updated information subsequently influences resource decisions.

Proposition 1 (Boundedness of the coupled state). Assume that F_R and F_I map the feasible state-action domains to bounded sets under bounded disturbances, the transmission delay satisfies $d(t) \leq d_{\max}$, and at least one successful update occurs within any M consecutive decision intervals. Then the coupled state $S_R(t), S_I(t), \Delta(t)$ remains bound.

Proof. Under the first assumption, $S_R(t)$ and $S_I(t)$ remain finite. From Eq. (14), AoI can increase for at most M consecutive intervals before an update is received; therefore,

$$\Delta(t) \leq M\delta + d_{\max} \quad (15)$$

Hence, all components of the coupled state are bounded. Because Eq. (8) combines normalized bounded variables with non-negative coefficients summing to one, $W(t)$ is also bounded. Therefore, the state representation supplied to the decision policy remains finite under the stated conditions.

This proposition establishes boundedness under the assumptions of STCF rather than asymptotic stability of an unrestricted physical production system.

Resource variations may increase processing and transmission delays, resulting in larger AoI values. Conversely, outdated information can reduce state estimation accuracy and lead to inefficient resource decisions. Therefore, Eqs. (12)–(14) provide the explicit closed-loop coupling relationship between resource evolution and information freshness used for adaptive resource-information co-evolution.

2.5 AI-Enabled Resource-Information Co-evolutionary Optimization

The resource-information synchronization problem is formulated as a Dec-POMDP with a task-specific co-evolutionary decision structure, where multiple agents collaboratively learn policies for coordinating resource allocation and information acquisition [17]. The framework considers energy consumption, information freshness, and system resilience under uncertain conditions.

The state space $\mathcal{SSS}$ incorporates hypergraph embedding, AoI, energy consumption, and manufacturing conditions. Consistent with Sections 2.3 and 2.4, the state at time t is defined as

$$s_t = [z_H(t), S_R(t), S_I(t), \Delta(t), E_{\text{total}}(t), \Omega(t)]. \quad (16)$$

where $z_H(t)$ is the dynamic hypergraph embedding, and the remaining terms denote the resource state, information state, AoI, energy consumption, and resilience state, respectively.

The action space $\mathcal{A}$ includes sensor sampling and production resource scheduling. The joint action is $a_t = a_R(t), a_I(t)$, where the resource agent controls production-resource scheduling and the information agent controls sensor sampling and information acquisition. The reward function is

$$r_t = -\alpha \frac{E_{\text{total}}(t)}{E_{\text{max}}} - \beta \frac{\Delta(t)}{\Delta_{\text{max}}} + \gamma \Omega(t) \quad (17)$$

where $\alpha, \beta, \gamma \in [0,1]$ are weighing coefficients, and E_{max} and Δ_{max} are normalization constants.

The expected cumulative reward is

$$J(\theta) = \mathbb{E} \pi \theta [\sum_{t=0}^T \gamma_{\text{disc}}^t r_t(s_t, a_t)] \quad (18)$$

where γ_{disc} denotes the temporal discount factor and is distinguished from the reward coefficient γ .

The PPO clipped objective is adopted for stable optimization:

$$L_{\text{CLIP}}(\theta) = \mathbb{E}_t [\min(\rho_t(\theta) A_t, \text{clip}(\rho_t(\theta), 1 - \epsilon, 1 + \epsilon) A_t)] \quad (19)$$

where A_t is the advantage function and

$$\rho_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (20)$$

is the PPO probability ratio. The symbol $\rho_t(\theta)$ avoids confusion with the reward r_t in Eq. (17).

Through heterogeneous state representation and multi-agent policy optimization, the proposed framework learns adaptive resource-information co-evolutionary strategies. The dynamic hypergraph provides the shared state representation, while the two agents coordinate resource scheduling and information acquisition.

3. Results and Analysis

3.1 Experimental Setup

The proposed STCF was implemented using Python 3.8 and OpenAI Gym on a GPU-accelerated workstation. The experimental environment used an Intel Core i7-12700K CPU, 32 GB RAM, and an NVIDIA GeForce RTX 3060 GPU. Experiments were conducted using two publicly available industrial datasets: the AI4I 2020 Predictive Maintenance Dataset and NASA C-MAPSS. AI4I 2020 provides industrial equipment variables including temperature, rotational speed, torque, and tool wear for characterizing equipment states and degradation. NASA C-MAPSS contains multivariate turbofan sensor measurements under different operating conditions, providing sequential degradation patterns for dynamic state modeling.

Because neither dataset is a direct flexible-manufacturing scheduling dataset, their variables were mapped to manufacturing-state semantics before being incorporated into the STCF control environment. The mapping is summarized in Table 2. This mapping uses the original equipment measurements as state trajectories rather than treating them as direct scheduling, energy, or communication labels.

Dataset	Original variables	Manufacturing interpretation	STCF role
AI4I 2020	Air/process temperature	Thermal condition	Resource-state feature in $S_R(t)$
AI4I 2020	Rotational speed	Machine operating speed	Resource-state feature in $S_R(t)$
AI4I 2020	Torque	Mechanical workload	Resource-state feature in $S_R(t)$
AI4I 2020	Tool wear	Equipment degradation	Degradation feature in $S_R(t)$
NASA C-MAPSS	Operating-condition settings	Equipment operating regime	Dynamic operating-state feature
NASA C-MAPSS	Sensor measurements	Condition and degradation trajectory	Sequential input for $S_R(t)$

AI4I 2020 and C-MAPSS provide equipment-condition and sensor-state trajectories, which populate the machine/resource and sensing components of the STCF state. Product-, task-, and operator-related entities are represented at the control-environment level. The resulting evaluation therefore examines resource-information coordination driven by real equipment-state trajectories.

Raw measurements were cleaned and normalized using min-max scaling. Samples retained temporal order and were divided into training, validation, and test sets at 70:15:15 to prevent future-state leakage. Sequential states were constructed using a 20-observation sliding window with stride one. All available sensor variables were retained after normalization, while invalid or inconsistent records were removed before window construction. The datasets provided equipment-state trajectories rather than direct energy-consumption or communication-delay labels. Energy consumption was evaluated from resource-allocation behavior, while AoI was updated from information-generation and reception times according to Section 2.4. The same preprocessing and evaluation protocol was applied to both datasets and all methods.

Multi-Agent Proximal Policy Optimization (MAPPO) contained two cooperative agents: a resource-scheduling agent and an information-acquisition agent. The former adjusted production-resource allocation from equipment and energy states, whereas the latter controlled sensor sampling and information updates according to resource conditions and AoI. Both received the dynamic heterogeneous-hypergraph embedding together with normalized resource, information, AoI, energy, and resilience variables. A centralized critic was used during training, while each actor executed from local observations. The actors and critic each contained two fully connected 128-unit layers with ReLU activation.

Table 2. Dataset-to-manufacturing semantic mapping

STCF was compared with three baselines: a Rule-Based (RB) heuristic, Genetic Algorithm (GA) [18], and standard Deep Q-Network (DQN) [19]. All methods used the same preprocessing and experimental environment. MAPPO [20] was trained for 5,000 episodes using Adam, a learning rate of 3×10^{-4} , batch size 256, discount factor 0.99, PPO clipping coefficient 0.20, and generalized advantage estimation coefficient 0.95. Reward coefficients were $\alpha = 0.4$, $\beta = 0.4$, and $\gamma = 0.2$.

For resilience evaluation, disturbances were introduced after an initial stable period. Equipment degradation was represented by a 30–40% reduction in normalized resource state. Operational fluctuations were generated by adding zero-mean perturbations with $SD = 0.05$ to normalized temperature, rotational speed, torque, and available sensor measurements. Disturbance onset, duration, and perturbation sequences were identical across methods within each paired run. After disturbance removal, recovery was evaluated. For NASA C-MAPSS, the original degradation trend was retained, and perturbations were applied to normalized operating-condition variables; for AI4I, disturbances were applied to normalized operating sequences.

The disturbance settings are summarized in Table 3 for reproducibility. The Resilience Score was calculated using Eq. (5), with q_{ref} denoting pre-disturbance normalized resource-state performance and $q(t)$ its value during disturbance and recovery. The same trajectory was used in Section 3.5. The score is bounded to $[0,1]$, with larger values indicating lower cumulative loss and faster recovery. Energy consumption and AoI were reported separately.

Table 3. Perturbation configuration for resilience evaluation

Configuration	Target	Magnitude / distribution	Timing	Cross-method control
Resource degradation	Normalized resource state	30–40% reduction	Onset at approximately 40 s; removed at approximately 55 s for recovery	Identical onset and duration
Operational fluctuation	Normalized operating/sensor variables	Zero mean, $SD = 0.05$	During disturbance interval	Identical perturbation sequence
AI4I 2020	Normalized operating sequences	Same settings	Same disturbance–recovery protocol	Same sequence across

				methods
NASA C-MAPSS	Operating-condition variables	Same perturbation magnitude; original degradation retained	Same protocol	Same sequence across methods

The Resilience Score was calculated using Eq. (5), with q_{ref} denoting pre-disturbance normalized resource-state performance and $q(t)$ its value during disturbance and recovery. The same trajectory was used in Section 3.5. The score is bounded to $[0,1]$, with larger values indicating lower cumulative loss and faster recovery. Energy consumption and AoI were reported separately.

Each method was evaluated by over 10 independent runs. STCF and standard DQN were compared using two-tailed paired t-tests, with runs paired by identical disturbance sequences and random-seed indices. Significance was assessed at $\alpha = 0.05$. In the performance tables, denotes $p < 0.05$, denotes $p < 0.01$, and ns denotes $p \geq 0.05$. The same procedure was applied to energy consumption, Average AoI, and Resilience Score.

3.2 Performance Comparison

The proposed STCF was evaluated on the AI4I 2020 and NASA C-MAPSS datasets using three metrics: Total Energy Consumption (kWh), Average Age of Information (AoI, ms), and Resilience Score. These metrics characterize resource utilization efficiency, information freshness, and operational stability under dynamic equipment conditions.

As shown in Table 4, STCF provides a favorable balance across the three performance dimensions, although it does not achieve the best value for every individual metric on both datasets. On the AI4I 2020 dataset, compared with GA, STCF reduced energy consumption by 14.2% and Average AoI by 21.8%, while achieving the highest resilience score of 0.84. Compared with standard DQN, STCF further reduced energy consumption from 150.8 to 139.1 kWh and Average AoI from 108.4 to 98.2 ms, while increasing the resilience score from 0.71 to 0.84.

On NASA C-MAPSS, STCF achieved the lowest energy consumption (141.8 kWh) and the highest resilience score (0.81), whereas standard DQN achieved a lower Average AoI (102.6 ms) than STCF (105.4 ms). Thus, STCF does not dominate DQN on information freshness in this dataset. Instead, the results indicate a trade-off: STCF accepts a modest increase in AoI while reducing energy consumption and improving resilience. This behavior is consistent with the multi-objective reward design, which jointly considers energy use, information freshness, and resilience rather than optimizing AoI alone.

Overall, the results show that STCF provides a competitive balance between resource allocation, information freshness,

and resilience under changing equipment conditions. Its performance should therefore be interpreted as coordinated multi-objective behavior rather than uniform superiority on every metric. The specific contributions of the heterogeneous hypergraph, co-evolution mechanism, and resilience-related reward are examined separately through the ablation experiments in Section 3.4.

As defined in Section 3.1, statistical significance was assessed using two-tailed paired t-tests over 10 independent runs, with STCF compared directly against standard DQN for

each metric. On AI4I 2020, the differences in energy consumption, Average AoI, and Resilience Score were significant at $p < 0.01$. On NASA C-MAPSS, the improvement in Resilience Score was significant at $p < 0.01$, whereas the differences in energy consumption and Average AoI were not significant at $p \geq 0.05$. Table 4 reports significance markers consistently for all STCF–DQN comparisons.

Table 4. Quantitative Comparison of System Performance on AI4I 2020 and NASA C-MAPSS (Mean $\pm$ SD, $n=10$)

Dataset	Method	Energy Consumption (kWh) ↓	Average AoI (ms) ↓	Resilience Score ↑
AI4I 2020	RB	184.5 $\pm$ 5.2	142.3 $\pm$ 8.1	0.45 $\pm$ 0.04
	GA	162.1 $\pm$ 4.3	125.6 $\pm$ 6.5	0.62 $\pm$ 0.05
	Standard DQN	150.8 $\pm$ 3.9	108.4 $\pm$ 5.2	0.71 $\pm$ 0.03
	STCF (Ours)	139.1 $\pm$ 2.8 **	98.2 $\pm$ 4.1 **	0.84 $\pm$ 0.02 **
NASA C-MAPSS	RB	171.6 $\pm$ 6.1	146.8 $\pm$ 7.9	0.48 $\pm$ 0.05
	GA	154.3 $\pm$ 5.0	128.5 $\pm$ 6.7	0.63 $\pm$ 0.04
	Standard DQN	145.2 $\pm$ 4.2	102.6 $\pm$ 5.1	0.73 $\pm$ 0.03
	STCF (Ours)	141.8 $\pm$ 3.5 ns	105.4 $\pm$ 4.6 ns	0.81 $\pm$ 0.03 **

Note: Statistical markers refer to two-tailed paired t-tests between STCF and standard DQN over 10 independent runs. ** indicates $p < 0.01$; * indicates $p < 0.05$; ns indicates $p \geq 0.05$. Lower values are better for Energy Consumption and Average AoI, whereas higher values are better for Resilience Score.

3.3 Convergence and Stability Analysis

To analyze the optimization efficiency of STCF, cumulative reward curves were monitored over 5,000 training episodes. As shown in Figure 2, STCF achieves faster convergence than both GA and standard DQN. The proposed framework reaches a cumulative reward of approximately 100 earlier than the compared methods and subsequently maintains a higher reward level during the later stages of training.

The convergence behavior is consistent with the design of the heterogeneous hypergraph representation and MAPPO-based multi-agent optimization. The hypergraph provides a structured representation of high-order relationships among heterogeneous manufacturing entities, while PPO constrains

large policy updates during training. These mechanisms provide a plausible explanation for the more stable learning behavior observed for STCF, although Figure 2 evaluates the complete framework rather than isolating the convergence contribution of each component.

The convergence curves were obtained from the same 10 independent runs defined in Section 3.1, and the shaded regions in Figure 2 indicate ± 1 standard deviation around the mean cumulative reward. STCF shows a smaller reward variation across runs, indicating more stable learning behavior under different initialization conditions and dynamic manufacturing states. These results suggest that the proposed co-evolutionary optimization strategy improves both convergence speed and training reliability.

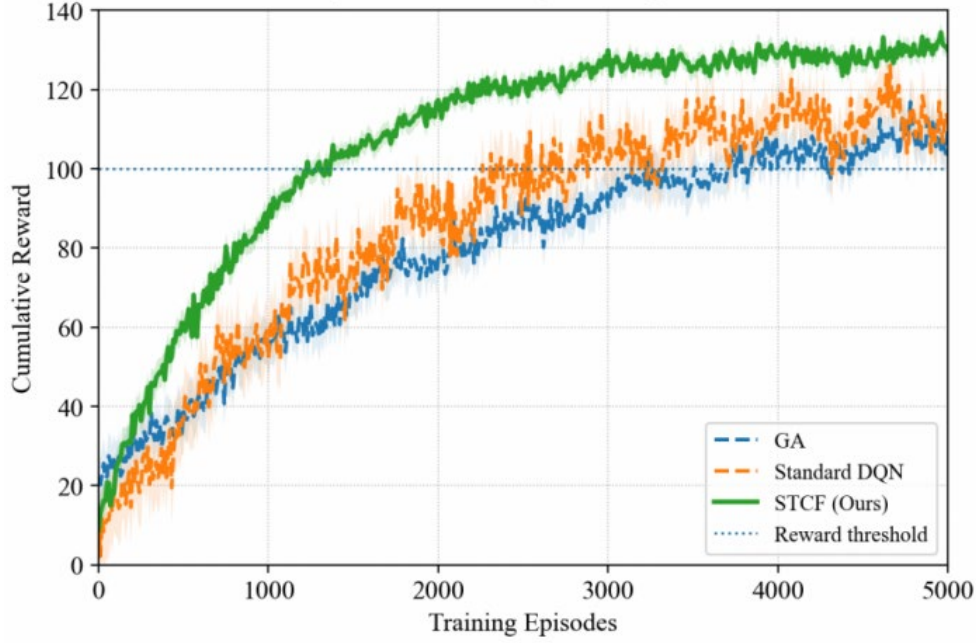


Figure 2. Optimization convergence comparison among GA, DQN, and STCF over 5,000 training episodes; shaded regions indicate ± 1 SD across 10 independent runs

3.4 Ablation Studies

To investigate the contribution of individual components, ablation experiments were conducted on both AI4I 2020 and NASA C-MAPSS datasets. Four configurations were compared: the complete STCF model; w/o Hypergraph, where heterogeneous hypergraph modeling is replaced by a standard graph representation; w/o Co-evolution, where the adaptive interaction between resource states and information updates is removed; and w/o Resilience Reward, where the resilience-related term $\Omega(t)$ is removed from the training reward while the Resilience Score remains an independent evaluation metric.

As shown in Table 5, removing individual components changes the balance among the three objectives and generally degrades at least one key performance dimension. This indicates that the design of each part plays a significant role in the coordinated optimization of resources and information. Without the heterogeneous hypergraph, energy consumption and Average AoI increase, while the Resilience Score decreases. This suggests that the hypergraph representation contributes to modeling the complex interrelationships among machines, products, operators, and sensors.

After removing the co-evolution mechanism, energy consumption and Average AoI increase and the Resilience

Score decreases on both datasets. The deterioration is particularly evident in information freshness and resilience. Since the system cannot adjust the information acquisition method according to resource changes, it becomes more difficult to maintain coordination between resource scheduling and information collection under changing manufacturing conditions.

Removing the resilience-related reward term slightly reduces energy consumption but produces a marked decrease in the independently evaluated Resilience Score. On AI4I 2020, the score decreases from 0.84 to 0.65, while on NASA C-MAPSS it decreases from 0.81 to 0.63. This result indicates a trade-off between short-term resource efficiency and disturbance tolerance: excluding the resilience objective can marginally reduce energy use but weakens the system's ability to maintain performance during equipment degradation and operational fluctuations.

Overall, the experimental results show that the heterogeneous hypergraph modeling, the resource and information collaboration mechanism, and the resilience optimization strategy make complementary contributions to STCF. The hypergraph supports high-order state representation, the co-evolution mechanism coordinates resource and information decisions, and the resilience-related reward improves performance retention under disturbances.

Table 5. Ablation experiment results of STCF on the AI4I 2020 and NASA C-MAPSS datasets (Mean $\pm$ SD, $n=10$)

Dataset	Variant	Energy Consumption (kWh) ↓	Average AoI (ms) ↓	Resilience Score ↑
AI4I 2020	STCF	139.1 $\pm$ 2.8	98.2 $\pm$ 4.1	0.84 $\pm$ 0.02
	w/o Hypergraph	144.7 $\pm$ 3.4	103.6 $\pm$ 4.5	0.77 $\pm$ 0.03

NASA C-MAPSS	w/o Co-evolution	146.2 ± 3.6	107.8 ± 4.8	0.71 ± 0.04
	w/o Resilience Reward	137.8 ± 3.1	99.5 ± 4.3	0.65 ± 0.04
	STCF	141.8 ± 3.5	105.4 ± 4.6	0.81 ± 0.03
	w/o Hypergraph	147.9 ± 3.8	109.7 ± 5.0	0.74 ± 0.04
	w/o Co-evolution	149.1 ± 4.0	113.2 ± 5.2	0.69 ± 0.04
	w/o Resilience Reward	140.6 ± 3.6	103.8 ± 4.7	0.63 ± 0.04

3.5 Mechanism and Resilience Analysis

The intelligent manufacturing system needs to promptly identify problems when the equipment's operating status changes and maintain the stability of the production process. Figure 3 illustrates the state-dependent information-acquisition response of STCF using the same disturbance protocol defined in Section 3.1 and summarized in Table 3.

During the stable operation phase (0–40 s), the production status showed no significant changes, and the system maintained a relatively low information-collection frequency. At approximately 40 s, the predefined equipment-degradation disturbance specified in Table 3 was introduced, reducing the normalized resource-state level by approximately 30–40%. As the resource-flow state decreased from approximately 1.0 to 0.6–0.7, STCF increased the information sampling rate from approximately 1 to nearly three times the normal level. The increase was generated by the learned information-acquisition policy rather than being manually prescribed, reflecting the coupling between resource-state deterioration and information-update decisions.

The disturbance was removed at approximately 55 s, initiating the recovery interval. As the resource state moved back toward its pre-disturbance level, the learned policy progressively reduced the sampling rate toward the initial level. This response limits unnecessary information acquisition after the abnormal condition subsides while preserving higher-frequency monitoring during the period of degraded equipment status.

The disturbance–response–recovery trajectory in Figure 3 follows the same onset, duration, perturbation, and recovery settings reported in Table 3 and is also used to calculate the Resilience Score according to Eq. (5). A smaller degradation and faster restoration therefore produce a higher score, while the score remains distinct from the instantaneous sampling-rate response shown in Figure 3. On AI4I 2020, STCF achieved a Resilience Score of 0.84 compared with 0.71 for DQN, corresponding to an approximately 18.3% relative improvement. These results indicate that the co-evolution mechanism enables information acquisition to adapt to equipment-state changes while supporting recovery under operational disturbances.

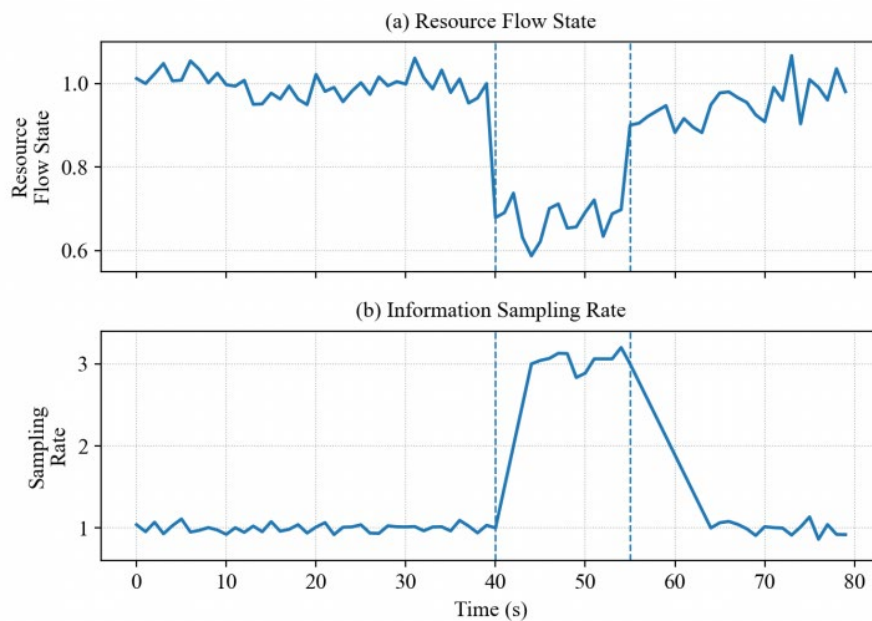


Figure 3. Resource-information co-evolution response under dynamic equipment conditions: (a) normalized resource-flow state and (b) information sampling rate. The disturbance protocol is identical to that summarized in Table 3, with disturbance onset at approximately 40 s and recovery beginning at approximately 55 s

3.6 Generalization and Robustness Analysis

To examine robustness within and beyond the training range, STCF was evaluated under 20%, 35%, and 50% resource degradation, 50-ms additional communication delay, 5% packet loss, and 10% missing sensor observations. The 35% degradation case represents an in-range disturbance condition, whereas the 20% and 50% degradation levels and the communication and sensing perturbations were not included during training. Results on AI4I 2020 are summarized in Table 6.

Table 6. Robustness and generalization under varied disturbance conditions

Condition	Energy consumption	Average AoI	Resilience Score
Reference condition	139.1 $\pm$ 2.8	98.2 $\pm$ 4.1	0.84 $\pm$ 0.02
20% degradation	140.6 $\pm$ 4.0	100.7 $\pm$ 4.5	0.82 $\pm$ 0.03
35% degradation	143.2 $\pm$ 4.4	104.9 $\pm$ 5.0	0.78 $\pm$ 0.04
50% degradation	148.5 $\pm$ 5.1	113.6 $\pm$ 5.8	0.69 $\pm$ 0.05
+50 ms communication delay	141.8 $\pm$ 4.2	107.3 $\pm$ 5.2	0.79 $\pm$ 0.04
5% packet loss	142.7 $\pm$ 4.6	110.1 $\pm$ 5.5	0.76 $\pm$ 0.04
10% missing sensor observations	145.1 $\pm$ 4.8	106.4 $\pm$ 5.3	0.74 $\pm$ 0.05

Performance declined progressively with disturbance severity, without abrupt failure under mild or moderate conditions. The Resilience Score decreased from 0.84 to 0.69 under 50% degradation, while communication delay, packet loss, and missing sensor observations mainly increased AoI and reduced recovery performance. These results indicate partial robustness to unseen disturbances, although out-of-range degradation and communication or sensing perturbations still cause noticeable performance loss.

3.7 Practical Deployment Analysis

The current experiments were conducted in a controlled evaluation environment and have not yet included closed-loop production-line validation. The environment assumes synchronized measurements, fixed decision intervals, bounded communication delay, reliable control execution, and normalized sensor inputs. Energy consumption is estimated from resource-allocation behavior rather than measured by industrial power meters. Device-specific protocols, PLC constraints, safety interlocks, and long-term sensor drift are not explicitly modeled.

A practical deployment can proceed in stages: offline training using historical production data, hardware-in-the-

loop validation with representative sensors and PLCs, and edge-gateway deployment in shadow mode without direct control. After validating latency, communication reliability, and decision consistency, limited closed-loop control can be introduced for non-critical resource and sampling decisions under predefined safety constraints and fallback rules. The control scope can then be expanded gradually. Quantization or knowledge distillation may reduce edge inference cost, while periodic offline retraining can adapt the policy to equipment and workload changes.

4. Conclusion

This study evaluated the effectiveness of resource-information co-evolution optimization in complex manufacturing systems under dynamic equipment conditions. Experimental results on two publicly available industrial datasets demonstrate that the proposed STCF achieves a balanced trade-off among resource efficiency, information freshness, and operational resilience compared with conventional optimization and reinforcement learning baselines. The results on the AI4I 2020 dataset show that STCF reduces total energy consumption by 14.2% and decreases the Age of Information by 21.8% compared with the genetic algorithm baseline, while achieving the highest resilience score among all compared methods. On NASA C-MAPSS, STCF achieved lower energy consumption and a higher Resilience Score than standard DQN, whereas DQN achieved a slightly lower Average AoI (102.6 ms versus 105.4 ms). Thus, the NASA results support the multi-objective value of STCF in balancing energy use, information freshness, and resilience rather than indicating uniform superiority on every individual metric. STCF achieved a Resilience Score of 0.81 under equipment degradation conditions, supporting its ability to maintain operational performance during changes in equipment status. By combining the results of the ablation experiments, it can be observed that the system performance declines after removing the heterogeneous dependency modeling, suggesting that the correlation information among different manufactured objects plays a crucial role in accurately describing the system state. Additionally, removing the resilience-related optimization term reduces the Resilience Score, showing its contribution to maintaining system performance under disturbances.

From the perspective of the training process, STCF reached a higher reward level more rapidly than the compared optimization methods while maintaining smaller fluctuations. Further analysis of the information-sampling process showed that the system adjusted sampling frequency according to resource-state changes, acquiring more information during abnormal conditions and reducing unnecessary collection during stable operation. The additional robustness analysis showed that performance declined as resource degradation increased and remained sensitive to unseen communication and sensing perturbations, indicating useful but bounded adaptability beyond the training conditions.

Several limitations remain. First, AI4I 2020 and NASA C-MAPSS primarily provide equipment-state and degradation trajectories rather than complete production-line scheduling, communication, and metered energy records; therefore, part of the resource-information interaction is constructed within the control environment. Second, although the robustness tests cover unseen degradation, communication delay, packet loss, and missing observations, generalization is still limited by the tested disturbance range and operating conditions. Third, the current evaluation does not include closed-loop deployment on a real production line, where asynchronous sensing, PLC constraints, network uncertainty, safety interlocks, and edge-computing limitations may affect policy execution. MADRL-related computational cost also remains relevant for resource-constrained industrial devices.

Future work will therefore evaluate STCF using real production logs containing richer scheduling, communication, and energy measurements, followed by hardware-in-the-loop validation with representative sensors, PLCs, and industrial networks. Subsequent studies will examine edge-gateway deployment in shadow mode before introducing limited closed-loop control under explicit safety and fallback constraints. Model quantization and knowledge distillation will be investigated to reduce inference cost, while broader cross-condition validation will be used to assess generalization across equipment types and production workloads.

References

- [1] Hu Y, Jia Q, Yao Y, Lee Y, Lee M, Wang C, et al. Industrial Internet of Things intelligence empowering smart manufacturing: a literature review. *IEEE Internet Things J.* 2024;11(11):19143-67. doi:10.1109/JIOT.2024.3367692.
- [2] Chi HR, Wu CK, Huang NF, Tsang KF, Radwan A. A survey of network automation for industrial Internet-of-Things toward Industry 5.0. *IEEE Trans Ind Inform.* 2023;19(2):2065-77. doi:10.1109/TII.2022.3215231.
- [3] Wang R, Cheung CF, Wang C. Heterogeneous hypergraph learning for analyzing surface defects in additive manufacturing process. *J Manuf Syst.* 2024;76:1-10. doi:10.1016/j.jmsy.2024.06.011.
- [4] Nagy L, Ruppert T, Löcklin A, Abonyi J. Hypergraph-based analysis and design of intelligent collaborative manufacturing space. *J Manuf Syst.* 2022;65:88-103. doi:10.1016/j.jmsy.2022.08.001.
- [5] Mo F, Chaplin JC, Sanderson D, Martínez-Arellano G, Ratchev S. Semantic models and knowledge graphs as manufacturing system reconfiguration enablers. *Robot Comput Integr Manuf.* 2024;86:102625. doi:10.1016/j.rcim.2023.102625.
- [6] Qin Z, Lu Y. Knowledge graph-enhanced multi-agent reinforcement learning for adaptive scheduling in smart manufacturing. *J Intell Manuf.* 2025;36(8):5943-66. doi:10.1007/s10845-024-02494-0.
- [7] Heik D, Bahrpeyma F, Reichelt D. Study on the application of single-agent and multi-agent reinforcement learning to dynamic scheduling in manufacturing environments with growing complexity: case study on the synthesis of an industrial IoT test bed. *J Manuf Syst.* 2024;77:525-57. doi:10.1016/j.jmsy.2024.09.019.
- [8] Li C, Chang Q, Fan HT. Multi-agent reinforcement learning for integrated manufacturing system-process control. *J Manuf Syst.* 2024;76:585-98. doi:10.1016/j.jmsy.2024.08.021.
- [9] Luo J, Yi S, Lin Z, Zhang H, Zhou J. Petri-net-based deep reinforcement learning for real-time scheduling of automated manufacturing systems. *J Manuf Syst.* 2024;74:995-1008. doi:10.1016/j.jmsy.2024.05.006.
- [10] Patel D, Sahu CK, Rai R. Security in modern manufacturing systems: integrating blockchain in artificial intelligence-assisted manufacturing. *Int J Prod Res.* 2024;62(3):1041-71. doi:10.1080/00207543.2023.2262050.
- [11] Peruzzini M, Prati E, Pellicciari M. A framework to design smart manufacturing systems for Industry 5.0 based on the human-automation symbiosis. *Int J Comput Integr Manuf.* 2024;37(10-11):1426-43. doi:10.1080/0951192X.2023.2257634.
- [12] Liu S, Zheng P, Bao J. Digital twin-based manufacturing system: a survey based on a novel reference model. *J Intell Manuf.* 2024;35(6):2517-46. doi:10.1007/s10845-023-02172-7.
- [13] Ji T, Xu X. Exploring the integration of cloud manufacturing and cyber-physical systems in the era of Industry 4.0 – an OPC UA approach. *Robot Comput Integr Manuf.* 2025;93:102927. doi:10.1016/j.rcim.2024.102927.
- [14] Tan J, Shi J, Wu L, Chen B, Tang H, Zhang C, et al. Embodied intelligence empowering customized manufacturing: architecture, opportunities, and challenges. *IEEE Access.* 2025;13:92740-55. doi:10.1109/ACCESS.2025.3572778.
- [15] Bozzi A, Graffione S, Jimenez JF, Sacile R, Zero E. A platoon-based approach for AGV scheduling and trajectory planning in fully automated production systems. *IEEE Trans Ind Inform.* 2025;21(1):594-603. doi:10.1109/TII.2024.3455413.
- [16] Liang J, Chan TT, Pan H. Minimizing age of collection for multiple access in wireless industrial Internet of Things. *IEEE Internet Things J.* 2024;11(2):2753-66. doi:10.1109/JIOT.2023.3293142.
- [17] Li Y, Liu Q, Li X, Gao L. Manufacturing resource-based self-organizing scheduling using multi-agent system and deep reinforcement learning. *J Manuf Syst.* 2025;79:179-98. doi:10.1016/j.jmsy.2025.01.004.
- [18] Lin L, Hao XC, Gen M, Jo JB. Network modeling and evolutionary optimization for scheduling in manufacturing. *J Intell Manuf.* 2012;23(6):2237-53. doi:10.1007/s10845-011-0569-6.
- [19] Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, et al. Human-level control through deep reinforcement learning. *Nature.* 2015;518(7540):529-33. doi:10.1038/nature14236.
- [20] Yu C, Velu A, Vinitky E, Gao J, Wang Y, Bayen A, et al. The surprising effectiveness of PPO in cooperative multi-agent games. In: Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A, editors. *Advances in neural information processing systems*. Vol. 35. Red Hook (NY): Curran Associates, Inc.; 2022. p. 24611-24. doi:10.52202/068431-1787.