SCIENCE CHINA
Information Sciences

• Supplementary File •

LaTac: latent tactics for robust multi-agent coordination under intermittent communication

Enguang Yao[1], Qidong Liu[1,2,3*], Jiajia Hou[1], Mingfei Sun[4], Zongzhang Zhang[5] & Mingliang Xu[1,2,3*]

[1]*School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou 450001, China*
[2]*National Supercomputing Center In Zhengzhou, Zhengzhou 450001, China*
[3]*Engineering Research Center of Intelligent Swarm Systems, Ministry of Education, Zhengzhou 450001, China*
[4]*Department of Computer Science, The University of Manchester, United Kingdom*
[5]*National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China*

Appendix A Method

Appendix A.1 Preliminaries

Given a multi-agent coordination task with intermittent communication, formally defined as a tuple: $G = \langle N, S, A, P, R, \Omega, O, \gamma, \omega \rangle$, where N is the finite set of n agents; $s \in S$ is the global state of the environment; and $\gamma \in [0, 1)$ is the discount factor; at each timestep t, each agent $i \in \{1, 2, \cdots, n\}$ obtains a partial observation $o_i \in O$ generated according to the observation function $\Omega(s, i)$, and separately receives additional information o_{-i} from its teammates via communication. We define a binary communication mask $\boldsymbol{\mathcal{I}}_i$ for each agent i, where $\boldsymbol{\mathcal{I}}_i(j) = 1$ if agent i is able to receive information from agent j at the current timestep; otherwise, $\boldsymbol{\mathcal{I}}_i(j) = 0$. To simulate the stochastic nature of intermittent communication, we set a parameter $\omega \in [0, 1]$ where $\forall j \in \{1, 2, \cdots, n\} - \{i\}$, $\boldsymbol{\mathcal{I}}_i(j)$ is independently set to 0 with information dropout probability ω. Based on the above information, each agent selects an action $a_i \in A$ according to a policy π, i.e., $a_i \sim \pi(a_i|o_i, o_{-i})$, resulting in a joint action $\mathbf{a} = [a_1, \cdots, a_n]^\top \in A^n$. The state then transitions to the next state s' according to the transition function $P(s'|s, \mathbf{a})$. Finally, all agents receive a shared reward $r = R(s, \mathbf{a})$ from the environment.

To address the above problems, we propose an implicit coordination framework based on the latent tactics, where the latent tactics dynamically assign roles to each agent based on the current situation. These roles guide agents toward robust decision-making through role-conditioned policy adaptation. Formally, we propose the following definition of the latent tactics and the roles.

Definition 1. (Latent Tactics and Roles) *Latent Tactics refer to the emergent role-based coordination patterns in multi-agent systems. Rather than being manually specified, they are implicitly learned through end-to-end role assignment. To formalize this concept, we define a latent tactic (LT) as a joint configuration of roles across agents, denoted by* $LT = [\rho_0, \rho_1, \rho_2, \cdots, \rho_n]$*, where each agent* i *selects a role* ρ_i *represented by a trainable embedding vector* $\mathbf{z}_i$ *drawn from a shared role embedding matrix* $\mathbf{Z}$*. Notably,* ρ_0 *denotes the default role, which serves as a non-specialized assignment when no explicit role selection is applied.*

Appendix A.2 Coordinative decision-making model and latent tactics training

In this phase, we propose a coordination model tailored for stable communication conditions, where inter-agent information is reliably available. The model consists of three key modules: an *Aggregator* for decentralized information aggregation, a *Commander* for learning latent tactics and assigning roles accordingly, and an *Actor* for producing action values conditioned on the agent's partial observation and its assigned role. We now proceed to detail these three core components.

Aggregator. Building on attention-based architectures developed for stable communication settings [1], we employ multi-head attention (MHA) [2] to aggregate teammate observations in a decentralized manner under stable communication, enabling agents to focus on the most relevant inputs. For each agent i, its own observation o_i is linearly projected into a query vector via a learnable matrix $\mathbf{W}_q^{i\ell}$ in each attention head $\ell \in \{1, 2, \cdots, L\}$. The information received from other agents, denoted as o_{-i}, is defined as $o_{-i} = \{o_j \mid j \neq i \wedge \boldsymbol{\mathcal{I}}_i(j) = 1\}$. Here, the communication mask $\boldsymbol{\mathcal{I}}_i$ is set to an all-one vector, indicating full information availability under stable communication scenarios. Then, $\forall o_j \in o_{-i}$, the input o_j is projected into a key vector via the matrix $\mathbf{W}_k^{i\ell}$, and the attention weight between the query o_i and the key o_j is computed as:

$$\mu_{ij}^{\ell} \propto \frac{(\mathbf{W}_q^{i\ell} o_i)^\top \cdot (\mathbf{W}_k^{i\ell} o_j)}{\sqrt{d_K}}, \tag{A1}$$

where d_K denotes the dimensionality of the key vectors and serves as a scaling factor in the dot-product attention mechanism to prevent the softmax outputs from becoming excessively peaked, which may lead to vanishing gradients during training.

* Corresponding author (email: ieqdliu@zzu.edu.cn, iexumingliang@zzu.edu.cn)

For each attention head ℓ, the agent-specific aggregated information is computed as a weighted sum over the value projections of the received information:

$$\mathbf{e}_i^\ell = \sum_{o_j \in o_{-i}} \mu_{ij}^\ell \cdot (\mathbf{W}_v^{i\ell} o_j), \tag{A2}$$

where $\mathbf{W}_v^{i\ell}$ denotes the learnable projection matrix that transforms o_j into its corresponding value representation in attention head ℓ. Then, the outputs from all L attention heads are concatenated and passed through a shared output projection matrix W_o to obtain the final aggregated feature vector for agent i:

$$\mathbf{e}_i = \text{Concat}(\mathbf{e}_i^1, \mathbf{e}_i^2, \cdots, \mathbf{e}_i^L)\mathbf{W}_o^i. \tag{A3}$$

To mitigate overfitting to idealized communication settings, we perturb the aggregated feature vector $\mathbf{e}_i$ for each agent i using the reparameterization trick from variational autoencoders (VAEs) [3], injecting structured Gaussian noise during training. This promotes the learning of more generalizable representations and provides a more realistic foundation for subsequent tactic-level role inference, facilitating transfer to intermittent communication conditions introduced later. Formally, we have:

$$\hat{\mathbf{e}}_i = \mu(\mathbf{e}_i) + \sigma(\mathbf{e}_i) \odot \eta, \tag{A4}$$

where $\mu(\mathbf{e}_i)$ and $\sigma(\mathbf{e}_i)$ are the mean and standard deviation vectors derived from $\mathbf{e}_i$, $\eta \sim \mathcal{N}(0, 1)$ is a noise vector sampled from a standard normal distribution, and $\odot$ denotes element-wise multiplication.

Commander. Through the aforementioned efficient information aggregation, the *Aggregator* provides agents with a more comprehensive and dynamic perception of the environment. Inspired by team-based gaming paradigms where human players collectively refine their understanding through routine training, culminating in the emergence of sophisticated latent tactics. These tactics are subsequently employed as strategic directives within cooperative scenarios. Consequently, we propose a *Commander* that leverages the aggregated information from the *Aggregator* to learn latent tactics. This module assigns appropriate roles to individual agents based on these learned tactics, thereby facilitating coordinated multi-agent cooperation. To extract temporally enriched latent representations for role inference, we define a role-inference encoder $f_\rho(\cdot, \cdot)$, composed of an MLP followed by a gated recurrent unit (GRU) [4]. Given the aggregated information $\hat{\mathbf{e}}_i$, the updated latent state is computed as:

$$\widetilde{\mathbf{h}}_i^\rho = f_\rho(\hat{\mathbf{e}}_i, \mathbf{h}_i^\rho), \tag{A5}$$

where $\mathbf{h}_i^\rho$ and $\widetilde{\mathbf{h}}_i^\rho$ are the previous and the current hidden states of the GRU, respectively. In this way, we can fully leverage temporal information and integrate historical data, thereby significantly enhancing the accuracy and relevance of the derived tactics. Accordingly, the probability distribution over roles for agent i is computed based on $\widetilde{h}_i^\rho$ as follows:

$$\hat{\mathcal{P}}_i = \text{softmax}(\text{MLP}(\widetilde{h}_i^\rho \mathbf{W}_\rho^i)), \tag{A6}$$

where W^ρ is the trainable parameter. Subsequently, the $\epsilon-$greedy method is employed to select roles for each agent during the training stage, thereby balancing the exploration-exploitation trade-off, i.e.,

$$\rho_i = \begin{cases} \arg\max_{\rho \in \Psi} \hat{\mathcal{P}}_i(\rho) & \text{with probability } 1-\epsilon, \\ \rho \sim \text{Random}(\Psi) & \text{with probability } \epsilon, \end{cases} \tag{A7}$$

where Ψ is the set of roles. In contrast, during the testing phase, it is generally expected that agents minimize or entirely eliminate exploration. Typically, we set ϵ to 0, which indicates the agents consistently select the role with the highest estimated probability according to the learned policy.

Once a discrete role ρ_i is selected by agent i as part of the latent tactics, its corresponding semantic representation $\mathbf{z}_i$ is retrieved from the trainable role embedding matrix $\mathbf{Z}$, by indexing the row associated with ρ_i. This continuous embedding serves as a role-specific feature descriptor, enabling each agent to ground its coordination strategy in a shared latent space.

Actor. The *Actor* module is responsible for generating the action a_i for each agent i, based on its partial observation and the role representation assigned by the *Commander*. To estimate Q-values based on both local observations and assigned roles, we define a parametric function $f_a(\cdot, \cdot)$ implemented as an MLP followed by a GRU. This function processes the concatenated input $[\mathbf{e}_i, \mathbf{z}_i]$ and returns a hidden state used for computing the Q-value vector over the action space. The resulting hidden state is then projected to the action space via a linear transformation. Formally, we have:

$$\widetilde{\mathbf{h}}_i^a = f_a([\mathbf{e}_i, \mathbf{z}_i], \mathbf{h}_i^a), \tag{A8}$$

$$Q_i = \mathbf{W}_A^i \widetilde{h}_i^a + \mathbf{b}_A^i, \tag{A9}$$

where $[\mathbf{e}_i, \mathbf{z}_i]$ is an operation for concatenating $\mathbf{e}_i$ and $\mathbf{z}_i$, and $\mathbf{h}_i^a$ and $\widetilde{\mathbf{h}}_i^a$ are the previous and the current hidden states of the GRU in *Actor*, respectively. $\mathbf{W}_A^i$ and $\mathbf{b}_A^i$ are learnable parameters that project the hidden representation to action values. In accordance with the aforementioned Eq. (A7), we also employ the ϵ-greedy method for selecting an action a_i, with the associated Q-value represented by Q_i.

Training. We enhance multi-agent coordination by extending the value decomposition framework of QMIX [5], which estimates the global action-value Q_{tot} via a monotonic mixing network over individual agent utilities Q_i, conditioned on

the global state s. In our approach, latent role representations are introduced into this framework to enable more expressive credit assignment and refined cooperative behavior.

Specifically, we incorporate an attention mechanism to draw contextual dependencies between the global state and a set of learned role representations. We first process the sequence of historical states $(s^0, s^1, \cdots, s^t)$ with a GRU to produce a state embedding $\mathbf{s}$, which encapsulates temporal information. Formally, following the standard attention formulation, we map a query and a set of key-value pairs to a weighted output. We designate the state embedding $\mathbf{s} \in \mathbb{R}^{d_s \times d}$ as the query, and the role representations $\mathbf{z} = [\mathbf{z}_1, \mathbf{z}_2, \cdots, \mathbf{z}_n]^\top \in \mathbb{R}^{n \times d}$ as both the key and value, where d is the dimension of role representation and d_s is the length of state embedding. The attention output $\boldsymbol{\xi}$ is computed as:

$$\boldsymbol{\xi}^{\ell'} = \sum_{i=1}^{n} \text{softmax}\left(\frac{\mathbf{s}\mathbf{W}_q^{i\ell'} \cdot [\mathbf{z}_i \mathbf{W}_k^{i\ell'}]^\top}{\sqrt{d_{K'}}}\right)(\mathbf{z}_i \mathbf{W}_v^{i\ell'}), \tag{A10}$$

$$\boldsymbol{\xi} = \text{Concat}(\boldsymbol{\xi}^1, \boldsymbol{\xi}^2, \cdots, \boldsymbol{\xi}^{L'})\mathbf{W}_o^{i'}, \tag{A11}$$

where $\boldsymbol{\xi}^{\ell'}(\ell' \in \{1, 2, \cdots, L'\})$ represents the output of each attention head, and $\mathbf{W}_q^{\ell'}, \mathbf{W}_k^{\ell'}, \mathbf{W}_v^{\ell'}$, and $\mathbf{W}_o'$ are trainable parameters. And $\sqrt{d_{K'}}$ serves to normalize the dot-product in attention computation, mitigating the risk of vanishing gradients and improving training stability. The final attention output $\boldsymbol{\xi}$ is combined with the global state s to dynamically condition the mixing network parameters through a hypernetwork [6], denoted as β. This framework is first utilized to train the *Actor* by minimizing the temporal difference loss [7], which can be described as:

$$\mathcal{L}_{\text{rl}}(\zeta, \beta) = \mathbb{E}_{(s,\mathbf{a},r,s')\sim\mathcal{D}}[(r + \gamma \max_{\mathbf{a}'} \bar{Q}_{\text{tot}}(s', \mathbf{a}' \mid \zeta, \beta) - Q_{\text{tot}}(s, \mathbf{a} \mid \zeta, \beta))^2], \tag{A12}$$

where ζ is the parameter of the *Actor* and $\bar{Q}_{\text{tot}}$ is the target network, s' and $\mathbf{a}'$ are the state and the joint action of the next timestep. And the expectation is estimated with uniform samples from the same replay buffer $\mathcal{D}$.

Notably, although the *Actor* consumes role representations as input, these are selected discretely by the *Commander*. Once a role ρ is assigned, the corresponding representation from $\mathbf{Z}$ is retrieved and passed to the *Actor*. Since this selection process is discrete and non-differentiable, gradient flow from the policy loss $\mathcal{L}_{\text{rl}}$ cannot be backpropagated to the upstream *Commander* and *Aggregator* module. To overcome the gradient blockage introduced by the non-differentiable role selection, we draw inspiration from adversarial training paradigms [8] and adopt an alternating optimization strategy. In this scheme, the *Commander* and *Aggregator* are updated while keeping the *Actor* fixed, thereby allowing the upstream modules to adapt their outputs in response to a stable policy. The corresponding objective is formulated as:

$$\mathcal{L}_{\text{rl}}(\theta, \phi, \beta) = \mathbb{E}_{(s,\mathbf{a},r,s')\sim\mathcal{D}}[(r + \gamma \max_{\mathbf{a}'} \bar{Q}_{\text{tot}}(s', \mathbf{a}' \mid \theta, \phi, \beta) - Q_{\text{tot}}(s, \mathbf{a} \mid \theta, \phi, \beta))^2], \tag{A13}$$

where θ and ϕ are the parameters of the *Aggregator* and the *Commander*, respectively.

Additionally, to preserve the semantic expressiveness of role representations, we introduce an auxiliary regularization objective. While roles derived from latent tactics are designed to be distinct, a learnable embedding matrix $\mathbf{Z}$ is susceptible to representation collapse, a phenomenon where embeddings converge to a non-differentiable solution. To counteract this, we propose a diversity loss that encourages the set of role embeddings to form a quasi-orthogonal basis [9]. This is achieved by penalizing the deviation of the embeddings' pairwise cosine similarity matrix from the identity matrix. Formally, this regularization loss $\mathcal{L}_{\text{orth}}$ is defined as:

$$\mathcal{L}_{\text{orth}} = \left\| \hat{\mathbf{Z}}\hat{\mathbf{Z}}^\top - \mathbf{I} \right\|_F^2, \tag{A14}$$

where $\hat{\mathbf{Z}}$ is the matrix of L2-normalized role embeddings derived from the learnable matrix $\mathbf{Z}$, $\mathbf{I}$ is the identity matrix matching the dimensionality of $\hat{\mathbf{Z}}\hat{\mathbf{Z}}^\top$, and $|\cdot|_F^2$ denotes the squared Frobenius norm.

Thus, the teacher model is optimized with a joint objective that combines the reinforcement learning loss and the orthogonality regularization:

$$\mathcal{L}_{\text{tot}} = \mathcal{L}_{\text{rl}} + \alpha \mathcal{L}_{\text{orth}}, \tag{A15}$$

where α is the adjustable hyperparameter for the regularization loss function.

Appendix A.3 Knowledge distillation for intermittent communication environments

To enable robust coordination under intermittent communication—where inter-agent information is partial and sporadically available—we adopt a representation-level distillation strategy that introduces a representation-level distillation framework that transfers knowledge from the stable communication setting. Although both training (stable communication) and deployment (intermittent communication) settings share a modular architecture consisting of an *Aggregator*, a *Commander*, and an *Actor*, the shift in communication reliability necessitates adjustments in how these modules process and interpret information. These adaptations preserve structural alignment while enabling the model to internalize robust coordination strategies through targeted distillation.

Firstly, under intermittent communication, each agent i receives information from its teammates based on a dynamically sampled binary communication mask $\boldsymbol{\mathcal{I}}_i$. At each timestep, for each $j \neq i$, the entry $\boldsymbol{\mathcal{I}}_i(j)$ is independently set to 0 with dropout probability $\omega > 0\%$, simulating information unavailable due to unreliable communication channels. This results in a time-varying, sparse information set o_{-i}, in contrast to the complete inputs used during training under stable

communication. In addition, in contrast to the stable communication phase, where controlled perturbations are introduced to improve robustness, no additional noise is injected under intermittent communication. This is because information dropout and incomplete reception already introduce sufficient uncertainty into the input space. In such a scenario, artificially adding further noise would not meaningfully enhance robustness and may even obscure the true coordination signals. The *Commander* thus performs role inference based directly on the naturally degraded and dynamically varying observations.

Training. To mitigate the performance degradation caused by degraded inputs in intermittent communication, we introduce a representation-level distillation objective to transfer robust internal representations from the model trained under stable communication. Specifically, we enforce consistency by minimizing the divergence between the intermediate outputs of the *Aggregator* and *Commander* modules across both settings. This encourages the intermittent communication model to emulate the latent feature representations and role-selection distributions learned by its stable-communication counterpart. Formally, the distillation objective consists of two components:

$$\mathcal{L}_{\mathrm{e}} = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{e}_i - \hat{\mathbf{e}}_i\|_2^2 \,, \tag{A16}$$

$$\mathcal{L}_{\mathrm{r}} = \frac{1}{n}\sum_{i=1}^{n} D_{KL}(\mathcal{P}_i \parallel \hat{\mathcal{P}}_i), \tag{A17}$$

where $\mathbf{e}_i$ and $\hat{\mathbf{e}}_i$ represent the aggregated information produced by the *Aggregator* module under stable and intermittent communication, respectively. Likewise, $\mathcal{P}_i$ and $\hat{\mathcal{P}}_i$ denote the role-selection probability distributions output by the corresponding *Commander* module.

To form the overall distillation objective, the two components—representation-level aggregation loss $\mathcal{L}_{\mathrm{e}}$ and role inference loss $\mathcal{L}_{\mathrm{r}}$—are jointly optimized. We combine them into a unified loss function with tunable weights to balance their relative contributions:

$$\mathcal{L}_{\mathrm{distill}} = \mathcal{L}_{\mathrm{e}} + \lambda_r \mathcal{L}_{\mathrm{r}}, \tag{A18}$$

where λ_r is the hyperparameter that controls the emphasis on feature-level and role-level alignments. This formulation mirrors the structure of soft-target distillation in supervised settings, but focuses exclusively on latent representation transfer in the absence of ground-truth labels. Moreover, $\mathcal{L}_{\mathrm{distill}}$ jointly guides the model deployed under intermittent communication to approximate the aggregated information and tactic-level role semantics learned under stable communication setting.

Appendix B Experiments

In this section, we conduct a series of experiments to evaluate the effectiveness of our proposed framework LaTac under intermittent communication conditions, focusing on three key questions. i) Can LaTac enhance model robustness in complex intermittent communication conditions (Sec. Appendix B.1)? ii) Can we learn meaningful latent tactics, and what are the respective contributions of different modules to the performance improvements (Sec. Appendix B.2)? iii) Is the LaTac model sensitive to hyperparameters, and in scenarios with varying degrees of information loss, how does LaTac select roles to consistently demonstrate robust and superior performance (Sec. Appendix B.3)?

We employ the StarCraft II micromanagement (SMAC) benchmark [10] and Multi-agent Particle Environment (MPE) [11] as our testbeds. SMAC provides a complex real-time strategy environment, and it comprises a series of StarCraft II micro-scenarios designed to assess the capacity of independent agents to coordinate and solve intricate tasks. We test LaTac on a range of maps including several hard and super hard scenarios, which demand fine-grained coordination and highlight the benefits of our coordination framework. To ensure fair evaluation, all experiments are conducted in StarCraft II version 2.4.10 at difficulty level 7. In MPE, we evaluate on Cooperative Navigation, as well as two predator-prey tasks (3v1 and 6v2), where agents must coordinate under partial observability to achieve spatial goals or capture fast-moving adversaries controlled by hand-coded heuristics.

To ensure fair robustness, each experiment was repeated with five random seeds, and we report the mean performance with 95% confidence intervals. The code can be available at https://github.com/yaoenguang/LaTac.

Appendix B.1 Comparative experiments

We evaluate the performance of LaTac against several state-of-the-art baselines: 1) FoX [12] which mitigates the curse of dimensionality in exploration to enhance multi-agent coordination; 2) RODE [13] which clusters actions based on their effects to form role-based subtasks, thereby boosting learning efficiency; 3) DCC [14] which employs a bi-level structure and mutual information constraints to facilitate both intra- and inter-subtask coordination; 4) MAIC [15], which facilitates multi-agent communication by generating targeted incentive messages that directly influence teammates' value functions, thereby enabling explicit coordination beyond conventional observation exchange. 5) NDQ [16], which minimizes inter-agent communication by jointly optimizing mutual information and entropy regularizers, enabling efficient coordination via sparse message exchange; 6) ROMA [17] which is a role-oriented MARL framework where agent roles emerge through learning; 7) QMIX [5] which decomposes the global action-value function; 8) VDN [18] which sums individual agent Q-values to estimate the joint action-value function.

FoX, RODE, MAIC, NDQ, DCC, and ROMA were implemented by extending the open-source PyMARL framework, while the official pymarl2 [19] project was used for QMIX and VDN. Since these methods were not originally designed for

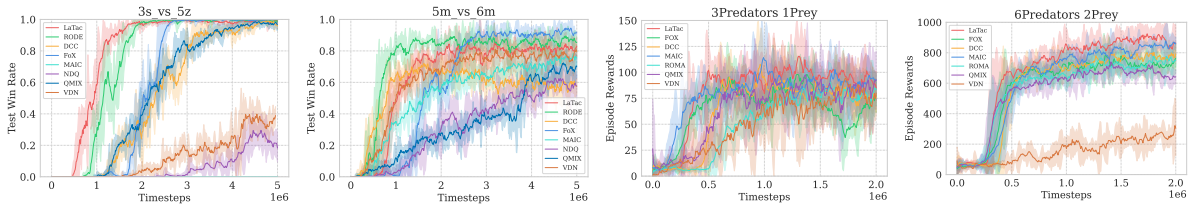

Figure B1 Performance comparison between **LaTac** and baselines on the SMAC and MPE benchmarks.

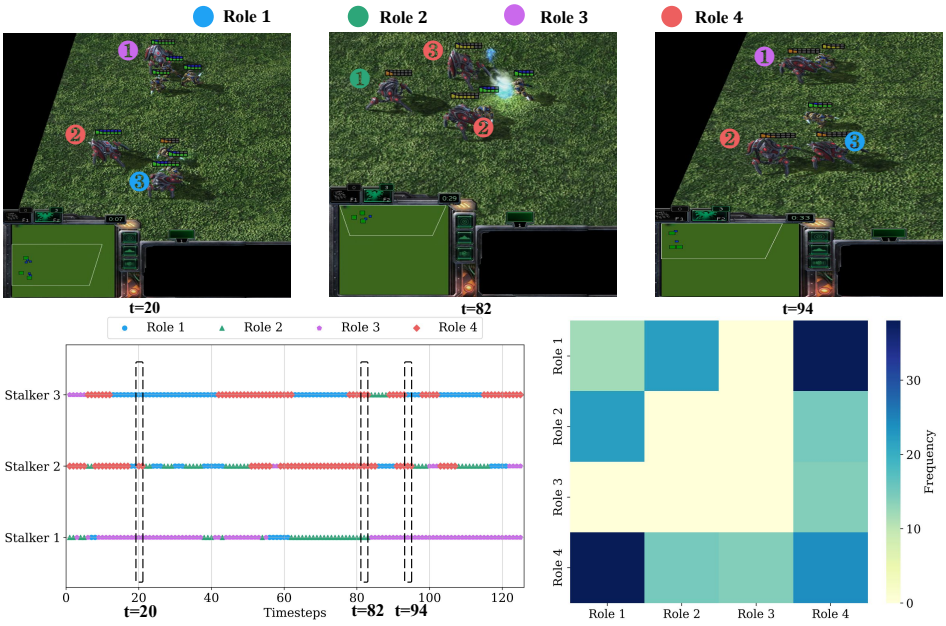

Figure B2 In an episode of the *3s_vs_5z* scenario, the tactic assignments are visualized through rendering scenes at three timesteps: $t = 20$, $t = 82$, and $t = 94$. In these snapshots, each allied agent is identified by a colored number, where the digit denotes the agent's ID and the color indicates its assigned role, as detailed in the chart at the lower left. This chart illustrates the dynamic role assignments during the game, while the heatmap at the lower right shows the frequency of various role choices made by the agents throughout the game.

intermittent communication, we integrated the *Aggregator* from LaTac, allowing agents to go beyond mere local observations and exchange information with teammates. This modification enables more robust coordination in the presence of communication disruptions.

To simulate intermittent communication, we set the information dropout probability ω to 40%, causing 40% of communications to be randomly dropped. For LaTac, we set the number of roles to 16, with the dimension of the role representation $\mathbf{Z}$ set matching that of the observation space.

The left two subfigures of Fig. B1 present a comparison of LaTac against various baseline algorithms across SMAC maps. The results demonstrate that LaTac performs excellently on most maps. Notably, on the super hard *6h_vs_8z* map, LaTac and RODE significantly outperform other algorithms, with LaTac exhibiting the best performance. This highlights the importance of role representations in multi-agent systems for addressing challenging maps and demonstrates LaTac's robustness in intermittent communication scenarios. However, on the *5m_vs_6m* map, LaTac does not achieve optimal performance, potentially due to the smaller map size, which reduces the influence of communication. In summary, these results provide compelling evidence that our framework effectively enhances multi-agent coordination, particularly in environment with intermittent communication.

Furthermore, we evaluate LaTac in the MPE benchmark to assess its robustness in coordinating agents under intermittent communication. As shown in the right two subfigures of Fig. B1, LaTac consistently outperforms all baselines on Cooperative Navigation and both predator-prey tasks, achieving higher final rewards. MAIC, a state-of-the-art communication framework tailored for fully connected settings, performs competitively when communication is stable but suffers notable degradation under intermittent conditions—highlighting an exposure bias where models trained with full information fail to maintain coordination when message availability becomes sparse. In contrast, structured baselines such as ROMA, DCC, and FoX incorporate inductive priors like role assignment, subtask abstraction, and formation control, which improve coordination but still rely on relatively reliable communication. Traditional value decomposition approaches like VDN and QMIX exhibit the weakest performance, due to their limited capacity to capture agent-specific diversity and communication dynamics. Notably, LaTac demonstrates a distinct advantage in handling intermittent communication, as its tactic-level abstraction enables agents to reason over latent coordination behaviors and make informed decisions even when teammate

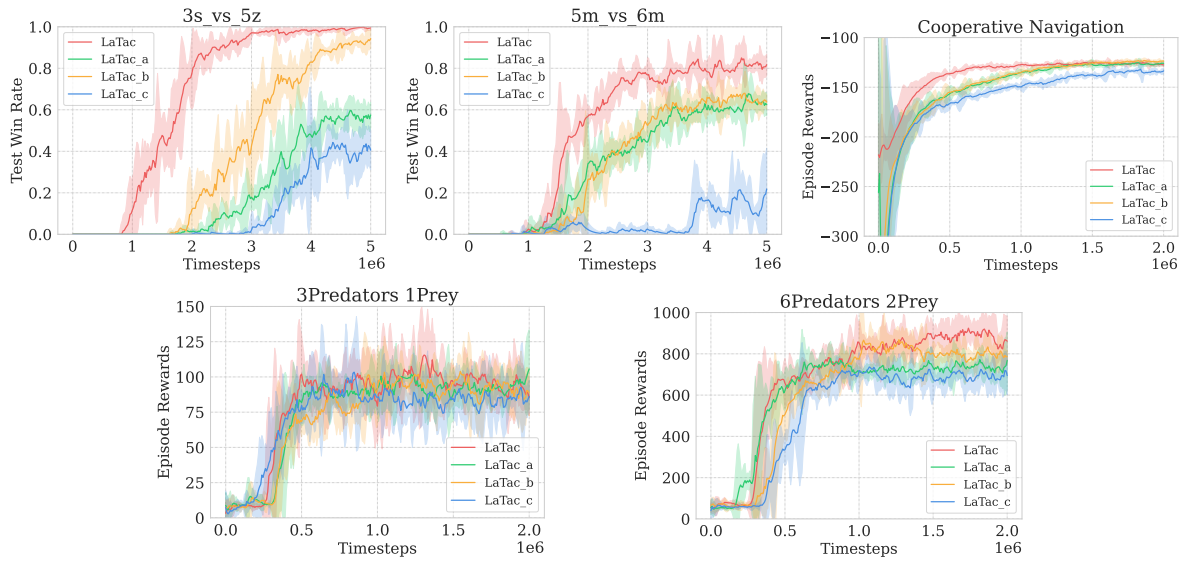

Figure B3 Ablation study on the core components of LaTac. LaTac_b removes the paradigm trained under stable communication, disabling representation-level distillation. LaTac_a excludes the *Aggregator* module, while LaTac_c removes the *Commander* module. Each variant isolates the contribution of a specific component to the overall coordination performance.

information is incomplete, thereby enhancing robustness under intermittent communication.

Appendix B.2 Ablation study

To answer the second question, we analyzed the impact of latent tactics on coordination and performance among agents by visualizing an example from the *3s_vs_5z* map in the SMAC benchmark. We configured the system to incorporate four distinct roles. The experimental results presented in Fig. B2 offer several interesting insights: Firstly, Stalkers demonstrate adaptive behavior by adjusting their roles to minimize losses. For instance, at timestamp $t = 82$, Stalker 1, with lower health, retreats to the rear for ranged attacks, while Stalkers 2 and 3, with higher health, advance to the frontline to act as shields, effectively blocking enemy advances. Secondly, latent tactics emerge within local Stalker populations, as evidenced by the heatmap in the bottom-right corner of Fig. B2, where role 1 and role 4 exhibit the highest frequency of occurrence. By examining game screenshots at $t = 20$ and $t = 94$, we observe that two allied Stalkers form an latent tactic, coordinating a pincer attack on enemy Zealots. Thirdly, Stalker role-switching exhibits a degree of stability, as illustrated in the same heatmap. To ensure stable coordination among multiple agents, each agent tends to maintain a specific role for an extended period rather than switching frequently, aligning with the conclusions drawn in the RODE [13] paper. These findings contribute to a deeper understanding of role emergence and coordination patterns in MARL systems, paving the way for future research to enhance teamwork and adaptability in complex multi-agent environment.

Furthermore, we performed an ablation analysis within the SMAC and MPE benchmarks by evaluating three variants of LaTac. As shown in Fig. B3, all three ablated variants—LaTac_a, LaTac_b, and LaTac_c exhibit varying degrees of performance degradation compared to the full LaTac framework. We first examine LaTac_b, which disables the teacher model that trained under stable communication conditions and thus removes the knowledge distillation mechanism. This leads to a substantial drop in performance, underscoring the pivotal role of teacher-guided supervision in transferring stable coordination patterns. In the absence of such guidance, the student model struggles to reconstruct coherent tactics under intermittent communication, thereby validating the necessity of the framework design based on knowledge distillation. In addition, LaTac_a shows a decline due to the removal of the *Aggregator* module, which reduces the agents' ability to gather comprehensive environmental information. However, LaTac_c suffers the most significant performance drop because it eliminates the *Commander* module, demonstrating the importance of latent tactics in enabling effective multi-agent coordination. As previously analyzed, the relatively small map size may limit the overall impact of communication, which explains why the removal of latent tactics has a more pronounced effect on coordination compared to the *Aggregator* module.

Appendix B.3 Parameter sensitivity analysis

First, we investigated the robustness of our model, specifically whether it can achieve robust decisions under varying intermittent communication conditions. We set $\omega = \{40\%, 80\%, 100\%\}$, where ω indicates that a random ω-percent of teammate information will be dropped during communication. The experimental results are shown in Fig. B4. Initially, we observe that a lower ω value leads to better performance, highlighting the crucial role of communication in multi-agent cooperative decision-making. However, as timestep t progresses, we find that our algorithm performs well even when some or all communication is dropped, matching or even surpassing the performance of the 40% communication scenario. This demonstrates the robustness of our algorithm in the face of intermittent communication.

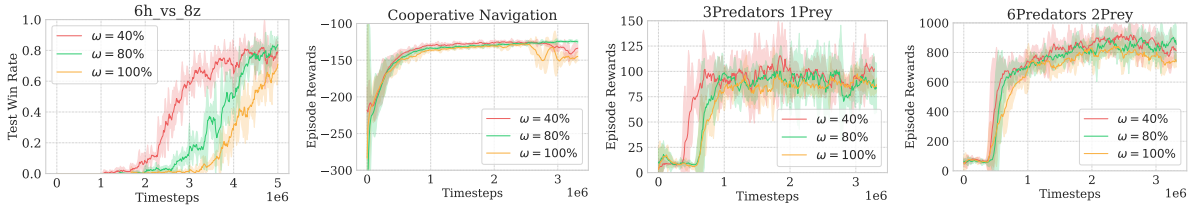

Figure B4 Ablation study evaluating LaTac performance under varying levels of information loss conducted on the SMAC and MPE benchmarks.

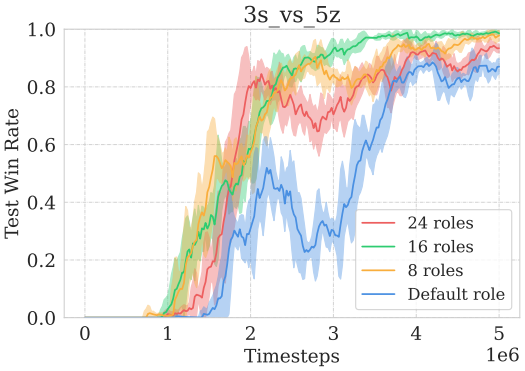

Figure B5 Performance comparison of LaTac under varying intermittent communication conditions on the *3s_vs_5z* map, from a shared default role to increasingly diverse specialized roles.

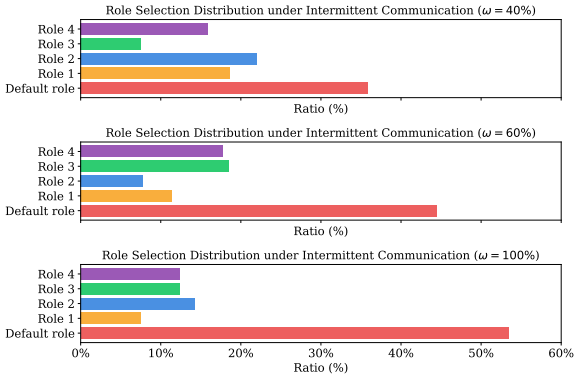

Figure B6 Ratio of default role selection under varying intermittent communication conditions on the *3s_vs_5z* map.

Second, we investigated the impact of different role counts on of our model. To this end, we compare four model variants with no explicit roles (i.e., all agents share the default role ρ_0), and others with 8, 16, and 24 roles. As shown in Fig. B5, the variant with 16 roles yields the highest performance. This suggests that moderate role granularity offers sufficient behavioral diversity to support specialization, while avoiding the redundancy and interference that can arise from overly fine-grained role decompositions. In contrast, limited role richness restricts the model's capacity to differentiate agent behavior, whereas excessive roles introduce unnecessary complexity, potentially hindering coordination stability.

Furthermore, we investigated the probability of agents selecting the default role under varying levels of intermittent communication during testing. As shown in Fig. B6, the results reveal a clear trend: when faced with significant information loss, agents are more likely to adopt the default role, relying primarily on their own observations for decision-making process due to uncertainty about their teammates' statuses. These findings highlight the complementary roles of role diversity and the default role: while diverse roles enable agents to achieve more flexible and adaptive coordination, the default role provides a consistent strategy that helps agents navigate scenarios with constrained communication.

Moreover, to evaluate the stability of our method, we performed a sensitivity analysis for the principal hyperparameters, namely the loss weight α (Eq. A15) and the regularization coefficient λ_r (Eq. A18), within the 3s_vs_5z environment. As illustrated in Figure B7, the results demonstrate the model's robustness. The model exhibits low sensitivity to the choice of α, maintaining near-optimal performance for values across a wide spectrum from 1 to 100. While the model shows greater sensitivity to the regularization coefficient λ_r, it consistently achieves high win rates for values within the range [0.1, 100]; only a significantly small value ($\lambda_r = 0.01$) results in degraded convergence. This empirical investigation confirms that the effectiveness of our proposed method is not critically dependent on fine-grained hyperparameter tuning, thereby underscoring its general robustness.

Appendix C Related work

Appendix C.1 Multi-agent communication

In the domain of multi-agent systems (MAS), effective communication is essential for coordination. Early studies [20, 21] employ a broadcasting approach, allowing all agents to receive partial observations or messages from every other agent within the environment. However, these methods incur significant communication bandwidth and costs, potentially leading to disturbances caused by communication noise. Some studies [1, 15, 22–25] have shifted focus toward selective peer-to-peer communication by developing mechanisms that determine what information is communicated, when it is shared, and with whom. Another important research direction emphasizes communication robustness, aiming to ensure coordination under unreliable or degraded channels. Approaches in this vein include multi-view consistency validation to assess message credibility [26], and adversarial training schemes that expose agents to adaptive perturbations in communication during learning, thereby enhancing policy resilience [27].

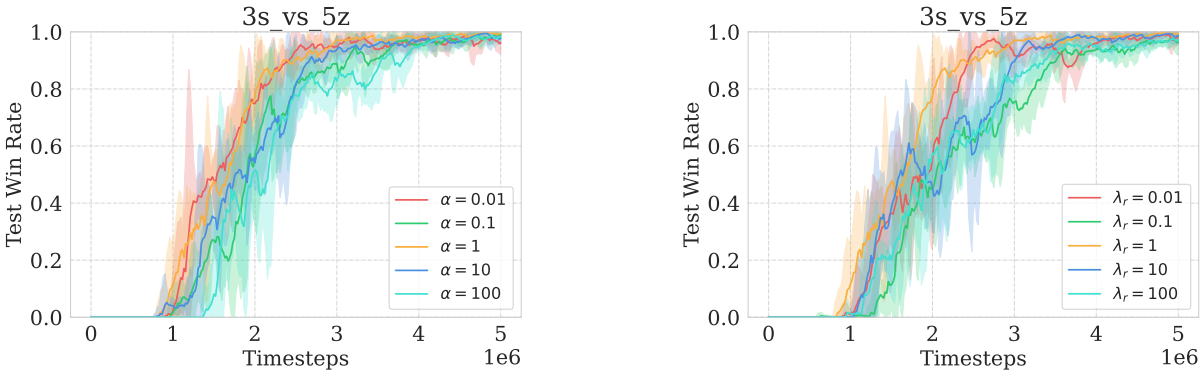

Figure B7 Sensitivity analysis of LaTac to the hyperparameters α and λ_r on the 3s_vs_5z map in SMAC. The test win rate remains stable across wide parameter ranges, indicating that LaTac is robust to both the role-orthogonality and teacher–student alignment coefficients.

However, real-world environments often involve aperiodically intermittent communication, where message availability is unpredictable and information flow may be disrupted [28]. This scenario resembles a persistent "fog of war" [29,30], where agents must act with fragmented situational awareness. Recent efforts, such as AICoRL [31], attempt to mitigate this by reconstructing missing global features via variational inference, but the reliance on approximated centralized representations may introduce misalignment with local contexts. Additionally, Deng et al. [32] explore distributed optimization for nonlinear MAS under intermittent communication, though their solution is tailored to control-theoretic consensus tracking rather than adaptive coordination in dynamic, decentralized environment. While these approaches provide useful insights for handling communication unreliability, they either rely on centralized reconstruction, impose restrictive control assumptions, or neglect the inherent variability in agent roles and coordination patterns. In contrast, our work is inspired by real-world scenarios where agents must adaptively form tactical coordination under unstable information flow. This calls for a framework that not only tolerates partial observability, but also preserves the flexibility to model latent, role-dependent interaction structures under decentralized conditions.

Appendix C.2 Structure-aware multi-agent coordination

Structure-aware coordination has emerged as a promising direction in multi-agent reinforcement learning, particularly through the use of role-based abstractions and spatially informed policy structures [17, 33]. These approaches aim to decompose complex cooperative tasks into specialized behaviors across agents, enabling more efficient coordination. For instance, RODE [13] reduces the joint action space by constraining each role to a predefined action subset, while SIRD [34] formulates role discovery as a hierarchical clustering problem in the action space. ACORM [35] enhances coordination by maximizing mutual information between agents and their roles, thereby encouraging behavioral diversity and facilitating skill transfer. FoX [12] extends this paradigm by leveraging formation-based coordination patterns, encoding spatial relations and dependencies into policy learning to promote structured cooperation without relying on explicit role assignments. Unlike these prior approaches that typically rely on either static clustering or explicit positional priors, our method introduces latent tactics that dynamically encode coordination intent, enabling adaptive role emergence under intermittent communication and task variability.

Parallel to structural modeling, value-based multi-agent reinforcement learning (MARL) provides a scalable framework for decentralized decision-making under partial observability. VDN [18] and QMIX [5] represent two widely adopted paradigms in this space. While VDN aggregates agent Q-values linearly, QMIX employs a mixing network to capture nonlinear inter-agent dependencies, both operating under the centralized training with decentralized execution (CTDE) paradigm. Building upon these foundations, our framework integrates role representations directly into Q-value computation through an attention-based aggregation mechanism, thereby unifying structure-aware coordination with value-based decision making and improving performance in complex cooperative tasks.

References

1 Guan C, Chen F, Yuan L, et al. Efficient multi-agent communication via self-supervised information aggregation. In: Advances in Neural Information Processing Systems, 2022. 1020–1033
2 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Advances in Neural Information Processing Systems, 2017. 5998–6008
3 Doersch C. Tutorial on variational autoencoders. arXiv:1606.05908, 2016.
4 Cho K, van Merrienboer B, Gulcehre C, et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. In: Conference on Empirical Methods in Natural Language Processing, 2014. 1724–1734
5 Rashid T, Samvelyan M, Schroeder De Witt C, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning. Journal of Machine Learning Research, 2020. 21(178):1–51
6 Ha D, Dai A, Le Q V. Hypernetworks. arXiv:1609.09106, 2016.
7 Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. Nature, 2015. 518(7540):529–533

8 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In: Advances in Neural Information Processing Systems, 2014. 2672–2680
9 Li S, Jia K, Wen Y W, et al. Orthogonal deep neural networks. IEEE transactions on pattern analysis and machine intelligence, 2019. 43(4):1352–1368
10 Samvelyan M, Rashid T, de Witt C S, et al. The starcraft multi-agent challenge. In: International Conference on Autonomous Agents and MultiAgent Systems, 2019. 2186–2188
11 Mordatch I and Abbeel P. Emergence of grounded compositional language in multi-agent populations. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2018. 1495-1502
12 Jo Y, Lee S, Yeom J, et al. Fox: Formation-aware exploration in multi-agent reinforcement learning. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2024. 12985–12994
13 Wang T H, Gupta T, Mahajan A, et al. Rode: Learning roles to decompose multi-agent tasks. In: International Conference on Learning Representations, 2021.
14 Li C, Dong S K, Yang S D, et al. Coordinating multi-agent reinforcement learning via dual collaborative constraints. Neural Networks, 2025. 182:106858
15 Yuan L, Wang J H, Zhang F X, et al. Multi-agent incentive communication via decentralized teammate modeling. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2022. 9466–9474
16 Wang T H, Wang J H, Zheng C Y, et al. Learning nearly decomposable value functions via communication minimization. In: International Conference on Learning Representations, 2020.
17 Wang T H, Dong H, Lesser V, et al. Roma: multi-agent reinforcement learning with emergent roles. In: International Conference on Machine Learning, 2020. 9876–9886
18 Sunehag P, Lever G, Gruslys A, et al. Value-decomposition networks for cooperative multi-agent learning based on team reward. In: International Conference on Autonomous Agents and MultiAgent Systems, 2018. 2085–2087
19 Hu J, Wang S Y, Jiang S Y, et al. Rethinking the implementation tricks and monotonicity constraint in cooperative multi-agent reinforcement learning. In: International Conference on Learning Representations, 2023.
20 Sukhbaatar S, Szlam A, Fergus R. Learning multiagent communication with backpropagation. In: Advances in Neural Information Processing Systems, 2016. 2244–2252
21 Foerster J, Assael Y M, de Freitas N, et al. Learning to communicate with deep multi-agent reinforcement learning. In: Advances in Neural Information Processing Systems, 2016. 2137–2145
22 Zhao C Z, Ze Y J, Dong J, et al. Dpmac: differentially private communication for cooperative multi-agent reinforcement learning. In: International Joint Conference on Artificial Intelligence, 2023. 4638–4646
23 Guo X D, Shi D M, Fan W H. Scalable communication for multi-agent reinforcement learning via transformerbased email mechanism. In: International Joint Conference on Artificial Intelligence, 2023. 126–134
24 Meng X R, Tan Y. Pmac: Personalized multi-agent communication. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2024. 17505–17513
25 Liu Z Y, Wan L P, Sui X, et al. Deep hierarchical communication graph in multi-agent reinforcement learning. In: International Joint Conference on Artificial Intelligence, 2023. 208–216
26 Yuan L, Jiang T, Li L H, et al. Robust cooperative multi-agent reinforcement learning via multi-view message certification. Science China Information Sciences, 2024. 67(4):142102
27 Yuan L, Chen F, Zhang Z Z, et al. Communication-robust multi-agent learning by adaptable auxiliary multi-agent adversary generation. Frontiers of Computer Science, 2024. 18(6):186331
28 Ge X H, Han Q L, Zhang X M, et al. Distributed coordination control of multi-agent systems under intermittent sampling and communication: A comprehensive survey. Science China Information Sciences, 2025. 68(5):151201
29 Tryhorn D, Dill R, D Hodson D, et al. Modeling fog of war effects in afsim. The Journal of Defense Modeling and Simulation, 2023. 20(2):131–146
30 Clausewitz C. On war. Penguin UK, 2003.
31 Fu L Y, Wang J, Luo H C. Multi-agent reinforcement learning for cooperative search under aperiodically intermittent communication. Expert Systems with Applications, 2025. 280:127526
32 Deng C, Xu L, Yang T, et al. Distributed cooperative optimization for nonlinear heterogeneous mass under intermittent communication. IEEE Transactions on Automatic Control, 2023. 69(4):2737–2744
33 Yuan H Q, Lu Z Q. Robust task representations for offline meta-reinforcement learning via contrastive learning. In: International Conference on Machine Learning, 2022. 25747–25759
34 Zeng X H, Peng H, Li A S. Effective and stable role-based multi-agent collaboration by structural information principles. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2023. 11772–11780
35 Hu Z C, Zhang Z Z, Li H X, Chen C L, et al. Attention-guided contrastive role representations for multi-agent reinforcement learning. In: International Conference on Learning Representations, 2024.