DIRECT: Decentralized Intention and Latent Rule Emergence for Multi-Agent Cooperation Under Intermittent Communication

Enguang Yao[1] , **Qidong Liu**[1,2*] , **Jialu Sun**[1] , **Kaibo Huang**[1,2] and **Mingliang Xu**[1,2*]
[1]School of Computer Science and Artificial Intelligence, Zhengzhou University, Zhengzhou, China
[2]Artificial Intelligence and Robotics Institute, Zhengzhou University, Zhengzhou, China
egyao@gs.zzu.edu.cn, ieqdliu@zzu.edu.cn, jialusun@gs.zzu.edu.cn, hkaibo@zzu.edu.cn, iexumingliang@zzu.edu.cn

Abstract

Most of the existing decentralized multi-agent systems facilitate coordination by achieving consensus. However, reaching global consensus among a large number of agents in complex settings is challenging because resolving a local disagreement often triggers cascading adjustments across the system. This problem is further exacerbated under intermittent communication. In the field of transportation, human drivers usually rely on *intentions* (e.g., turn signals) in conjunction with traffic *rules* (such as "yielding to through traffic") to achieve efficient cooperation, instead of establishing consensus with all surrounding vehicles. Inspired by this observation, we propose the $Intention \otimes Latent\ Rule$ paradigm to facilitate coordination beyond consensus, which is instantiated by **DIRECT** (**D**ecentralized **I**ntention and Latent **R**ule **E**mergence **C**ooperaT**i**on). Within this framework, agents forecast their intentions based on current actions, and the latent rule inference module derives latent-rule representations from available information to regulate competing intentions, enabling cooperative action selection even when communication is disrupted. Extensive experiments on challenging StarCraft II micromanagement and Multi-Agent Particle Environments demonstrate that DIRECT consistently outperforms state-of-the-art methods, achieving more robust coordination across a range of communication-degraded scenarios. Our code is available at https://github.com/AAI-ZZU/DIRECT-master.

1 Introduction

The Centralized Training with Decentralized Execution (CTDE) is a widely adopted paradigm for cooperative multi-agent reinforcement learning. Despite its progress, CTDE relies on purely local observations during execution, which introduces uncertainty about the state and teammates' information. This uncertainty frequently results in ineffective coordination and suboptimal joint strategies. Recently,

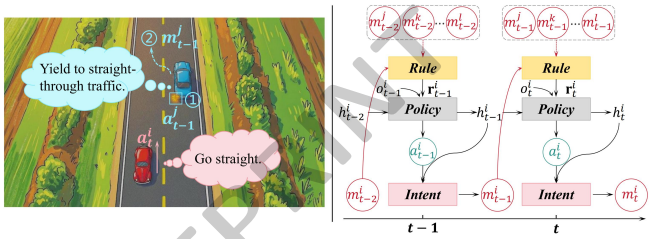

Figure 1: *Left:* A traffic scenario illustrating that decentralized coordination requires the interplay of communicated intentions (turn signals) and traffic rules (yielding). *Right:* A simplified overview of the proposed framework, comprising two key components: the *Intention Forecaster Module* (Intent) and the *Latent Rule Inference Module* (Rule).

more and more approaches increasingly incorporate explicit inter-agent communication into their algorithms to enhance coordination [Du *et al.*, 2025; Meng and Tan, 2024; Du *et al.*, 2024]. We divide these methods into two categories: i) state-aware coordination, where agents exchange their local observations or embeddings [Hu *et al.*, 2024; Guo *et al.*, 2023], and ii) intention-aware coordination, where agents communicate their long-term objectives [Xu *et al.*, 2023b; Song *et al.*, 2025] or sub-goals [Zhang *et al.*, 2025; Dann *et al.*, 2023] to teammates to enable strategic alignment. In this work, we focus on the latter strategy, as the former is often constrained by high communication costs, limiting its effectiveness in real-world applications with stringent bandwidth and real-time requirements [Roth *et al.*, 2006]. Moreover, excessive communication can introduce irrelevant or even harmful information [Guan *et al.*, 2022], potentially undermining the convergence of the learning process.

Although intention-aware approaches have achieved strong performance in several cooperative settings, their applicability becomes limited in *denied environments*. First, existing approaches assume reliable communication, whereas in *denied environments* characterized by the "fog of war" [Clausewitz, 2003], agents must coordinate under intermittent information [Yao *et al.*, 2025; Liu *et al.*, 2025]. Second, existing approaches generally encourage agents to reach global consensus to improve coordination [Li *et al.*, 2025; Xu *et al.*, 2023a]. However, achieving such agreement among all agents is often difficult. Resolving a local disagreement be-

*Corresponding authors

tween two agents may trigger cascading conflicts with the actions of others. To address these issues, we propose a novel $Intention \otimes Latent\ Rule$ paradigm. Our inspiration is drawn from human cooperative practices in traffic systems, as illustrated in the left panel of Fig. 1, where the agent efficiently resolves the local conflicts by coupling an intention (turn signal) with a shared traffic rule (yielding to through traffic), under which the turning vehicle yields unilaterally.

As shown in the right panel of Fig. 1, our method is driven by two core components. The *Intention Forecaster Module* maps the agent's local action history into a predictive intention signal, thereby establishing a communication grounding for future behavior [Lo *et al.*, 2024; Lin *et al.*, 2021; Udagawa and Aizawa, 2019]. Moreover, the *Latent Rule Inference Module* treats the communication link as a time-varying topology. By employing a gating mechanism to selectively process fragmentary signals from currently available neighbors, this module facilitates the emergence of the latent rules. Our main contributions are summarized as follows:

- We propose a *beyond-consensus* paradigm for decentralized multi-agent coordination, formalized as $Intention \otimes Latent\ Rule$. This paradigm leverages latent rules to mitigate conflicts among shared intentions, thereby bypassing the need for global consensus.
- We develop **DIRECT** (**D**ecentralized **I**ntention and Latent **R**ule-**E**mergent **C**ooperaT**i**on), a framework that employs an action-grounded *Intention Forecaster Module* to project the agent's plan into a shared semantic space, ensuring that communicated intentions remain behaviorally grounded. These intentions are processed by an attention-based *Latent Rule Inference Module*, which jointly considers their temporal freshness and semantic relevance, thereby inferring latent-rule representations that facilitate coordinated decision-making.
- Empirical results on StarCraft II and MPE benchmarks demonstrate that DIRECT significantly outperforms state-of-the-art baselines, exhibiting coordination under varying levels of intermittent communication.

2 Related Work

Research on communication-enabled MARL can be broadly organized along several dimensions: when communication is triggered based on its expected utility [Guo *et al.*, 2023], who communicates through structured or graph-based topologies [Du *et al.*, 2025; Hu *et al.*, 2024; Xu *et al.*, 2023b; Xie *et al.*, 2025], how received information is integrated into decision-making [Sun *et al.*, 2024a], and what information is exchanged, including compact semantic representations of local observations or latent features [Meng and Tan, 2024; Du *et al.*, 2024]. In this line of work, intention-based communication has emerged as a prominent paradigm, focusing on sharing representations of anticipated future behavior.

Recent intention-based approaches adopt diverse representations, ranging from explicit forecasts of future trajectories or action sequences [Kim *et al.*, 2021; Fang *et al.*, 2023], to more abstract sub-goals or task-critical intermediate states [Zhang *et al.*, 2025], as well as consensus-driven formulations that infer shared group-level guidance without transmitting explicit plans [Li *et al.*, 2025; Xu *et al.*, 2023a]. These methods generally assume stable communication and promote coordination by driving agents toward global consensus. In contrast, our approach grounds communicated intentions in agents' executed actions and leverages an emergent latent rule to regulate conflicts among intentions, enabling coordination under intermittent communication.

From the perspective of communication constraints, another line of recent work studies multi-agent coordination under imperfect communication conditions, such as limited bandwidth and temporal delays. Bandwidth-oriented methods address transmission budgets and semantic compression [Saleh and Rayna, 2025; Yu *et al.*, 2024; Chen *et al.*, 2025; Sun *et al.*, 2024b], while latency- and asynchrony-aware approaches focus on mitigating delayed or misaligned information arrivals [Yuan *et al.*, 2023; Song *et al.*, 2025]. These methods improve robustness to compressed or delayed communication, but generally assume persistent link availability, where information eventually arrives. In contrast, our work considers nonpersistent connectivity, in which communication links may fail entirely and information can be absent for extended periods.

3 Preliminary

A Decentralized Partially Observable Markov Decision Process (Dec-POMDP) is defined by the tuple $G = \langle N, \mathcal{S}, \mathcal{O}, \Omega, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma \rangle$, where $N = \{1, \dots, n\}$ denotes the set of agents, and $s \in \mathcal{S}$ is the global state. At each timestep t, each agent $i \in N$ receives a local observation $o_t^i \sim \Omega(\cdot \mid s_t, i)$ and selects an action $a_t^i \in \mathcal{A}^i$. The resulting joint action $\mathbf{a}_t = \langle a_t^1, \dots, a_t^n \rangle \in \mathcal{A}$, where $\mathcal{A} = \prod_{i \in N} \mathcal{A}^i$, induces a state transition $s_{t+1} \sim \mathcal{P}(\cdot \mid s_t, \mathbf{a}_t)$ and yields a shared reward $r_t = \mathcal{R}(s_t, \mathbf{a}_t)$. The objective is to maximize the expected discounted return $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$.

In this paper, we formalize intermittent communication as a time-varying connectivity process, extending the standard formulation to an Intermittent Communication Dec-POMDP (IC-Dec-POMDP). This is defined by the augmented tuple $G_{IC} = \langle N, \mathcal{S}, \mathcal{O}, \Omega, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \mathcal{C} \rangle$, where $\mathcal{C}$ characterizes the stochastic availability of communication links. Specifically, the connectivity for each directed pair $(j \rightarrow i)$ at timestep t is modeled as an independent Bernoulli process: $C_t^{ji} \sim \text{Bernoulli}(1 - \omega)$. Here, $C_t^{ji} = 1$ denotes successful transmission, 0 indicates link failure, and $\omega \in [0, 1]$ represents the failure probability. Accordingly, the communication configuration $\mathcal{C}$ is formalized as $\mathcal{C} = \{C_t^{ji} : t \geq 0; i, j \in N\}$.

4 Method

In this section, we introduce DIRECT, a framework that realizes the $Intention \otimes Latent\ Rule$ formulation for decentralized coordination under intermittent communication. The key idea is to mitigate the impact of conflicting intentions via the emergent latent rule, rather than relying on a hand-crafted fixed rule. As illustrated in Fig. 2, DIRECT comprises two core components: an *Intention Forecaster Module*, which derives a communicated intention representation for each agent from its recent behavior and temporal context, and a *Latent*

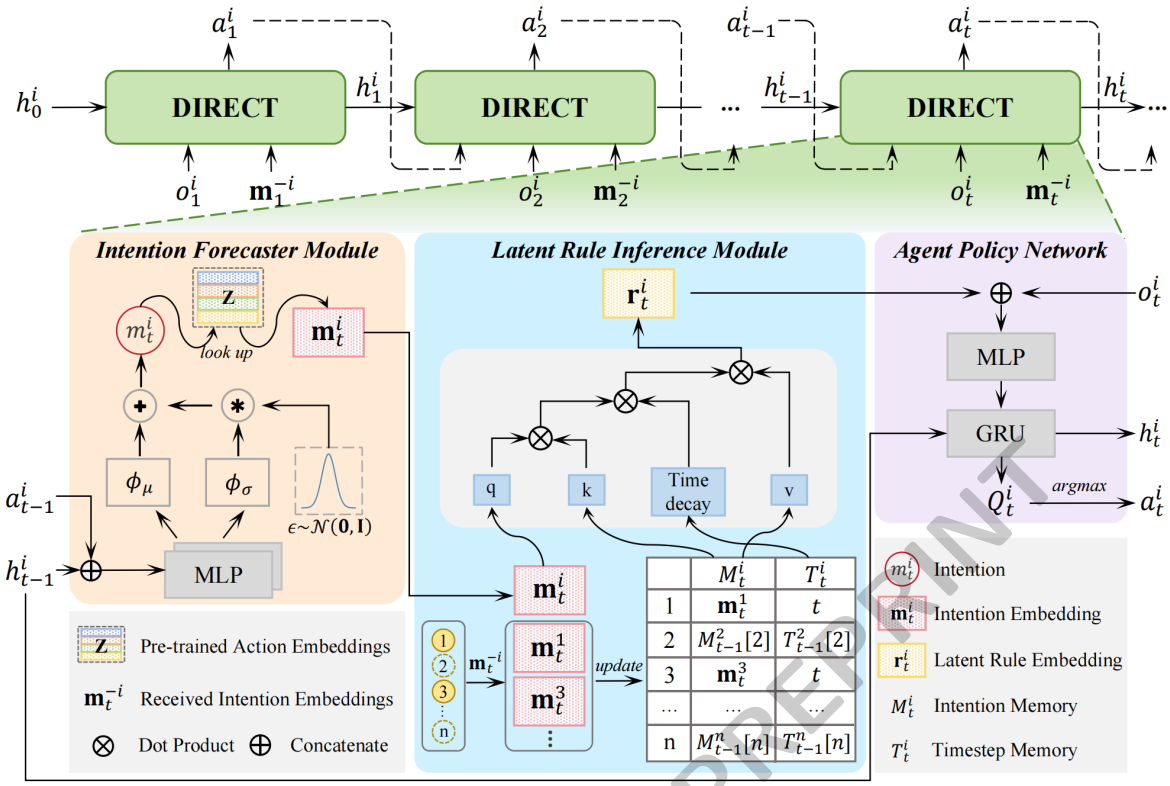

Figure 2: Overview of the proposed DIRECT framework. The top panel illustrates the recurrent information flow across timesteps, while the bottom panel details the architectural workflow at timestep t under the *Intention* $\otimes$ *Latent Rule* formulation. Specifically, the *Intention Forecaster Module* predicts an intention embedding $\mathbf{m}_t^i$ for each agent based on its recent action a_{t-1}^i. To handle communication volatility, the *Latent Rule Inference Module* maintains an episodic memory (M_t^i, T_t^i) of received information and applies time-decay weighting to account for the varying recency and contextual relevance of stored intentions. Based on this memory, the module generates a local latent-rule representation for agent i, which serves as a coordination prior for the *Agent Policy Network* to adjust local action selection.

Rule Inference Module, which instantiates this shared logic into a context-dependent local representation to regulate how these intentions influence local decision-making.

4.1 Intention Forecaster Module

The *Intention Forecaster Module* is designed to derive an intention from the agent's executed action. At timestep t, the module conditions on the agent's immediate history by concatenating the recurrent hidden state h_{t-1}^i with the executed action a_{t-1}^i. The resulting vector $x_t^i = [h_{t-1}^i, a_{t-1}^i]$ encodes a compact representation of the agent's recent behavior. This context vector x_t^i is then passed through a non-linear encoder ϕ_{enc} to obtain a feature representation $e_t^i = \phi_{\text{enc}}(x_t^i)$, which is used to parameterize the diagonal Gaussian distribution over the intention space via two linear heads ϕ_μ and ϕ_σ:

$$\mu_t^i = \phi_\mu(e_t^i), \quad \log \sigma_t^i = \phi_\sigma(e_t^i). \tag{1}$$

Then, the intention is obtained via the reparameterization trick to enable gradient-based optimization:

$$m_t^i = \mu_t^i + \epsilon \odot \exp(\log \sigma_t^i), \quad \text{where } \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}). \tag{2}$$

To regularize the intention distribution, we minimize the Kullback–Leibler (KL) divergence between the predicted distribution $q(m_t^i \mid x_t^i)$ and a standard Gaussian prior $p(m_t^i)$:

$$\mathcal{L}_{KL} = D_{KL}\big(q(m_t^i \mid x_t^i) \,\|\, p(m_t^i)\big). \tag{3}$$

To ensure a shared semantic interpretation of intentions, we construct a unified action embedding space that serves as a grounding interface between intentions and executable actions. Specifically, following the design of RODE [Wang *et al.*, 2021], we pre-train a single action encoder ϕ_e that maps each action $a \in \mathcal{A}$ to a d-dimensional embedding $\mathbf{z}_a = \phi_e(a)$. The resulting embedding matrix $\mathbf{Z} \in \mathbb{R}^{|\mathcal{A}| \times d}$ is shared by all agents. The encoder ϕ_e is trained within a predictive model to capture the dynamical and reward consequences of actions by jointly predicting the next local observation and the global reward, encouraging the embeddings to capture the functional effects of actions on environment dynamics. After pre-training, ϕ_e is frozen and used as a fixed semantic reference.

With the shared encoder ϕ_e frozen, we project the predicted intention $m_t^i \in \mathbb{R}^{|\mathcal{A}|}$ into the shared action embedding space, yielding a grounded intention embedding $\mathbf{m}_t^i \in \mathbb{R}^d$. Specifically, $\mathbf{m}_t^i$ is calculated as the expected action embedding under the softmax distribution defined by m_t^i:

$$\mathbf{m}_t^i = \mathbb{E}_{a \sim P_t^i}[\mathbf{z}_a], \quad P_t^i(a) \propto \exp(m_t^i(a)/\tau), \tag{4}$$

where $m_t^i(a)$ denotes the logit score associated with action a, and τ is a temperature parameter.

4.2 Latent Rule Inference Module

The *Latent Rule Inference Module* operates under the intermittent communication setting defined in Sec. 3. In this setting, each agent i maintains an intention memory $\mathcal{M}_t^i = (M_t^i, T_t^i)$, consisting of a content component M_t^i and a temporal index component T_t^i. For each teammate j, $M_t^i[j]$ stores the most recently received grounded intention from agent j, while $T_t^i[j]$ records the timestep at which that entry was received. This memory is updated according to the communication protocol $\mathcal{C}$. Whenever information from agent j is successfully delivered to agent i at timestep t, the corresponding memory entry is refreshed; otherwise, the memory entry remains unchanged:

$$(M_t^i[j], T_t^i[j]) \leftarrow \begin{cases} (\mathbf{m}_t^j, t), & \text{if } C_t^{ji} = 1, \\ (M_{t-1}^i[j], T_{t-1}^i[j]), & \text{otherwise.} \end{cases} \quad (5)$$

Building on the intention memory $\mathcal{M}_t^i$, we derive $\mathbf{r}_t^i$ as a local instantiation of the latent rule. In our framework, the latent rule is defined as a globally shared coordination logic parameterized by the *Latent Rule Inference Module*, while $\mathbf{r}_t^i$ represents its private, context-dependent realization for agent i at timestep t. By weighting teammate intentions based on their current context and recency, this module allows each agent to adaptively modulate how diverse intentions influence local decision-making under intermittent communication. Specifically, the relevance score is computed as:

$$\tilde{\alpha}_t^{ij} = \exp\left(\frac{(W_Q \mathbf{m}_t^i)^\top (W_K M_t^i[j])}{\sqrt{d}} - \beta(t - T_t^i[j]) \right), \quad (6)$$

where $W_Q, W_K \in \mathbb{R}^{d \times d}$ are learnable projections, and $\beta > 0$ controls temporal decay. The normalized weights are obtained via:

$$\alpha_t^{ij} = \frac{\tilde{\alpha}_t^{ij}}{\sum_{k \in N_t^i} \tilde{\alpha}_t^{ik}}. \quad (7)$$

The resulting latent representation $\mathbf{r}_t^i$ for agent i is then represented as a weighted combination of teammate intentions:

$$\mathbf{r}_t^i = \sum_{j \in N_t^i} \alpha_t^{ij} W_V M_t^i[j], \quad (8)$$

where $W_V \in \mathbb{R}^{d \times d}$ maps intentions into the latent rule space. Rather than a fixed rule, $\mathbf{r}_t^i$ serves as a coordination prior that biases the agent's subsequent decision-making toward mitigating inter-agent conflicts.

4.3 Agent Policy Network

To incorporate the inferred latent rule representation into local decision-making, the agent conditions its policy on both the representation $\mathbf{r}_t^i$ and the observation o_t^i. Specifically, $\mathbf{r}_t^i$ and o_t^i are concatenated and passed through an MLP for feature projection, and the resulting representation is then used to update the internal hidden state at each timestep t as:

$$h_t^i = \mathrm{GRU}([\mathbf{r}_t^i, o_t^i], h_{t-1}^i), \quad (9)$$

from which per-action value estimates are produced via:

$$Q_t^i(\cdot) = f_Q(h_t^i). \quad (10)$$

During decentralized execution, the agent selects its action according to:

$$a_t^i = \arg\max_a Q_t^i(a), \quad (11)$$

while centralized training aggregates individual action-value functions through the QMIX mixer [Rashid *et al.*, 2020].

4.4 Training

The overall training objective augments the standard QMIX loss with an auxiliary regularization term on the latent intention space. Specifically, the global reinforcement learning objective minimizes the temporal-difference (TD) error:

$$\mathcal{L}_{rl} = \mathbb{E}_{\mathcal{D}}\left[(y_t - Q_{tot}(\mathbf{s}_t, \mathbf{a}_t))^2\right], \quad (12)$$

where $y_t = r_t + \gamma \max_{\mathbf{a}_{t+1}} \bar{Q}_{tot}(\mathbf{s}_{t+1}, \mathbf{a}_{t+1})$ is the TD target, $\bar{Q}_{tot}$ is the target network, and $\mathcal{D}$ denotes the replay buffer of collected transitions.

The complete training objective is then given by:

$$\mathcal{L}_{tot} = \mathcal{L}_{rl} + \lambda_{KL} \mathcal{L}_{KL}, \quad (13)$$

where $\mathcal{L}_{KL}$ is the auxiliary KL regularization term defined in Sec. 4.1, and λ_{KL} controls its relative contribution.

5 Experiment

In this section, we empirically evaluate DIRECT on two diverse benchmarks: the StarCraft II micromanagement (SMAC) benchmark [Samvelyan *et al.*, 2019] and the Multi-Agent Particle Environment (MPE) [Mordatch and Abbeel, 2018]. Through these experiments, we aim to answer three core questions regarding the framework's effectiveness: (1) Does DIRECT outperform state-of-the-art communication- and intention-based baselines, particularly in complex cooperative scenarios? (See Sec. 5.1). (2) Can the *Intention* $\otimes$ *Latent Rule* paradigm effectively translate communicated intentions into coordinated behavior? (See Sec. 5.2). (3) Can DIRECT robustly sustain coordinated behavior across different degrees of intermittent communication? (See Sec. 5.3).

Baselines. We compare DIRECT against a set of baseline methods drawn from three representative categories. First, we consider intention- and consensus-oriented methods, including M2I2 [Sun *et al.*, 2024a], which models teammate intentions via masked state prediction, and COLA [Xu *et al.*, 2023a], which learns the implicit consensus from local observations. Second, we further include general communication frameworks that focus on learning structured and efficient information exchange, such as ROCO [Xie *et al.*, 2025], MAIC [Yuan *et al.*, 2022], and NDQ [Wang *et al.*, 2020]. Third, as non-communicative baselines, we adopt QMIX [Rashid *et al.*, 2020] to assess the necessity of explicit information exchange.

5.1 Performance Comparison

To ensure fair robustness, each experiment was repeated with five random seeds, and we report the mean performance with 95% confidence intervals. Besides, to simulate intermittent communication, we set the link failure probability ω to 0.6 across all experiments. For the baselines, M2I2, COLA, ROCO, MAIC, and NDQ were implemented based on the

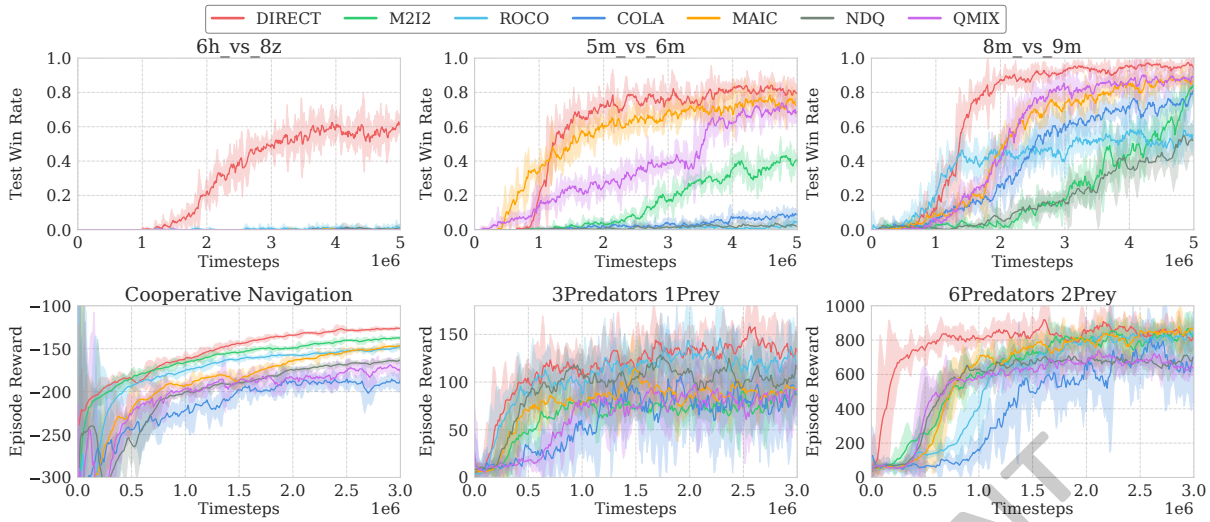

Figure 3: Performance comparison between DIRECT and baselines on the SMAC and MPE benchmarks.

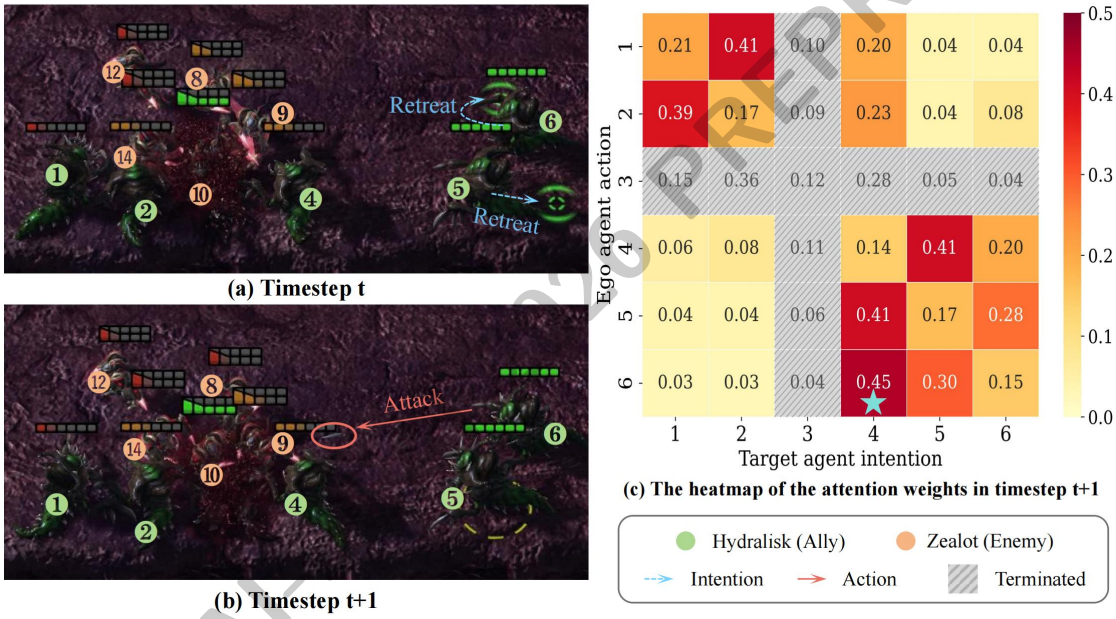

Figure 4: Intention conflict resolution via latent rule emergence in the SMAC *6h_vs_8z* scenario. (a–b) Consecutive timesteps showing conflicting intentions at t and coordinated action adjustment at $t+1$. (c) Attention weight heatmap at $t+1$, illustrating how agents selectively attend to teammates' intentions when resolving conflicts.

open-source PyMARL framework. Additionally, we adapted QMIX to accommodate the communication setting by concatenating received information with their local observations. This ensures they operate under the same information availability constraints as other baselines.

The top row of Fig. 3 presents the performance on the SMAC benchmark. DIRECT consistently outperforms baselines across all maps, achieving higher final win rates and more stable training behavior. This advantage is most pronounced on the heterogeneous map *6h_vs_8z*, a scenario demanding precise coordination among agents with heterogeneous unit capabilities. While baselines struggle near a zero win rate—likely due to their inability to resolve conflicting intentions without regulation—DIRECT achieves a 60% win rate. This validates the efficacy of the latent rule in modulating disparate action-grounded signals. On symmetric maps like *5m_vs_6m* and *8m_vs_9m*, DIRECT similarly maintains its lead through more stable learning dynamics.

The bottom row of Fig. 3 reports results on continuous control tasks. Across all evaluated environments, DIRECT consistently achieves higher episode rewards than the baseline methods. In contrast, approaches such as COLA and ROCO exhibit pronounced performance oscillations, particularly under intermittent communication, whereas DIRECT maintains

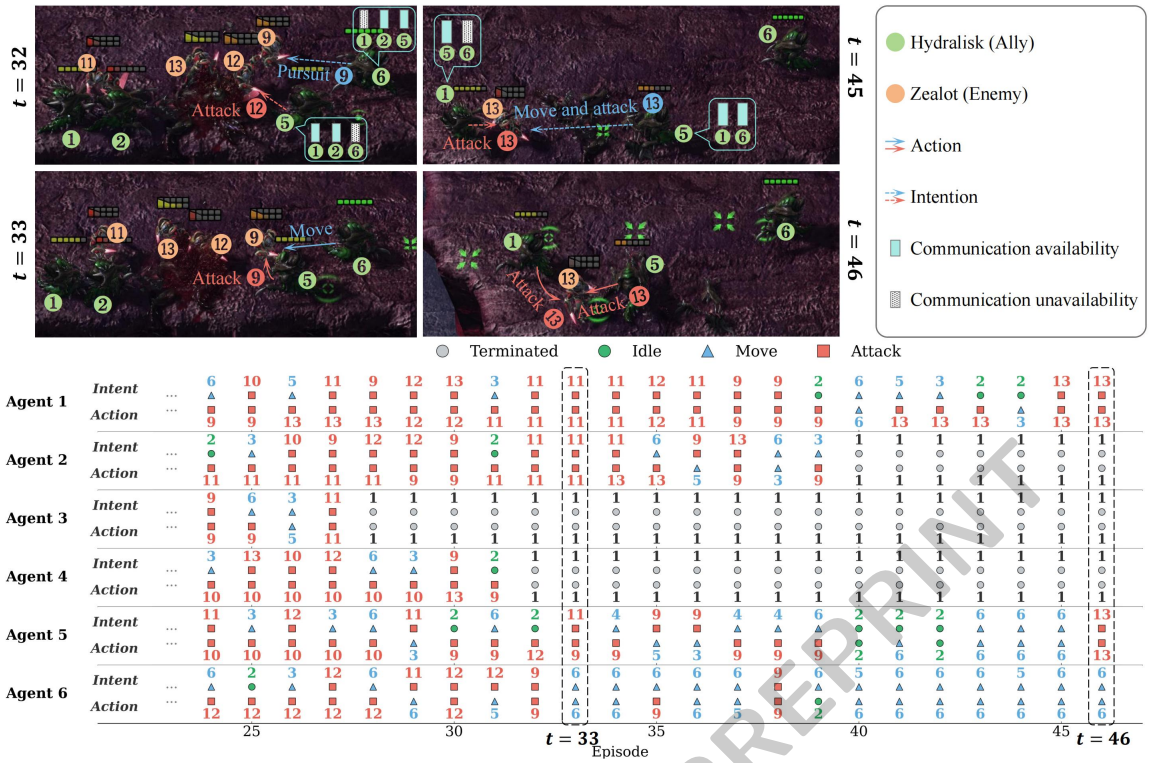

Figure 5: Representative snapshots and timeline analysis from the *6h_vs_8z* scenario. The top panels visualize selected timesteps, showing communicated action-grounded intentions as dashed arrows and the corresponding executed actions at the subsequent timestep as solid arrows. The bottom panel summarizes, for each agent, the temporal sequences of communicated intentions (upper row) and executed actions (lower row) over the latter half of a selected episode.

smoother reward trajectories throughout training. This consistent behavior indicates greater robustness to inconsistent teammate information and demonstrates that the overall design of DIRECT supports stable coordination in continuous control tasks, enabling agents to learn reliable policies even when communication is intermittent.

5.2 Case Study

To answer the second question, we use two visual case studies to show how the *Intention* $\otimes$ *Latent Rule* paradigm enables effective coordination from communicated intentions. Fig. 4 illustrates how coordination in DIRECT emerges through a latent rule in the SMAC *6h_vs_8z* scenario. At timestep t, allied Hydralisk 5 and Hydralisk 6 communicate retreat intentions, while Hydralisk 4 is surrounded by enemy Zealots. At timestep $t+1$, Hydralisk 5 and Hydralisk 6 adjust their actions from retreat to attack, converging on a compatible response that alleviates the local disadvantage. This coordinated adjustment indicates that an emergent latent rule mitigates conflicting intentions by incorporating both communicated information and dynamic environmental state. The attention heatmap further shows that, during action selection, agents place higher emphasis on teammates whose communicated intentions are potentially incompatible with their own, indicating that the latent rule guides agents to adjust their actions when faced with incompatible teammate intentions, producing coherent joint actions.

Fig. 5 offers several interesting insights on how latent rules mediate the translation from communicated intentions to coordinated actions under intermittent communication. First, temporary mismatches between intentions and actions naturally arise from sudden changes in the environment, and the latent rule helps agents adjust their actions accordingly while preserving coordinated team behavior. At $t = 32$, allied Hydralisk 6 communicates an intention to pursue enemy Zealot 9, while Hydralisk 5 initially targets Zealot 12. As Zealot 12 moves out of favorable engagement range, Hydralisk 5 reallocates its target to Zealot 9 at $t = 33$, reflecting a coordinated adjustment driven by both teammate intention and updated local state. Second, when agents infer mutually compatible intentions, the latent rule reinforces this alignment and induces coordinated execution. At $t = 45$, allied Hydralisk 1 and Hydralisk 5 communicate aligned intentions toward enemy Zealot 13, which are directly realized at $t = 46$ as a coordinated joint attack. Third, over the latter half of the selected episode, communicated intentions exhibit a clear alignment with executed actions over time. As shown in the bottom panel of Fig. 5, this alignment is reflected by the dominant color correspondence between intention and action sequences, indicating that intentions serve as high-level behavioral abstractions that guide coordination without constraining execution. Together, these insights show that the latent rule both stabilizes coordination under local perturbations and accelerates alignment when shared objectives emerge.

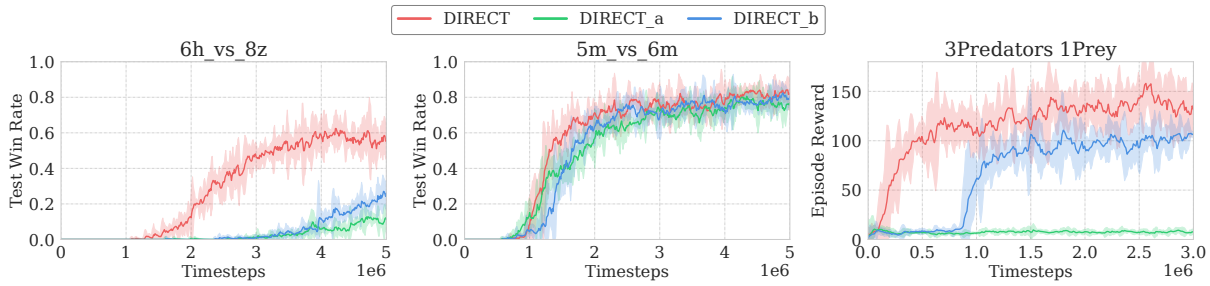

Figure 6: Ablation study on the core components of DIRECT. DIRECT_a removes the pretrained action encoder in *Intention Forecaster Module*, eliminating the action-grounded intention representation. DIRECT_b removes the *Latent Rule Inference Module*, directly using communicated intentions for action selection without latent rule emergence.

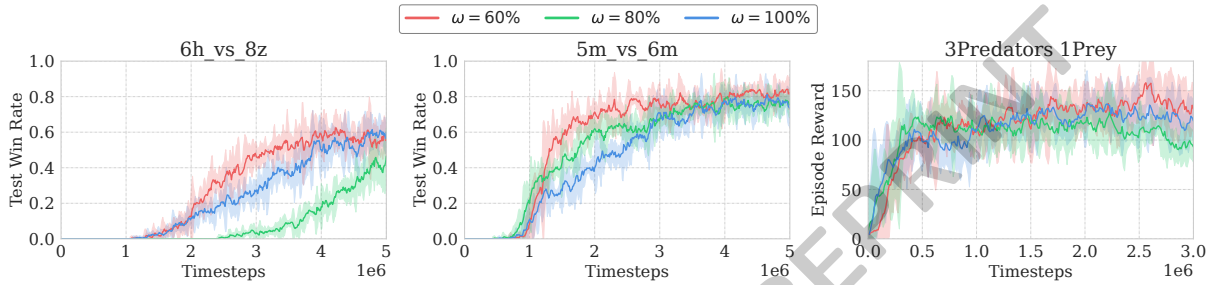

Figure 7: Sensitivity analysis of DIRECT under varying link failure probabilities ω on the SMAC and MPE benchmarks.

5.3 Ablation and Sensitivity Analysis

We evaluate the impact of key components in Fig. 6. Removing the pre-trained action encoder (DIRECT_a) leads to a significant performance collapse on the complex *6h_vs_8z* map, indicating that grounding communicated information in actions is essential for effective coordination. Without this grounding, agents struggle to interpret the intention behind information received from teammates. In contrast, removing the *Latent Rule Inference Module* (DIRECT_b) results also degrades performance on complex tasks, showing that action-grounded intentions alone are not sufficient. Without a mechanism to account for outdated information, coordination becomes unstable. By combining both components, DIRECT achieves consistently strong performance, demonstrating that action grounding and latent rule regulation play complementary roles in supporting robust coordination.

As shown in Fig. 7, we evaluate the robustness of DIRECT under varying link failure probabilities ω. As ω decreases, agents receive intention more frequently, which generally corresponds to better coordination performance. Interestingly, highly intermittent communication does not consistently outperform the no-communication setting with $\omega = 100\%$. One possible explanation is that sporadic intention updates can lead to mismatches between communicated intentions and the actions agents actually execute, causing the exchanged information to act as noise. In contrast, when communication is entirely absent, agents rely on a consistent local information stream, which can result in more stable learning behavior. Importantly, DIRECT does not exhibit noticeable performance degradation under such noisy communication conditions. Its performance remains comparable to the stable baseline, suggesting that the framework can effectively tolerate intention–action mismatches without compromising learning stability.

6 Conclusion

In this paper, we proposed DIRECT, a decentralized communication framework for cooperative multi-agent reinforcement learning under intermittent communication. Bypassing the need for global agreement, DIRECT follows the $Intention \otimes Latent\ Rule$ paradigm, in which agents share behavior-grounded intentions, and an emergent latent rule is used to mitigate conflicts among these communicated intentions when making decisions. By grounding intentions in executed actions and adapting to partial and temporally fragmented information, DIRECT ensures robust cooperation even in intermittent communication environments. Extensive empirical analyses demonstrate that DIRECT achieves robust coordination under intermittent communication. Rather than enforcing global consensus, the results show that coordination emerges from how the latent rule mitigates conflicts among communicated intentions during decision-making. In particular, when communicated intentions temporarily diverge from subsequent actions due to dynamic state changes, coordinated behavior is maintained despite these discrepancies. A promising direction for future work is to integrate Vision–Language–Action (VLA) models as a shared grounding interface for intentions. By aligning DIRECT's intention embeddings with the pre-trained action representations of VLA, the $Intention \otimes Latent\ Rule$ paradigm could be extended to open-ended embodied settings, enabling zero-shot coordination with previously unseen partners.

Acknowledgments

This work is supported by National Natural Science Foundation of China (Grant No. 62276238, U24A20326, 62325602, 62036010), the Foundation of Henan Educational Committee (Grant No. 25HASTIT034), the Natural Science Foundation of Henan (Grant No. 232300421095).

References

[Chen *et al.*, 2025] Jinchao Chen, Qiuhao Shu, Yantao Lu, and et al. Qctf: A quantized communication and transferable fusion framework for multi-agent collaborative perception. *IEEE Transactions on Intelligent Transportation Systems*, 26:15013 – 15027, 2025.

[Clausewitz, 2003] Carl Clausewitz. *On war*. Penguin UK, 2003.

[Dann *et al.*, 2023] Michael Dann, Yuan Yao, Natasha Alechina, and et al. Multi-agent intention recognition and progression. In *IJCAI*, pages 91–99, 2023.

[Du *et al.*, 2024] Wei Du, Shifei Ding, Lili Guo, and et al. Expressive multi-agent communication via identity-aware learning. In *AAAI*, pages 17354–17361, 2024.

[Du *et al.*, 2025] Wei Du, Shifei Ding, Wei Guo, and et al. Multi-agent communication with information preserving graph contrastive learning. In *IJCAI*, pages 64–71, 2025.

[Fang *et al.*, 2023] Yuchen Fang, Zhenggang Tang, Kan Ren, and et al. Learning multi-agent intention-aware communication for optimal multi-order execution in finance. In *KDD*, pages 4003–4012, 2023.

[Guan *et al.*, 2022] Cong Guan, Feng Chen, Lei Yuan, and et al. Efficient multi-agent communication via self-supervised information aggregation. In *NeurIPS*, pages 1020–1033, 2022.

[Guo *et al.*, 2023] Xudong Guo, Daming Shi, and Wenhui Fan. Scalable communication for multi-agent reinforcement learning via transformer-based email mechanism. In *IJCAI*, pages 126–134, 2023.

[Hu *et al.*, 2024] Shengchao Hu, Li Shen, Ya Zhang, and et al. Learning multi-agent communication from graph modeling perspective. In *ICLR*, 2024.

[Kim *et al.*, 2021] Woojun Kim, Jongeui Park, and Youngchul Sung. Communication in multi-agent reinforcement learning: intention sharing. In *ICLR*, 2021.

[Li *et al.*, 2025] Dapeng Li, Na Lou, Zhiwei Xu, and et al. Efficient communication in multi-agent reinforcement learning with implicit consensus generation. *AAAI*, pages 23240–23248, 2025.

[Lin *et al.*, 2021] Toru Lin, Jacob Huh, Christopher Stauffer, and et al. Learning to ground multi-agent communication with autoencoders. *NeurIPS*, pages 15230–15242, 2021.

[Liu *et al.*, 2025] Qidong Liu, Shaoyao Niu, Chaoyue Liu, and et al. Llt: Learning latent tactics in uncertainty cooperative multi-agent reinforcement learning. In *YAC*, pages 1326–1332, 2025.

[Lo *et al.*, 2024] Yat Long Lo, Biswa Sengupta, Jakob Nicolaus Foerster, and et al. Learning multi-agent communication with contrastive learning. In *ICLR*, 2024.

[Meng and Tan, 2024] Xiangrui Meng and Ying Tan. Pmac: Personalized multi-agent communication. In *AAAI*, pages 17505–17513, 2024.

[Mordatch and Abbeel, 2018] Igor Mordatch and Pieter Abbeel. Emergence of grounded compositional language in multi-agent populations. In *AAAI*, pages 1495–1502, 2018.

[Rashid *et al.*, 2020] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, and et al. Monotonic value function factorisation for deep multi-agent reinforcement learning. *Journal of Machine Learning Research*, 21(178):1–51, 2020.

[Roth *et al.*, 2006] Maayan Roth, Reid G. Simmons, and Manuela Veloso. What to communicate? execution-time decision in multi-agent pomdps. In *Distributed Autonomous Robotic Systems 7*, pages 177–186, 2006.

[Saleh and Rayna, 2025] Soudijani Saleh and Dimitrova Rayna. Synthesis of communication policies for multi-agent systems robust to communication restrictions. In *IJCAI*, pages 257–266, 2025.

[Samvelyan *et al.*, 2019] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, and et al. The starcraft multi-agent challenge. In *AAMAS*, pages 2186–2188, 2019.

[Song *et al.*, 2025] Shoucheng Song, Youfang Lin, Sheng Han, and et al. Code: Communication delay-tolerant multi-agent collaboration via dual alignment of intent and timeliness. In *AAAI*, pages 23304–23312, 2025.

[Sun *et al.*, 2024a] Chuxiong Sun, Peng He, Qirui Ji, Zehua Zang, Jiangmeng Li, Rui Wang, and Wei Wang. M2i2: Learning efficient multi-agent communication via masked state modeling and intention inference. *arXiv preprint arXiv:2501.00312*, 2024.

[Sun *et al.*, 2024b] Qingshuang Sun, Denis Steckelmacher, and Yuan et al. Yao. Dynamic size message scheduling for multi-agent communication under limited bandwidth. *IEEE Transactions on Mobile Computing*, 23:15080–15097, 2024.

[Udagawa and Aizawa, 2019] Takuma Udagawa and Akiko Aizawa. A natural language corpus of common grounding under continuous and partially-observable context. In *AAAI*, pages 7120–7127, 2019.

[Wang *et al.*, 2020] Tonghan Wang, Jianhao Wang, Chongyi Zheng, and et al. Learning nearly decomposable value functions via communication minimization. In *ICLR*, 2020.

[Wang *et al.*, 2021] Tonghan Wang, Tarun Gupta, Anuj Mahajan, and et al. Rode: Learning roles to decompose multi-agent tasks. In *ICLR*, 2021.

[Xie *et al.*, 2025] Zaipeng Xie, Sitong Shen, Yaowu Wang, and et al. Roco: Role-oriented communication for efficient

multi-agent reinforcement learning. *Expert Systems with Applications*, page 129421, 2025.

[Xu *et al.*, 2023a] Zhiwei Xu, Bin Zhang, Dapeng Li, and et al. Consensus learning for cooperative multi-agent reinforcement learning. In *AAAI*, pages 11726–11734, 2023.

[Xu *et al.*, 2023b] Zhiwei Xu, Bin Zhang, Dapeng Li, and et al. Deep hierarchical communication graph in multi-agent reinforcement learning. In *IJCAI*, pages 208–216, 2023.

[Yao *et al.*, 2025] Enguang Yao, Qidong Liu, Jiajia Hou, and et al. Latac: Latent tactics for robust multi-agent coordination under intermittent communication. *Science China Information Sciences*, 2025. http://engine.scichina.com/doi/10.1007/s11432-025-4724-0.

[Yu *et al.*, 2024] Lebin Yu, Qiexiang Wang, Yunbo Qiu, and et al. Effective multi-agent communication under limited bandwidth. *IEEE Transactions on Mobile Computing*, 23:7771–7784, 2024.

[Yuan *et al.*, 2022] Lei Yuan, Jianhao Wang, Fuxiang Zhang, and et al. Multi-agent incentive communication via decentralized teammate modeling. In *AAAI*, pages 9466–9474, 2022.

[Yuan *et al.*, 2023] Tingting Yuan, Hwei-Ming Chung, Jie Yuan, and et al. Dacom: Learning delay-aware communication for multi-agent reinforcement learning. In *AAAI*, pages 11763–11771, 2023.

[Zhang *et al.*, 2025] Zeren Zhang, Zhiwei Xu, Guangchong Zhou, and et al. Unveiling decision intention for cooperative multi-agent reinforcement learning. In *AAMAS*, pages 2345–2354, 2025.