On Diffusion Models for Multi-Agent Partial Observability: Shared Attractors, Error Bounds, and Composite Flow

Tonghan Wang[*]
Harvard University
Cambridge, MA 02138, USA
twang1@g.harvard.edu

Heng Dong[*]
Tsinghua University
Beijing, China
drdhxi@gmail.com

Yanchen Jiang
Harvard University
Cambridge, MA 02138, USA
yanchen_jiang@g.harvard.edu

David C. Parkes
Harvard University
Cambridge, MA 02138, USA
parkes@eecs.harvard.edu

Milind Tambe
Harvard University
Cambridge, MA 02138, USA
milind_tambe@harvard.edu

ABSTRACT

Multiagent systems grapple with partial observability (PO), and the decentralized POMDP (Dec-POMDP) model highlights the fundamental nature of this challenge. Whereas recent approaches to addressing PO have appealed to deep learning models, providing a rigorous understanding of how these models and their approximation errors affect agents' handling of PO and their interactions remain a challenge. In addressing this challenge, we investigate reconstructing global states from local action-observation histories in Dec-POMDPs using diffusion models. Wefi rstfi nd that diffusion models conditioned on local history represent possible states as stablefi xed points. In collectively observable (CO) Dec-POMDPs, individual diffusion models conditioned on agents' local histories share a uniquefi xed point corresponding to the global state, while in non-CO settings, sharedfi xed points yield a distribution of possible states given joint history. We furtherfi nd that, with deep learning approximation errors,fi xed points can deviate from true states and the deviation is negatively correlated to the Jacobian rank. Inspired by this low-rank property, we bound a deviation by constructing a surrogate linear regression model that approximates the local behavior of a diffusion model. With this bound, we propose a *composite diffusion process* iterating over agents with theoretical convergence guarantees to the true state.

KEYWORDS

Diffusion Model; Multi-Agent; Partial Observability; State Reconstruction from Observation; Dec-POMDP; Fixed Point

ACM Reference Format:
Tonghan Wang[*], Heng Dong[*], Yanchen Jiang, David C. Parkes, and Milind Tambe. 2025. On Diffusion Models for Multi-Agent Partial Observability: Shared Attractors, Error Bounds, and Composite Flow. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 18 pages.

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

Given that the ability of individual agents to perceive complete information about the global state is limited [2, 51, 62, 67], partial observability (PO) fundamentally characterizes the dynamics and interactions in multi-agent systems. Decentralized POMDPs (Dec-POMDPs) [50] highlight this information limitation, where many complex challenges are rooted in this issue, such as communication [16, 74], decentralized control [10, 82], cooperation[72, 77], and coordination [79, 85] under incomplete information.

Decades of research addressing PO in the context of these challenges [17, 31, 39, 66] have fostered the development of specialized sub-fields [16, 19, 22, 33, 52, 71, 87] within the multi-agent system, thereby shaping its current landscape. As a general way of handling the uncertainty due to PO, the concept of belief states is introduced to represent an agent's probabilistic state estimation based on local information[42, 47, 70]. While these methods effectively encapsulate uncertainty in some environments, traditionally, they may suffer from scalability issues due to the exponential growth in complexity of belief updates. Recent works use powerful deep learning models to address scalability, e.g., by directly predicting unseen state features [27, 47, 80, 81]. However, a rigorous understanding of how deep learning models and their approximation errors can impact agents' handling of PO and their interactions remains elusive – a gap we address in this paper using diffusion models.

Diffusion models [24, 29, 38, 64, 65] offer a novel promising avenue towards addressing uncertainty in Dec-POMDPs due to PO, specifically by learning the mapping from local histories to global states. The primary challenge in learning such mappings lies in its inherent stochasticity and problem scale. A single history may correspond to multiple possible states, resulting in a one-to-many, stochastic mapping. Diffusion models, with their ability to model stochastic processes through the iterative denoising inductive bias, offer new opportunities to address such stochasticity. Additionally, the spaces of histories and states are often continuous and high-dimensional, also making diffusion models well-suited due to their proven powerful representational capacity in such expansive spaces [5, 15, 18, 23, 25, 40, 60, 78].

In this paper, we conduct an in-depth investigation into the use of diffusion models to manage PO in Dec-POMDPs, offering theoretical understandings supported by empirical evidences to solve the new challenges in this effort. To meet the requirements of decentralized control, our study comprises two steps. First, each agent infers states using a diffusion model conditioned on its local history. In this phase, we address critical problems, including how a diffusion model represents multiple possible states given local history, how accurate this representation is when the denoiser network is over- and under-parameterized, and methods to quantify the agent's uncertainty regarding the state. For the second step, should uncertainty persist, we study how to resolve it and optimally determine the true state by merging diffusion processes of all agents.

Specifically, our contributions are as follows. Thefi rst contribution is about how diffusion models represent states. In scenarios with minimal deep learning approximation errors, for each state s consistent with local history τ, the diffusion model conditioned on τ learns to create a *stablefi xed point* at the location s. Then, the repeated application of the denoiser network induces a discrete-time flow that transports noisy inputs to these attractors.

When agent i's diffusion model conditioned on τ_i has a single fixed point, it can confidently infer the global state as this unique fixed point. Our second contribution relates to complex scenarios with multiplefi xed points. We establish that in collectively observable Dec-POMDPs, there exists a uniquefi xed point shared by all agents, corresponding precisely to the true global state. Moreover, in non-collectively observable Dec-POMDPs, where aggregating local information cannot fully reveal the true state, sharedfi xed points represent all possible states given the joint history, and diffusion models can reproduce their true posterior probabilities.

We then consider the influence of deep learning approximation errors typical in learning with large Dec-POMDPs. Wefi nd that the major impact is that agents'fi xed points might deviate from the true state. Our third contribution is to investigate the underlying causes of these deviations and propose a method to bound their norm. Our theoretical analyses and empirical evidence suggest that deviations are inversely correlated with the Jacobian rank of the denoiser network atfi xed points. This low-rank behavior enables us to approximate local behavior of diffusion models by a surrogate linear regression model, whose solution gives an upper bound to deviations. Empirical evidence supports the tightness of this bound.

The deviation offi xed points from states implies that it becomes impractical to determine the global state by intersecting thefi xed points, as deviations vary among agents. To solve this problem, our fourth contribution is to propose the concept of *composite diffusion* which denoises the input iteratively using denoiser networks conditioned on each agent's history. Theoretically, we prove that composite diffusion, regardless of the agents' order, converges to the true state with an error no larger than the deviation upper bound. We support the analyses by showing that composite diffusion leads to accurate global state estimation in the complex SMACv2 [14] benchmark across a variety of highly stochastic testing cases.

By providing a rigorous understanding of impacts of diffusion models on PO in Dec-POMDPs, this paper opens the door to newer algorithms for various multi-agent problems such as policy learning, coordination, and communication in complex environments.

Related Works. Diffusion models [24] have been extensively explored in single-agent settings, significantly advancing areas such as planning that require approximators of MDP dynamics [4, 26, 36], data synthesis for reinforcement learning [8, 41], and policy training on offline datasets [1, 9, 21]. In contrast, how to synergize diffusion models with multi-agent systems remains largely underexplored. Xu et al. [81] investigate diffusion models within Dec-POMDPs but do not focus on the inherent stochasticity in mapping local histories to states, nor do they resolve how to address the disagreements among agents regarding the true state. Similarly, these critical problems remain untouched in other deep learning approaches that attempt to learn low-dimensional state representations [27, 80] using variational autoencoder [11] or contrastive learning [7].

Orthogonal to our focus on studying the explicit reconstruction of states from local history, many sub-fields in multi-agent systems have developed innovative approaches to mitigate the effects of PO, such as modeling other agents [56, 86], intention inference [20, 34, 55], and communication [28, 32, 68] for better decision-making. In multi-agent reinforcement learning, it is popular to employ RNNs [22, 58, 71] or Transformers [12, 77] to process sequential history data, while incorporating global information during centralized training by techniques like value function decomposition [58, 69, 73] and global gradient approximation [75, 83].

2 DIFFUSION MODELS FOR DEC-POMDPS

A Dec-POMDP [3, 50] is a tuple G=$\langle C, \mathcal{S}, \mathcal{A}, P, R, \Omega, O, n, \gamma\rangle$, where $\mathcal{A}$ is thefi nite action set, C is thefi nite set of n agents, $\gamma \in [0, 1)$ is the discount factor, and $s \in \mathcal{S} \subseteq \mathbb{R}^{|s|}$ is the true state. $|s|$ is the dimension of s. We consider partially observable settings and agent i only has access to an observation $o_i \in \Omega$ drawn according to the observation function $O(s, i)$. Each agent has a history $\tau_i \in \mathcal{T} \equiv (\Omega \times \mathcal{A})^* \times \Omega$. At each timestep, each agent i selects an action $a_i \in \mathcal{A}$, forming a joint action $\boldsymbol{a} \in \mathcal{A}^n$, leading to the next state s' according to the transition function $P(s'|s, \boldsymbol{a})$ and a shared reward $R(s, \boldsymbol{a})$ for each agent. The joint history of all agents is denoted by $\tau_{1:n}$, or $\boldsymbol{\tau}$ when the agent order is irrelevant.

When referring to local history without specifying that it pertains to a particular agent, we employ the notation τ. The mapping from τ to the corresponding global state s is a one-to-many mapping due to partial observability. We formalize this mapping as follows.

Definition 1 [History-State Mapping]. $S_G : \mathcal{T} \to 2^{\mathcal{S}}$ maps τ_i to the set of all possible states when agent i observes τ_i: $S_G(\tau_i) = \{s \mid p(s|\tau_i) > 0\}$, where $p(s|\tau_i)$ is the posterior state distribution. We say that the states in $S_G(\tau_i)$ are *consistent* with τ_i.

We use diffusion models to learn the mapping S_G and reproduce the posterior $p(s|\tau_i)$.

Diffusion models and scores. Given a training dataset $\mathcal{D} = \{(\tau_i^{(k)}, s^{(k)})\}_{k=1}^K$, where $s^{(k)} \in S_G(\tau_i^{(k)})$ is a state consistent with local history $\tau_i^{(k)}$, a score-based model $f_\theta : \mathcal{T} \times \mathbb{R}^{|s|} \to \mathbb{R}^{|s|}$ (also called a *denoiser network*) is trained to minimize

$$MSE(f_\theta, \sigma) = \mathbb{E}_{\tau_i, s, y \sim s+z}\left[\|s - f_\theta(\tau_i, y)\|^2\right], \quad (1)$$

Here $y = s + z$, where $z \sim \mathcal{N}(0, \sigma^2 I)$. We refer to s as clean states and y as noisy states. During training, noisy states are generated by injecting a randomly sampled noise with a noise level $\sigma > 0$ to s. The training involves the histories of all agents and the corresponding states. Our analyses in this paper are applicable to most neural network architectures, while in our experiments, we employ a simple fully-connected network with τ_i and y as inputs and denoised state as output (details in Appendix B). This network is shared among all the agents.

As shown in Appx. A.1, adapting the derivation from Kadkhodaie et al. [29], Miyasawa et al. [43], Robbins [59], the optimal denoiser network yields the expected state given the noisy input y:

$$f^\star(\tau_i, y) = \mathbb{E}_s\left[s \middle| y, \tau_i\right], \tag{2}$$

which is related to the conditional scores by

$$\nabla \log p_\sigma(y|\tau_i) = \frac{1}{\sigma^2}\left(\mathbb{E}_s\left[s \middle| y, \tau_i\right] - y\right). \tag{3}$$

Posterior state distribution and discrete-timefl ow. We are concerned with the estimated states and their distribution given history τ_i. We diffuse states consistent with τ_i to noise by a diffusion process characterized by the "variance-exploding" stochastic differential equation (SDE) [65]:

$$dy = g(t)d\mathbf{w}, \quad g(t) = \sqrt{\frac{d[\sigma^2(t)]}{dt}}, \tag{4}$$

where $\mathbf{w}$ is the standard Wiener process, *i.e.*, Brownian motion. Let $\sigma(t) = Ae^t$, where A is a constant, $t \in [0, T]$, and T is the maximum timestep. According to Eq. 4, we have $g^2(t) = 2\sigma^2(t)$. We generate states from noise by (reverse time) probabilityfl ow ordinary differential equation (ODE) conditioned on τ_i:

$$dy = -\sigma^2(t)\nabla_y \log p_t(y|\tau_i)dt, \tag{5}$$

which has the same marginal probability densities $\{p_t(y|\tau_i)\}_{t=0}^T$ as the time-reversal of the diffusing SDE in Eq. 4 [65]. Here dt represents an infinitesimal negative time step. We approximate the solution to Eq. 5 by iteration

$$y(t-1) = y(t) + \sigma^2(t)\nabla_{y(t)} \log p_t(y(t)|\tau_i) = f^\star(\tau_i, y(t)). \tag{6}$$

The second equality follows from Eq. 2, 3.

Rigorously, states are generated by applying a numerical solver to Eq. 5, and Eq. 6 brings discretization errors. To justify the use of this discretization, we show that we can stillfi nd the support of $p(s|\tau_i)$ (Thm. 1) and that errors of $p(s|\tau_i)$ can be bounded (Thm. 4).

In practice, we approximate the iteration in Eq. 6 by

$$y^{(\ell+1)} = f_\theta(\tau_i, y^{(\ell)}), \tag{7}$$

where ℓ is the iteration index, with increasing ℓ corresponding to decreasing t. Here, $f^\star$ is replaced by the trained score-based model f_θ. Deep learning approximation errors affect the accuracy of this iterative scheme, which is studied in Sec. 4. We formally describe our iteration algorithm by the following definition.

Definition 2 [Discrete-timefl ow]**.** A *discrete-timefl ow* $\phi_\ell(\tau_i, \theta) : \mathbb{N} \times \mathbb{R}^{|s|} \to \mathbb{R}^{|s|}$ conditioned on local history τ_i and denoiser network parameters θ is defined by $\phi_\ell(\tau_i, \theta)(y) = f_\theta(\tau_i, \phi_{\ell-1}(\tau_i, \theta)(y))$, with the initial condition $\phi_0(\tau_i, \theta)(y) = y$.

Intuitively, ϕ_ℓ generates a denoised state after applying the denoiser network for ℓ times, transporting a noisy state y to $y^{(\ell)} = \phi_\ell(\tau_i, \theta)(y)$. The distribution of these denoised states is given by the push-forward equation defined as follows.

Definition 3 [Push-forward equation]**.** The distribution of estimated states after applying f_θ for ℓ times is $p_\ell = [\phi_\ell(\tau_i, \theta)]_* p_0$, where p_0 is the distribution of noisy states, and the push-forward operator $*$ is defined by $[\phi_\ell]_* p_0(y) = p_0(\phi_\ell^{-1}(y)) \det\left[\partial\phi_\ell^{-1}/\partial y\right]$.

In this way, the estimated posterior state distribution given by the diffusion process is ($\ell \to \infty$ indicates $T \to \infty$):

$$p_\theta(s|\tau_i) = [\phi_{\ell\to\infty}(\tau_i, \theta)]_* p_0(s). \tag{8}$$

According to Definition 3, this distribution depends on the Jacobian $\partial\phi_\ell^{-1}/\partial y = (\partial\phi_\ell/\partial y)^{-1}$ [37]. As ϕ_ℓ is dependent on $f_\theta(\tau_i, y)$, our analysis would heavily utilize denoiser network Jacobian

$$J_f(y|\tau_i) = \nabla_y f_\theta(\tau_i, y) = \partial f/\partial y|_{(\tau_i, y)}. \tag{9}$$

$J_f(y|\tau_i)$ has eigenvalues $\lambda_k(y|\tau_i)$ and eigenvectors $e_k(y|\tau_i), k \in [|s|]$. Dependencies on y and τ_i will be omitted in these notations when they are unambiguous within the given context. An important property we use in this paper is the *Jacobian rank*, defined as the rank of the matrix $J^+(y|\tau) = \left(I - \frac{\partial f}{\partial y}(\tau, y)\right)^{-1} \frac{\partial f}{\partial \tau}(\tau, y)$.

3 STATES AS SHARED FIXED POINTS

We now present ourfi ndings on how diffusion models represent the one-to-many mapping from histories to states. Wefi rst consider the case with minimal influence of deep learning approximation errors in this section, and study more complex scenarios in Sec. 4.

3.1 Example

We start with a didactic example. Sensor networks are a classic problem in the multi-agent literature [44, 48, 84] inspired by real-world challenges [35]. The environment consists of multiple sensor agents and moving targets. Each agent can scan at most one nearby area per timestep, and two agents must scan an area simultaneously to track a target. Since we are studying the case with minimal influence of deep learning approximation errors in this section, we use a small sensor network with 2×2 sensor agents (1^{st} column of Fig. 1, with each sensor represented by a circle) and 1 target. There are four possible states, each represented by a one-hot vector indicating the target's true location.

Collectively observable (CO) Dec-POMDPs. Thefi rst row of Fig. 1 illustrates a CO Dec-POMDP [54]. Each agent's observation $o \in \mathbb{R}^2$ includes a separate dimension for each nearby area, with a value 1 if the target is present and 0 otherwise. In this example, the target is in `Area 1`. The right side plots changes in thefi rst two state dimensions during diffusion. Each arrow starts from a possible noisy state, which, together with local history, is the input to a denoiser network. The network outputs a denoised state, marking the endpoint of the arrow. `Agent 3` and 4 have uncertainty because they cannot observe the target in their nearby areas. This uncertainty is reflected in the vectorfi elds. Conditioned on their histories (length=1), there are twofi xed points, each representing a possible state. For example, `Agent 3` cannot distinguish whether the target is in `Area 1` or `Area 4`; correspondingly, there are two attractors $y = (1, 0, 0, 0)$ and $(0, 0, 0, 1)$. Moreover, we observe that thefl ow has an equal probability of converging to these twofi xed points, matching the true posterior state distribution given the history. `Agent 1` and 2 know the true state because they observe the target. Correspondingly, there is only onefi xed point $(1, 0, 0, 0)$ given their history. *More importantly, we note that the only commonfi xed point shared by all agents is the true state* $(1, 0, 0, 0)$.

Non-collectively observable Dec-POMDPs. The second row of Fig. 1 illustrates a non-CO Dec-POMDPs. Agent observations are

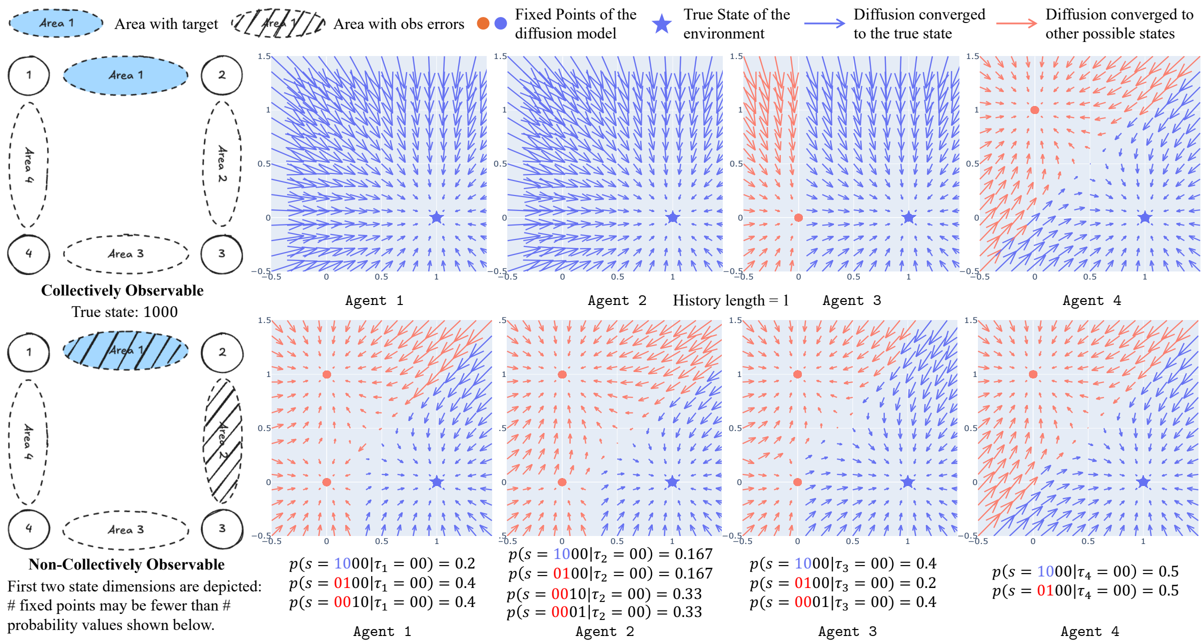


Figure 1: *With minimal deep learning approximation errors, a diffusion process represents states consistent with local history τ_i (length=1 in this figure) as their attractors, provably equivalent to stable fixed points of the denoiser network $f_\theta(\tau_i, \cdot)$.* Arrows point to denoiser network outputs $y' = f_\theta(\tau_i, y)$ from input noisy states y. The first two dimensions of y' and y are shown. Top row: In collectively observable (CO) Dec-POMDPs, a unique fixed point is shared by all agents, which is also the true state. Bottom row: In non-CO Dec-POMDPs, shared fixed points are all states consistent with joint history τ, and diffusion models reproduce the posterior state distribution $p(s|\tau_i)$ under appropriate distributions of input noisy states.

the same, but when the target is in `Area 1` or 2, nearby sensors fail to observe it with 50% probability. For example, when the true state is $(1, 0, 0, 0)$, the observation of `Agent 1` can be either $(1, 0)$ or $(0, 0)$ with equal probability. This environment is non-CO because it is possible for all agents observe $(0, 0)$, in which case even aggregating all local information does not reveal the target's location, as shown in Fig. 1. We observe a significant difference from the first row: there are multiple shared fixed points: $(1, 0, 0, 0)$ and $(0, 1, 0, 0)$. This reflects the best inference the agents can achieve: `Agent 3` and 4 is sure that the target is not in `Area 3` or 4. However, the aggregated local information cannot tell whether the target is in `Area 1` or 2.

3.2 Infer State Locally: Stable Fixed Points

Although simple, the sensor network example encapsulates our findings which will be discussed in this section. Our first finding pertains to how diffusion models represent states, *i.e.*, how each agent infers states based on its own history. We begin by showing that these individual diffusion processes converge.

Theorem 1. *[Converged Diffusion] In the absence of approximation errors, repeatedly applying the denoiser network $f_\theta(\tau_i, y)$ converges to a state s that is consistent with τ_i and has a dominate posterior probability given y: $\phi_\infty(\tau_i, \theta)(y) = s = \arg\max_{s \in S_G(\tau_i)} p(s|y, \tau_i)$.*

Theorem 1 formally underpins the observation in Fig. 1. Based on this, we give the sufficient and necessary conditions of how a diffusion model represents states.

Theorem 2. *[State Representation] The diffusion model represents states consistent with τ_i by the* attractors *of its flow:*

$$\hat{S}_G(\tau_i) = \left\{ y^* \mid [\phi_\infty(\tau_i, \theta)]_* p_0(y^*) > 0 \right\}, \tag{10}$$

which are equivalent to the stable fixed points $\mathcal{F}_\phi(\tau_i)$ of $f_\theta(\tau_i, \cdot)$,

$$\mathcal{F}_\phi(\tau_i) = \{y^* \mid y^* = f_\theta(\tau_i, y^*), |\lambda_{\max}(y^*|\tau_i)| < 1\}. \tag{11}$$

Here, $\lambda_{\max}(y^*|\tau_i)$ is the largest eigenvalue of the Jacobian $J_f(y^*|\tau_i)$.

Proved in Appx. A.2, Theorems 1 and 2 align with recent research [6, 53] showing that diffusion models are able to produce samples from data distributions with bounded support on a low-dimensional data manifold. Since Theorem 2 shows $\mathcal{F}_\phi(\tau_i) = \hat{S}_G(\tau_i)$ always hold, the terms *attractor* and *fixed pointed* will be used interchangeably. In the absence of approximation errors, the stable fixed points are the states consistent with τ_i: $\mathcal{F}_\phi(\tau_i) = S_G(\tau_i)$.

3.3 Infer State Globally: Shared Fixed Points

Theorems 1 and 2 show how an individual diffusion model represents the inference of agent i about states. In the simplest scenario described in the finding below, this local inference suffices to determine the true global state.

Finding 1. If an individual diffusion model has only one stable fixed point y^*, agent i is able to infer that $s = y^*$.

A unique fixed point implies that only one state is consistent with τ_i. Therefore, agent i can unambiguously determine the global state, eliminating the need for communication with others.

We focus primarily on more complex scenarios where agents are uncertain about the state, *i.e.*, individual diffusion models have multiple fixed points. We first show that, this uncertainty can be resolved in collectively observable Dec-POMDPs.

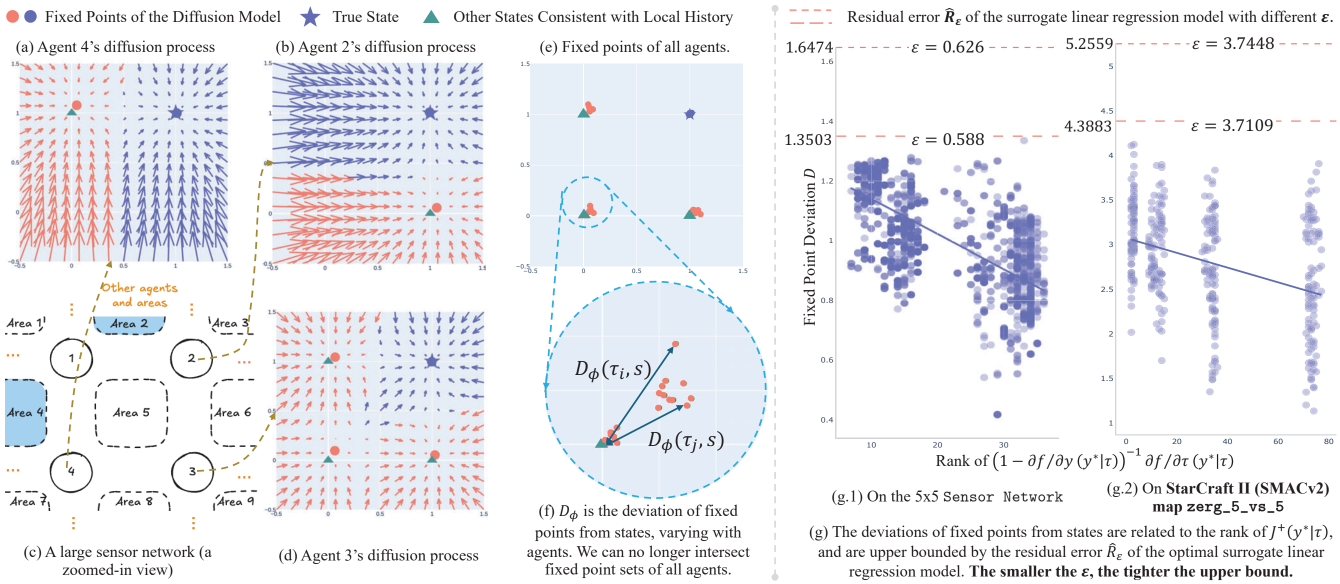


Figure 2: Deep learning approximation errors cause fixed points to deviate from true states. Deviation norms are related to the Jacobian rank and can be upper bounded by a surrogate linear model. (a,b,d) In the 5×5 sensor network (with a zoomed-in view in (c)), we show changes in state dimensions corresponding to `Area2` and 4 during diffusion. (e,f) The impact of these deviations becomes evident when fixed points of all agents are displayed together in a single panel: the true state can no longer be determined by intersecting fixed point sets of all agents, as it is possible that $\cap_i \mathcal{F}_\phi(\tau_i) = \emptyset$. (g) Empirical evidence from SMACv2 and the 5×5 sensor network shows that deviation norms negatively correlate to Jacobian ranks and are tightly upper bounded by optimal residual errors of the surrogate linear model.

Theorem 3. *Without approximation errors, in collectively observable Dec-POMDPs, the intersection of the fixed point sets of all agents is the true state s:* $\cap_i \mathcal{F}_\phi(\tau_i) = \{s\}$.

Conversely, non-collectively observable Dec-POMDPs are more complicated, as agents are collectively unable to uniquely determine the state. However, we can still show that diffusion models identify all states consistent with the joint history and can reproduce the posterior probability of these states.

Theorem 4. *Without approximation errors, in non-collectively observable Dec-POMDPs, the intersection of fixed point sets is all states consistent with joint history:* $\cap_i \mathcal{F}_\phi(\tau_i) = \{s \mid p(s|\boldsymbol{\tau}) > 0\}$. *The true posterior probability* $p(s|\boldsymbol{\tau})$ *can be recovered with appropriate prior distributions* $p(y)$ *of initial noisy states* y.

Proved in Appx. A.3, Theorems 3 and 4 establish a communication protocol–agents share their fixed points when necessary to maximally resolve uncertainty about states. A potential application of these results is to train a denoiser during centralized training and use inferred states or their distributions in decentralized execution, thereby enabling new MARL algorithms (please refer to Appx. C).

4 DEVIATED FIXED POINTS

We now consider the influence of deep learning approximation errors. **(1)** We find that the major impact of these errors is that fixed points deviate from states (Fig. 2(a-f)). **(2)** Theoretically, we identify Jacobian ranks as the primary factor driving this deviation (Theorem 5). Empirical results (Fig. 2(g)) support this finding and pinpoint design choices that influence Jacobian ranks (Fig. 3). **(3)** Inspired by the low-rank property, we construct a surrogate linear regression model (Finding 3) to bound the deviation (Theorem 6). **(4)** This deviation bound helps prove the convergence of a novel composite diffusion process (Sec. 5).

4.1 Example

We begin with a concrete example that expands the sensor network to include 5×5 agents and 2 targets, with each sensor configured to scan 4 nearby areas. Fig. 2(c) zooms in on the section of the map containing two targets. For a fair comparison against the example in the previous section, we keep the network architecture and training agenda unchanged (details in Appx. B).

In Fig. 2(a,b,d), we show the discrete-time flow induced by the diffusion models in this task. The increased problem size puts an extra burden on diffusion models. For example, in Fig. 2(d), the diffusion model of `Agent 3` converges to four fixed points (blue and red circles), which are not strictly overlapped with possible states (blue star and green triangles), indicating that fixed points deviate from clean states. For better visualization, in Fig. 2(f), we put the fixed points of all agents together and zoom in on the fixed points around (0,0). It shows that diffusion models of different agents exhibit distinct fixed points with varied deviations. In this way, *it is no longer practical to calculate the intersection of agents' fixed point sets to obtain the true state, as the intersection would be empty.*

4.2 Jacobian Rank and Surrogate Linear Model

Empirical Observations. To understand why diffusion models can no longer represent the consistent states $S_G(\tau_i)$ accurately, we take a closer look at the fixed points learned by the diffusion models in the 5×5 sensor network (Fig. 2(c)) and the `zerg_5_vs_5` map from the complex, highly stochastic SMACv2 benchmark [14].

Specifically, we look at the Jacobian rank (the rank of $J^+(y|\tau) = (I - \partial f/\partial y(\tau, y))^{-1}\, \partial f/\partial\tau(\tau, y)$) at fixed points of individual agents.

Fig. 2(g) shows the relationship between fixed points' Jacobian ranks and their deviations from states (in ℓ_2 norm). A clear negative correlation is observed between these two variables, *i.e.*, fixed points with lower Jacobian ranks are more likely to exhibit large deviations from true states.

Theoretical Understanding. We now formally analyze the underlying reason for this negative correlation and thereby show why diffusion models might not be able to represent all states accurately. We first define the deviation of fixed points from states as follows.

Definition 4 [Deviation of fixed points from states]**.** We define the error between a true state s and its corresponding fixed point by

$$D_\phi(\tau_i, s) = s - \phi_\infty(\tau_i, \theta)(s). \tag{12}$$

Here, $\phi_\infty(\tau_i, \theta)(s)$ is the attractor to which the diffusion process conditioned on τ_i converges when initialized from state s.

Our analysis begins with the following theorem that characterizes the influence of τ on the fixed points, i.e., how the fixed point changes when τ changes.

Theorem 5. *Let $y^* \in \mathcal{F}_\phi(\tau)$ be a fixed point corresponding to history τ. When τ changes to $\tau' = \tau + \Delta\tau$, the fixed point shifts to $y^{*\prime} = y^* + \Delta y^*$. If the changes in Jacobian satisfies $\|J_f(\tau', y^{*\prime}) - J_f(\tau, y^*)\|_F < \epsilon$ for a small ϵ, we have*

$$\Delta y^* \approx \left(I - \frac{\partial f}{\partial y}(\tau, y^*)\right)^{-1} \frac{\partial f}{\partial \tau}(\tau, y^*)\Delta\tau. \tag{13}$$

The approximation error in Eq. (13) is bounded by $\|\mathtt{Err}(\Delta y^*)\| \le \frac{M\epsilon^2}{2(1-\lambda_{\max})m^2}$, where $\lambda_{\max}$ is the largest eigenvalue of Jacobian $J_f(y^*|\tau_i)$ at y^*, and M, m is the upper/lower bound on the norm of Hessian. Due to its $O(\epsilon^2)$ magnitude, this error is negligible when ϵ is small.

Finding 2. The major takeaway of Theorem 5 emerges from Eq. (13). This equation implies that a diffusion model behaves like a (locally) linear model $\Delta y^* \approx J^+(y^*|\tau)\Delta\tau$ with weights

$$J^+(y^*|\tau) = \left(I - \frac{\partial f}{\partial y}(\tau, y^*)\right)^{-1} \frac{\partial f}{\partial \tau}(\tau, y^*) \tag{14}$$

to approximate the shifts of fixed points when local history changes.

In Corollary 5.1, we expand Eq. (13) in the special case where the denoiser network is a fully-connected network.

Corollary 5.1. *If $f_\theta(\tau, y) = g\left(\sigma\left((W_\tau, W_y)\begin{pmatrix}\tau\\ y\end{pmatrix} + b\right)\right)$ is a fully connected network. $\sigma(\cdot)$ is an element-wise activation function and $g(\cdot)$ represents the subsequent fully connected layers, which may introduce additional non-linearities following the first layer, we have $\Delta y^* \approx U(I-\Lambda)^{-1}\Lambda U^\top W_y^+ W_\tau \Delta\tau$, where $U\Lambda U^\top$ is eigen-decomposition of Jacobian $J_f(y^*|\tau_i)$, W_y^+ is Moore–Penrose inverse $W_y^+ = W_y^\top(W_y W_y^\top)^{-1}$.*

Proved in Appx. A.4, Corollary 5.1 examines a specific network architecture, while the other analyses in this paper apply to any denoiser architecture. Corollary 5.1 assumes J_f is symmetric and non-negative, which is approximately true for learned denoisers [46] and can be proved to hold for the optimal denoiser [29].

Based on Theorem 5, we use proof by contradiction to show that diffusion models might not have enough capacity to represent all states accurately. We start with a local history τ and one of its consistent states s, assuming that the diffusion model conditioned on τ has enough capacity to exactly represent s by a fixed point y^*. We then consider other histories near τ: $\mathcal{T}_\epsilon = \{\tau' \mid \|J_f(\tau', y^{*\prime}) - J_f(\tau, y^*)\|_F \le \epsilon\}$. Due to Finding 2, the diffusion model is trained to (locally) solve the following optimization problem.

Definition 5 [Surrogate Local Linear Regression Model]**.** For a local history τ and a consistent state s, let $\mathcal{M}_\epsilon \subset \mathcal{D}$ be a sample set containing (τ', s') in the training dataset $\mathcal{D}$ satisfying $\|J_f(\tau, y^*) - J_f(\tau', y^{*\prime})\|_F \le \epsilon$. The surrogate linear regression problem is:

$$\mathcal{R}_\epsilon : \quad \underset{W \in \mathbb{R}^{|s|\times|\tau|}}{\arg\min} \sum_{(\tau', s') \in \mathcal{M}_\epsilon} \|W\Delta\tau - \Delta s\|^2, \tag{15}$$

where $\Delta\tau = \tau - \tau'$, $\Delta s = s - s'$. The residual error of the optimal solution to $\mathcal{R}_\epsilon$ is $\hat{R}_\epsilon$. The number of linearly independent s' in $\mathcal{M}_\epsilon$ is $r(\mathcal{M}_\epsilon) = \dim(\mathrm{span}\,\{s' \mid (\tau', s') \in \mathcal{M}_\epsilon\})$.

Intuitively, given $\Delta\tau$, the denoiser network learns $J^+(y^*|\tau)$ to minimize the difference between Δy^* and the groundtruth Δs. The surrogate $\mathcal{R}_\epsilon$ provides the best *linear* solution to this optimization problem. The question is whether the denoiser network, locally, has enough capacity to perform better than this solution.

Finding 3. If the denoiser network f_θ is over-parameterized with the maximum possible rank of $J^+(y^*|\tau)$ (Eq. (14)) larger than the number of linearly independent state samples in the local regression problem $\mathcal{R}_\epsilon$:

$$\mathrm{rank}(J^+(y^*|\tau)) \ge r(\mathcal{M}_\epsilon), \tag{16}$$

then the linear regression problem $\mathcal{R}_\epsilon$ is underdetermined, indicating that the denoiser network has enough capacity to represent the history-state mapping S_G.

On the other hand, if the denoiser network is under-parameterized, leaving $\mathrm{rank}(J^+(y^*|\tau)) < r(\mathcal{M}_\epsilon)$, then we have an over-determined regression problem that inevitably induces residual errors, leading to deviations of fixed points from states.

Evidence 3.1. We empirically verify our findings in the 5×5 sensor network. In Fig. 3-middle, we increase the number of states, so that $r(\mathcal{M}_\epsilon)$ increases. With a fixed network size, we can see that $D_\phi(\tau, s)$ increases. We also find that increasing the network width can increase the rank of $J^+(y^*|\tau)$, and correspondingly decrease D_ϕ as shown in Fig. 3 left. This is not trivial as the input dimension is fixed and is smaller than the network width, which means the rank of $J^+(y^*|\tau)$ is actually upper bounded by the input dimension.

4.3 Bounded Deviations

We now discuss how to bound the deviations D_ϕ.

Finding 4. When ϵ is large, the expressivity of a diffusion model is more powerful than the surrogate regression model $\mathcal{R}_\epsilon$. This is because the denoiser network becomes more non-linear as the Jacobian changes significantly. In this case $\hat{R}_\epsilon$ is larger than D_ϕ, as $\hat{R}_\epsilon$ is the residual of a linear model, while D_ϕ is the residual of a non-linear model. On the other hand, when we decrease ϵ, the denoiser network is approaching linear and its capacity is getting close to the linear regression model, so D_ϕ is getting close to $\hat{R}_\epsilon$.

We formally present this finding in the following theorem.

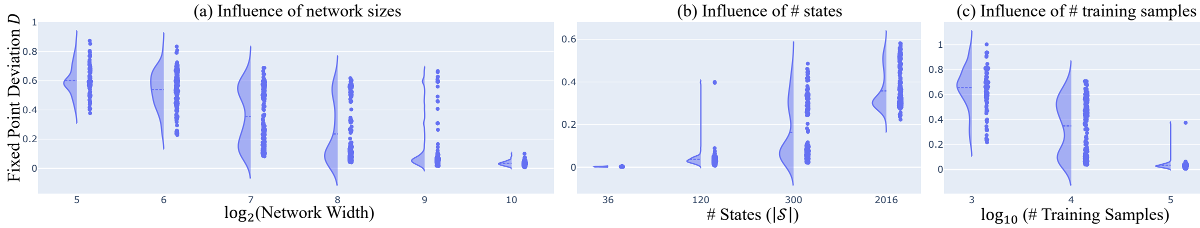


Figure 3: Practical factors contributing to low Jacobian ranks (which correlate negatively with deviations of fixed points from true states) include narrow network architectures, large state space sizes, and small numbers of training samples. In each panel, a blue point represents the deviation of a fixed point, and distributions of these deviations are displayed on the left.

Theorem 6. *[Upper Bounded Deviation] Let $\mathcal{M} = \{(\tau^{(k)}, s^{(k)})\}_{k=1}^{K_1}$. If there exist τ and τ' in $\mathcal{M}$ where $\|J_f(\tau', y^{*\prime}) - J_f(\tau, y^*)\|_F > \epsilon$, then for $(\tau, s) \in \mathcal{M}$, we have*

$$D_\phi(\tau, s) < \hat{R}(\mathcal{M}) = \mathrm{Tr}(\Sigma_s) - \mathrm{Tr}(\Sigma_{s\tau}\Sigma_\tau^{-1}\Sigma_{\tau s}). \quad (17)$$

Here $\Sigma_s = \mathbb{E}_{s\sim\mathcal{M}}[(s - \mathbb{E}[s])(s - \mathbb{E}[s])^\top]$, cross covariance $\Sigma_{s\tau} = \mathbb{E}_{(\tau,s)\sim\mathcal{M}}[(s - \mathbb{E}[s])(\tau - \mathbb{E}[\tau])^\top]$, and $\Sigma_{s\tau} = \Sigma_{\tau s}^\top$. $\hat{R}(\mathcal{M})$ is the residual error of the optimal linear regression model on $\mathcal{M}$.

Proved in Appx. A.5, Theorem 6 upper bounds fixed point deviations by constructing a surrogate regression problem. By decreasing the ϵ, we can tighten this upper bound, which helps prove the convergence of our composite diffusion process in the next section.

Evidence 4.1. We provide empirical evidence to support Finding 4 and Theorem 6. The experiments are conducted on the 5×5 sensor network (Fig. 2 (g.1)) and the `zerg_5_vs_5` map (Fig. 2 (g.2)) from the complex, highly stochastic SMACv2 benchmark [14]. In (g.1), two dashed horizontal lines show the residual errors $\hat{R}_\epsilon$ of the surrogate linear regression model under two different ϵ values. When $\epsilon = 0.626$, the linear regression model exhibits weaker representational capacity, with $\hat{R}_{0.626} = 1.6474$ significantly larger than the deviations of the fixed points. By contrast, when we decrease ϵ to 0.588, $\hat{R}_{0.588}$ provides a tight upper bound for the deviations. The case in SMACv2 is similar, *indicating that surrogate models effectively approximate local behavior of diffusion models.*

5 COMPOSITE DIFFUSION

As we discussed in Sec. 4, when fixed points deviate from clean states, it is impractical to obtain true states by intersecting all agents' fixed points. Instead, we propose to use *composite diffusion*.

Definition 6 [Composite diffusion]. Let $[i_1, i_2, \cdots, i_n] \in \mathcal{P}([n])$ be a permutation of n agents. The composite diffusion conditioned on $\tau_{i_{1:n}}$ iteratively applies individual denoiser models based on each agent's history: $f_\theta(\tau_{i_{1:n}}, y) = f_\theta(\tau_{i_1}, f_\theta(\tau_{i_2}, \cdots f_\theta(\tau_{i_n}, y)))$, inducing a *discrete-time composite flow* $\phi_\ell(\tau_{i_{1:n}}, \theta) : \mathbb{N} \times \mathbb{R}^{|s|} \to \mathbb{R}^{|s|}$, $\phi_\ell(\tau_{i_{1:n}}, \theta)(y) = f(\tau_{i_{(\ell)_n}}, \phi_{\ell-1}(\tau_{i_{1:n}}, \theta)(y))$, where $(\ell)_n = (\ell \bmod n)$, and $\phi_0(\tau_{i_{1:n}}, \theta)(y) = y$.

We then use composite diffusion to estimate true states.

5.1 Composite Diffusion Yields True States

Composite diffusion algorithm. Our algorithm has two steps. Step 1 (Composite denoising): Sample a set of Gaussian noise $\{y^{(k,0)}\}_{k=0}^{K_2}$ from $\mathcal{N}(0, \sigma^2 I)$. Apply composite diffusion to each $y^{(k,0)}$, resulting in a sequence of denoised states $(y^{(k,1)}, \cdots, y^{(k,L)})$.

Step 2 (Condition check): Check whether the following conditions hold. (1) For the last n elements in the sequence, $y^{(k,\ell)}$, $L - n < \ell \le L$, the largest eigenvalue of the denoiser Jacobian at $y^{(k,\ell)}$ satisfies $|\lambda_{\max}| < 1$. (2) Apply the individual denoiser conditioned on each agent's history repeatedly to the last element $y^{(k,L)}$ until converge. The change in $y^{(k,L)}$ should be smaller than $2D_\phi$ (bounded by Theorem 6). These condition checks guarantee convergence to the true state as proven in the following theorem and corollary.

Theorem 7. *[Composite Diffusion Approaches True State] In collectively observable Dec-POMDPs, for a sequence k satisfying the conditions in Step 2, the composite diffusion algorithm approaches the true state s with an error bound $\max_{1\le i\le n} D_\phi(\tau_i, s)$. Specifically, it converges to the convex hull of the agents' fixed points near s.*

Corollary 7.1. *In non-collectively observable Dec-POMDPs, with sufficiently enough initial samples, the composite diffusion algorithm approaches every possible global state s consistent with joint history $\tau_{i_{1:n}}$ with an upper error bound $\max_{1\le i\le n} D_\phi(\tau_i, s)$.*

Theorem 7 and Corollary 7.1 are proved in Appx. A.7. They highlight the advantages of composite diffusion summarized as follows.

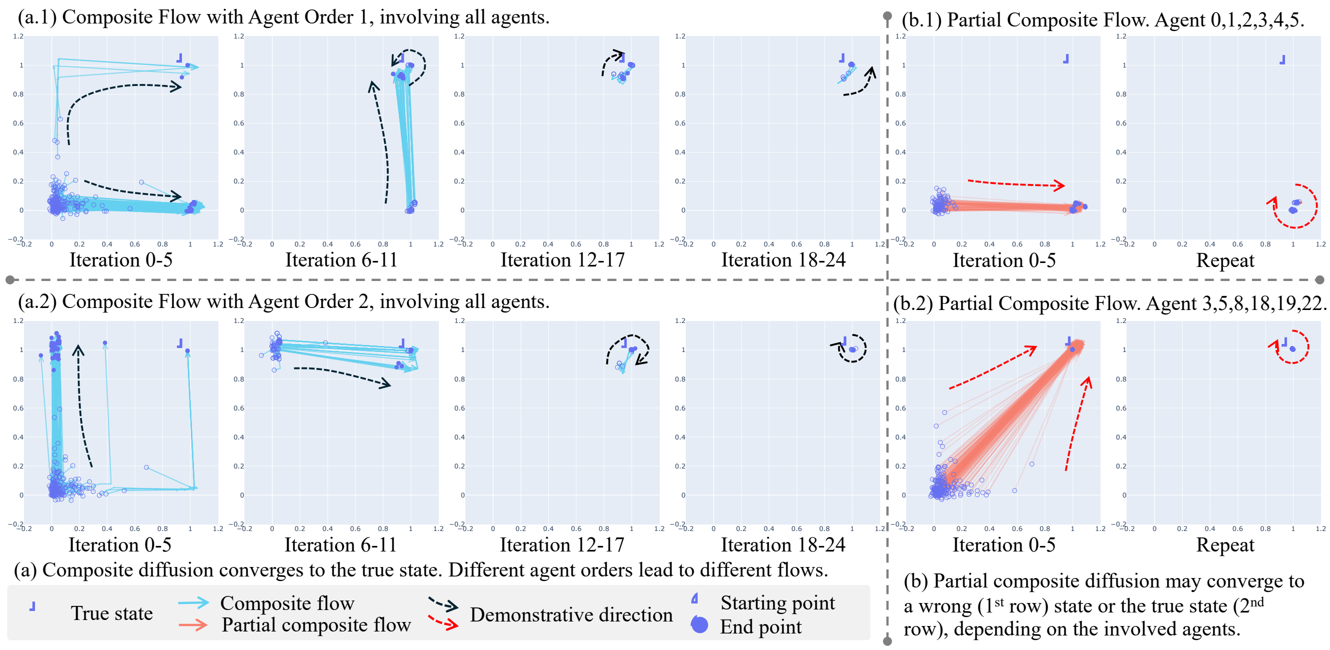
Finding 5. Unlike individual diffusion conditioned on the history of a single agent, composite diffusion can resolve the uncertainty when there are multiple fixed points. Even in the simple case with only one fixed point, composite diffusion can better estimate the true state, as proved in the following theorem (details in Appx. A.8).

Theorem 8. *When there is only one fixed point for the diffusion model conditioned on τ_i, $i \in [n]$, assume that the final element $y^{(L)}$ distributes uniformly, composite diffusion provides a more accurate global state estimation than individual diffusion:*

$$\mathbb{E}_{y^{(L)}}[\|y^{(L)} - s\|] \le \frac{1}{|\mathcal{F}_s|}\sum_{y_i^* \in \mathcal{F}_s} \|y_i^* - s\|. \quad (18)$$

Evidence 5.1. We evaluate the accuracy of composite diffusion (measured in peak signal-to-noise ratio, PSNR, [76]) against individual diffusion on SMACv2 [14]. Training data is collected by running MAPPO [83] (see Appx. B). For a fair comparison, both methods use the same number of denoising steps. Composite diffusion achieves higher PSNR (indicating lower errors) across all test cases, which examine various factors that can influence diffusion processes. This gap is provable (Theorem 8) even when individual diffusion adopts more powerful network architectures like in [81].

When agents share common information (*e.g.*, a portion of a state visible to all agents [49]), applying composite diffusion only to private information can reduce its overhead. We next discuss overhead reduction in general cases.



(a) Composite diffusion converges to the true state. Different agent orders lead to different flows.

(b) Partial composite diffusion may converge to a wrong (1st row) state or the true state (2nd row), depending on the involved agents.

Figure 4: Evolution of denoised state distributions (first two dimensions) during composite diffusion processes, initialized with various noisy states, in the 5×5 sensor network. Each panel shows the changes (from open circles to closed circles) over 6 denoising iterations, with each iteration conditioned on the history of a single agent, *e.g.*, in iteration 0-5, six agents in the corresponding order are involved. (a) Composite diffusion converges to the true state regardless of the agent ordering. (b) Partial composite diffusion may converge to incorrect states depending on the participating agents.

Table 1: True state estimation errors measured in PSNR. Higher PSNR values indicate lower errors. Individual diffusion exhibits a provable performance gap compared to composite diffusion.

Setting	Alg.	Network Width			# Training Samples			Obs. History Length			Obs. Sight Range		Tasks		
		1024	4096	8192	10	100K	500K	1	5	10	5	9	Zerg	Terran	Protoss
Train	Individual	13.28	18.71	28.20	56.45	28.20	35.29	26.39	28.20	31.86	26.29	28.20	28.20	29.01	27.79
	Composite	**15.98**	**22.18**	**30.77**	**56.77**	**30.77**	**36.39**	**27.88**	**30.77**	**33.13**	**27.92**	**30.77**	**30.77**	**30.68**	**30.20**
Test	Individual	13.37	15.37	17.42	11.32	17.43	23.22	20.77	17.43	16.54	15.94	17.43	17.42	17.28	17.24
	Composite	**16.07**	**18.49**	**20.40**	**12.38**	**20.40**	**25.49**	**23.54**	**20.40**	**18.94**	**17.90**	**20.40**	**20.40**	**18.59**	**20.08**

5.2 Partial Composite Diffusion

Composite diffusion requires a communication chain in which each agent receives the output of the preceding agent's denoiser network and sends its own denoising output to the subsequent agent. These messages reside in $\mathbb{R}^{|s|}$. In very large systems, it is possible to trade off this communication overhead against state estimation accuracy by involving only a subset of agents in composite diffusion. We thereby define *partial* composite diffusion $f_\theta(\tau_{i_{1:k}}, \cdot)$ and *partial* compositefl ow $\phi_\ell(\tau_{i_{1:k}}, \theta)$, in which $k < n$ and $[i_1, i_2, \cdots, i_k] \in \mathbb{P}([k])$ is a permutation of the considered k agents. We analyze its convergence property in Corollary 7.2.

Corollary 7.2. *With sufficient initial samples, the partial composite diffusion algorithm approaches every possible global state s consistent with* $\tau_{i_{1:k}}, k < n$*, with an upper error bound* $\max_{1\leq t\leq k} D_\phi(\tau_{i_t}, s)$.

Finding 6. Composite diffusion $f_\theta(\tau_{i_{1:n}}, y)$ converges to the true global state, while partial composite diffusion $f_\theta(\tau_{i_{1:k}}, y), k < n$ may converge to wrong states, depending on the participating agents.

Evidence 6.1. Fig. 4 shows the evolution of denoised state distributions (focusing on thefi rst two dimensions) during (partial) composite diffusion processes in the 5×5 sensor network. Initial states are sampled from $\mathcal{N}(0, I)$. Compositefl ows reliably discover the true state regardless of the agent ordering. In contrast, a partial compositefl ow stabilizes at a state consistent with participants' histories, with the accuracy depending on the participating agents.

6 CLOSING REMARKS

This paper provides thefi rst rigorous understanding of how deep learning models and their approximation errors can impact agents' handling of PO in Dec-POMDPs. We expect that this work can establish a general framework for addressing the challenges posed by PO across various multi-agent sub-fields. As an initial demonstration of these possibilities, Appx. C provides an example where integrating diffusion models with policy learning can further enhance the performance of multi-agent RL algorithms, such as MAPPO [83].

ACKNOWLEDGMENTS

We sincerely thank Stefano Ermon, Dheeraj Nagaraj, Lingkai Kong, and the anonymous reviewers of AAMAS 2025 for their valuable inputs, which significantly helped improve the quality of this paper.

REFERENCES

[1] Suzan Ece Ada, Erhan Oztop, and Emre Ugur. 2024. Diffusion policies for out-of-distribution generalization in offline reinforcement learning. *IEEE Robotics and Automation Letters* (2024).

[2] Christopher Amato, Girish Chowdhary, Alborz Geramifard, N Kemal Üre, and Mykel J Kochenderfer. 2013. Decentralized control of partially observable Markov decision processes. In *52nd IEEE Conference on Decision and Control*. IEEE, 2398–2405.

[3] Daniel S Bernstein, Robert Givan, Neil Immerman, and Shlomo Zilberstein. 2002. The complexity of decentralized control of Markov decision processes. *Mathematics of operations research* 27, 4 (2002), 819–840.

[4] Johann Brehmer, Joey Bose, Pim De Haan, and Taco S Cohen. 2024. EDGI: Equivariant diffusion for planning with embodied agents. *Advances in Neural Information Processing Systems* 36 (2024).

[5] Duygu Ceylan, Chun-Hao P Huang, and Niloy J Mitra. 2023. Pix2video: Video editing using image diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 23206–23217.

[6] Sitan Chen, Sinho Chewi, Jerry Li, Yuanzhi Li, Adil Salim, and Anru Zhang. 2023. Sampling is as easy as learning the score: theory for diffusion models with minimal data assumptions. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=zyLVMgsZ0U_

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*. PMLR, 1597–1607.

[8] Zoey Chen, Sho Kiami, Abhishek Gupta, and Vikash Kumar. 2023. Genaug: Retargeting behaviors to unseen situations via generative augmentation. *arXiv preprint arXiv:2302.06671* (2023).

[9] Cheng Chi, Zhenjia Xu, Siyuan Feng, Eric Cousineau, Yilun Du, Benjamin Burchfiel, Russ Tedrake, and Shuran Song. 2023. Diffusion policy: Visuomotor policy learning via action diffusion. *The International Journal of Robotics Research* (2023), 02783649241273668.

[10] Maria Carmela De Gennaro and Ali Jadbabaie. 2006. Decentralized control of connectivity for multi-agent systems. In *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE, 3628–3633.

[11] Carl Doersch. 2016. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908* (2016).

[12] Heng Dong, Tonghan Wang, Jiayuan Liu, and Chongjie Zhang. 2022. Low-rank modular reinforcement learning via muscle synergy. *Advances in Neural Information Processing Systems* 35 (2022), 19861–19873.

[13] Bradley Efron. 2011. Tweedie's formula and selection bias. *J. Amer. Statist. Assoc.* 106, 496 (2011), 1602–1614.

[14] Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob Nicolaus Foerster, and Shimon Whiteson. 2023. SMACv2: An Improved Benchmark for Cooperative Multi-Agent Reinforcement Learning. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. https://openreview.net/forum?id=5OjLGiJW3u

[15] Patrick Esser, Sumith Kulal, Andreas Blattmann, Rahim Entezari, Jonas Müller, Harry Saini, Yam Levi, Dominik Lorenz, Axel Sauer, Frederic Boesel, et al. 2024. Scaling rectifiedfl ow transformers for high-resolution image synthesis. In *Forty-first International Conference on Machine Learning*.

[16] Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems*. 2137–2145.

[17] Claudia V Goldman and Shlomo Zilberstein. 2003. Optimizing information exchange in cooperative multi-agent systems. In *Proceedings of the second international joint conference on Autonomous agents and multiagent systems*. 137–144.

[18] Yuwei Guo, Ceyuan Yang, Anyi Rao, Zhengyang Liang, Yaohui Wang, Yu Qiao, Maneesh Agrawala, Dahua Lin, and Bo Dai. 2024. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. In *The Twelfth International Conference on Learning Representations*.

[19] Dongqi Han, Kenji Doya, and Jun Tani. 2019. Variational recurrent models for solving partially observable control tasks. *arXiv preprint arXiv:1912.10703* (2019).

[20] Yanlin Han and Piotr Gmytrasiewicz. 2018. Learning others' intentional models in multi-agent settings using interactive POMDPs. *Advances in Neural Information Processing Systems* 31 (2018).

[21] Philippe Hansen-Estruch, Ilya Kostrikov, Michael Janner, Jakub Grudzien Kuba, and Sergey Levine. 2023. Idql: Implicit q-learning as an actor-critic method with diffusion policies. *arXiv preprint arXiv:2304.10573* (2023).

[22] Matthew Hausknecht and Peter Stone. 2015. Deep recurrent q-learning for partially observable mdps. In *2015 aaai fall symposium series*.

[23] Jonathan Ho, William Chan, Chitwan Saharia, Jay Whang, Ruiqi Gao, Alexey Gritsenko, Diederik P Kingma, Ben Poole, Mohammad Norouzi, David J Fleet, et al. 2022. Imagen video: High definition video generation with diffusion models. *arXiv preprint arXiv:2210.02303* (2022).

[24] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems* 33 (2020), 6840–6851.

[25] Jonathan Ho, Tim Salimans, Alexey Gritsenko, William Chan, Mohammad Norouzi, and David J Fleet. 2022. Video diffusion models. *Advances in Neural Information Processing Systems* 35 (2022), 8633–8646.

[26] Michael Janner, Yilun Du, Joshua B Tenenbaum, and Sergey Levine. 2022. Planning with diffusion forfl exible behavior synthesis. *arXiv preprint arXiv:2205.09991* (2022).

[27] Jiechuan Jiang, Chen Dun, Tiejun Huang, and Zongqing Lu. 2018. Graph convolutional reinforcement learning. *arXiv preprint arXiv:1810.09202* (2018).

[28] Jiechuan Jiang and Zongqing Lu. 2018. Learning attentional communication for multi-agent cooperation. In *Advances in Neural Information Processing Systems*. 7254–7264.

[29] Zahra Kadkhodaie, Florentin Guth, Eero P Simoncelli, and Stéphane Mallat. 2023. Generalization in diffusion models arises from geometry-adaptive harmonic representation. *arXiv preprint arXiv:2310.02557* (2023).

[30] Zahra Kadkhodaie and Eero P Simoncelli. 2020. Solving linear inverse problems using the prior implicit in a denoiser. *arXiv preprint arXiv:2007.13640* (2020).

[31] Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. 1998. Planning and acting in partially observable stochastic domains. *Artificial intelligence* 101, 1-2 (1998), 99–134.

[32] Yipeng Kang, Tonghan Wang, Qianlan Yang, Xiaoran Wu, and Chongjie Zhang. 2022. Non-linear coordination graphs. *Advances in Neural Information Processing Systems* 35 (2022), 25655–25666.

[33] Hsu Kao and Vijay Subramanian. 2022. Common information based approximate state representations in multi-agent reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*. PMLR, 6947–6967.

[34] Woojun Kim, Jongeui Park, and Youngchul Sung. 2020. Communication in multi-agent reinforcement learning: Intention sharing. In *International conference on learning representations*.

[35] Victor Lesser, Charles L Ortiz, and Milind Tambe. 2003. *Distributed sensor networks: A multiagent perspective*. Vol. 9. Springer Science & Business Media.

[36] Zhixuan Liang, Yao Mu, Mingyu Ding, Fei Ni, Masayoshi Tomizuka, and Ping Luo. 2023. Adaptdiffuser: Diffusion models as adaptive self-evolving planners. *arXiv preprint arXiv:2302.01877* (2023).

[37] Yaron Lipman, Ricky TQ Chen, Heli Ben-Hamu, Maximilian Nickel, and Matthew Le. 2022. Flow Matching for Generative Modeling. In *The Eleventh International Conference on Learning Representations*.

[38] Xingchao Liu, Chengyue Gong, et al. 2023. Flow Straight and Fast: Learning to Generate and Transfer Data with Rectified Flow. In *The Eleventh International Conference on Learning Representations*.

[39] John Loch and Satinder Singh. 1998. Using Eligibility Traces to Find the Best Memoryless Policy in Partially Observable Markov Decision Processes.. In *ICML*, Vol. 98. Citeseer, 323–331.

[40] Aaron Lou, Chenlin Meng, and Stefano Ermon. 2024. Discrete Diffusion Modeling by Estimating the Ratios of the Data Distribution. In *Forty-first International Conference on Machine Learning*.

[41] Cong Lu, Philip Ball, Yee Whye Teh, and Jack Parker-Holder. 2024. Synthetic experience replay. *Advances in Neural Information Processing Systems* 36 (2024).

[42] Liam C MacDermed and Charles L Isbell. 2013. Point based value iteration with optimal belief compression for Dec-POMDPs. *Advances in neural information processing systems* 26 (2013).

[43] Koichi Miyasawa et al. 1961. An empirical Bayes estimator of the mean of a normal population. *Bull. Inst. Internat. Statist* 38, 181-188 (1961), 1–2.

[44] Pragnesh Jay Modi, Hyuckchul Jung, Milind Tambe, Wei-Min Shen, and Shriniwas Kulkarni. 2001. A dynamic distributed constraint satisfaction approach to resource allocation. In *Principles and Practice of Constraint Programming—CP 2001: 7th International Conference, CP 2001 Paphos, Cyprus, November 26–December 1, 2001 Proceedings 7*. Springer, 685–700.

[45] Sreyas Mohan, Zahra Kadkhodaie, Eero P Simoncelli, and Carlos Fernandez-Granda. 2019. Robust and interpretable blind image denoising via bias-free convolutional neural networks. *arXiv preprint arXiv:1906.05478* (2019).

[46] Sreyas Mohan, Zahra Kadkhodaie, Eero P Simoncelli, and Carlos Fernandez-Granda. 2020. Robust And Interpretable Blind Image Denoising Via Bias-Free Convolutional Neural Networks. In *International Conference on Learning Representations*.

[47] Darius Muglich, Luisa M Zintgraf, Christian A Schroeder De Witt, Shimon Whiteson, and Jakob Foerster. 2022. Generalized beliefs for cooperative AI. In *International Conference on Machine Learning*. PMLR, 16062–16082.

[48] Ranjit Nair, Pradeep Varakantham, Milind Tambe, and Makoto Yokoo. 2005. Networked distributed POMDPs: A synthesis of distributed constraint optimization and POMDPs. In *AAAI*, Vol. 5. 133–139.

[49] Ashutosh Nayyar, Aditya Mahajan, and Demosthenis Teneketzis. 2013. Decentralized stochastic control with partial history sharing: A common information approach. *IEEE Trans. Automat. Control* 58, 7 (2013), 1644–1658.

[50] Frans A Oliehoek, Christopher Amato, et al. 2016. *A concise introduction to decentralized POMDPs*. Vol. 1. Springer.

[51] Shayegan Omidshafiei, Jason Pazis, Christopher Amato, Jonathan P How, and John Vian. 2017. Deep decentralized multi-task multi-agent reinforcement learning under partial observability. In *International Conference on Machine Learning*. PMLR, 2681–2690.

[52] Peng Peng, Ying Wen, Yaodong Yang, Quan Yuan, Zhenkun Tang, Haitao Long, and Jun Wang. 2017. Multiagent bidirectionally-coordinated nets: Emergence of human-level coordination in learning to play starcraft combat games. *arXiv preprint arXiv:1703.10069* (2017).

[53] Jakiw Pidstrigach. 2022. Score-based generative models detect manifolds. *Advances in Neural Information Processing Systems* 35 (2022), 35852–35865.

[54] David V Pynadath and Milind Tambe. 2002. The communicative multiagent team decision problem: Analyzing teamwork theories and models. *Journal of artificial intelligence research* 16 (2002), 389–423.

[55] Chao Qu, Hui Li, Chang Liu, Junwu Xiong, Wei Chu, Weiqiang Wang, Yuan Qi, Le Song, et al. 2020. Intention propagation for multi-agent reinforcement learning. (2020).

[56] Roberta Raileanu, Emily Denton, Arthur Szlam, and Rob Fergus. 2018. Modeling others using oneself in multi-agent reinforcement learning. In *International conference on machine learning*. PMLR, 4257–4266.

[57] Martin Raphan and Eero P Simoncelli. 2011. Least squares estimation without priors or supervision. *Neural computation* 23, 2 (2011), 374–420.

[58] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2018. QMIX: Monotonic Value Function Factorisation for Deep Multi-Agent Reinforcement Learning. In *International Conference on Machine Learning*. 4292–4301.

[59] Herbert E Robbins. 1992. An empirical Bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 388–394.

[60] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 10684–10695.

[61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. Springer, 234–241.

[62] Naci Saldi, Tamer Başar, and Maxim Raginsky. 2019. Approximate Nash equilibria in partially observed stochastic games with mean-field interactions. *Mathematics of Operations Research* 44, 3 (2019), 1006–1033.

[63] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philiph H. S. Torr, Jakob Foerster, and Shimon Whiteson. 2019. The StarCraft Multi-Agent Challenge. *CoRR* abs/1902.04043 (2019).

[64] Yang Song and Stefano Ermon. 2019. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems* 32 (2019).

[65] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. 2021. Score-Based Generative Modeling through Stochastic Differential Equations. In *International Conference on Learning Representations*.

[66] Matthijs TJ Spaan and Nikos Vlassis. 2005. Perseus: Randomized point-based value iteration for POMDPs. *Journal of artificial intelligence research* 24 (2005), 195–220.

[67] Sriram Srinivasan, Marc Lanctot, Vinicius Zambaldi, Julien Pérolat, Karl Tuyls, Rémi Munos, and Michael Bowling. 2018. Actor-critic policy optimization in partially observable multiagent environments. *Advances in neural information processing systems* 31 (2018).

[68] Sainbayar Sukhbaatar, Rob Fergus, et al. 2016. Learning multiagent communication with backpropagation. *Advances in neural information processing systems* 29 (2016).

[69] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. 2018. Value-decomposition networks for cooperative multi-agent learning based on team reward. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 2085–2087.

[70] Pradeep Varakantham, Ranjit Nair, Milind Tambe, and Makoto Yokoo. 2006. Winning back the cup for distributed POMDPs: planning over continuous belief spaces. In *Proceedings of thefi fth international joint conference on Autonomous agents and multiagent systems*. 289–296.

[71] Rose E Wang, Michael Everett, and Jonathan P How. 2020. R-MADDPG for partially observable environments and limited communication. *arXiv preprint arXiv:2002.06684* (2020).

[72] Tonghan Wang, Heng Dong, Victor Lesser, and Chongjie Zhang. 2020. ROMA: Multi-Agent Reinforcement Learning with Emergent Roles. In *Proceedings of the 37th International Conference on Machine Learning*.

[73] Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, and Chongjie Zhang. 2020. RODE: Learning Roles to Decompose Multi-Agent Tasks. In *International Conference on Learning Representations*.

[74] Tonghan Wang, Jianhao Wang, Chongyi Zheng, and Chongjie Zhang. 2019. Learning Nearly Decomposable Value Functions Via Communication Minimization. In *International Conference on Learning Representations*.

[75] Yihan Wang, Beining Han, Tonghan Wang, Heng Dong, and Chongjie Zhang. 2020. DOP: Off-policy multi-agent decomposed policy gradients. In *International conference on learning representations*.

[76] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* 13, 4 (2004), 600–612.

[77] Muning Wen, Jakub Kuba, Runji Lin, Weinan Zhang, Ying Wen, Jun Wang, and Yaodong Yang. 2022. Multi-agent reinforcement learning is a sequence modeling problem. *Advances in Neural Information Processing Systems* 35 (2022), 16509–16521.

[78] Xiaoran Wu, Zien Huang, and Chonghan Yu. 2024. Animating the Past: Reconstruct Trilobite via Video Generation. (2024). https://doi.org/10.12074/202410.00084

[79] Jing Xu, Fangwei Zhong, and Yizhou Wang. 2020. Learning multi-agent coordination for enhancing target coverage in directional sensor networks. *Advances in Neural Information Processing Systems* 33 (2020), 10053–10064.

[80] Zhiwei Xu, Yunpeng Bai, Dapeng Li, Bin Zhang, and Guoliang Fan. 2021. Side: State inference for partially observable cooperative multi-agent reinforcement learning. *arXiv preprint arXiv:2105.06228* (2021).

[81] Zhiwei Xu, Hangyu Mao, Nianmin Zhang, Xin Xin, Pengjie Ren, Dapeng Li, Bin Zhang, Guoliang Fan, Zhumin Chen, Changwei Wang, et al. 2024. Beyond Local Views: Global State Inference with Diffusion Models for Cooperative Multi-Agent Reinforcement Learning. *arXiv preprint arXiv:2408.09501* (2024).

[82] Peng Yang, Randy A Freeman, and Kevin M Lynch. 2008. Multi-agent coordination by decentralized estimation and control. *IEEE Trans. Automat. Control* 53, 11 (2008), 2480–2496.

[83] Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. 2022. The surprising effectiveness of ppo in cooperative multi-agent games. *Advances in Neural Information Processing Systems* 35 (2022), 24611–24624.

[84] Chongjie Zhang and Victor Lesser. 2011. Coordinated multi-agent reinforcement learning in networked distributed POMDPs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 25. 764–770.

[85] Haifeng Zhang, Weizhe Chen, Zeren Huang, Minne Li, Yaodong Yang, Weinan Zhang, and Jun Wang. 2020. Bi-level actor-critic for multi-agent coordination. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 7325–7332.

[86] Kaiqing Zhang, Zhuoran Yang, and Tamer Başar. 2021. Multi-agent reinforcement learning: A selective overview of theories and algorithms. *Handbook of reinforcement learning and control* (2021), 321–384.

[87] Weinan Zhang, Xihuai Wang, Jian Shen, and Ming Zhou. 2021. Model-based multi-agent policy optimization with adaptive opponent-wise rollouts. *arXiv preprint arXiv:2105.03363* (2021).