NEURAL INFORMATION PROCESSING SYSTEMS

Revisiting Multi-Agent World Modeling from a Diffusion-Inspired Perspective

Yang Zhang (Speaker)
Tsinghua University

Chongjie Zhang
Washington University in St. Louis

Chenjia Bai
Institute of AI, China Telecom

Background

- Two primary challenges in modeling Multi-Agent dynamics
  - Exponentially expanded joint action space
  - Additional inter-dependencies among agents (more complex than single-agent scenario)
- Current Multi-Agent world model dealt with these via:
  - Centralized modeling, but suffer from heavy computational cost
  - Decentralized modeling with CTDE principle (currently predominant)

Is it all we can do to deal with Multi-Agent dynamics ?

Background

- Potential limitations from Decentralized modeling with CTDE

1. Mismatch on the transition function estimation

individual modeling + additional communication modules

v.s.

the global state transition in Dec-POMDP or global MDP

2. No supervision signal for the aggregated feature (may hinder training)
3. Lack of efficient utilization of global state in modeling (least important point:)

Motivation

- Therefore,

"Can we develop a centralized modeling scheme that maintains consistency, while keeping computational complexity manageable? "

- Core insight lies in: *Uncertainty about the next global state progressively decreases as individual agent actions are revealed*.
- This observation mirrors the reverse process in diffusion models.

↓

😊 We can realize it via the diffusion process formulation.

Re-formulation

Assumption 1 (Diffusion-Inspired Decomposition of Multi-Agent Dynamics). *In our diffusion-inspired formulation with the descending order of agent id* $(n, n-1, \ldots, 1)$ *as the conditioning order, the global state transition* $P(s_{t+1}|s_t, a_t^{1:n})$ *yields the next state in a manner akin to a typical reverse diffusion process, i.e., satisfying*

$$P(s_{t+1}, s_{t+1}^{(1):(n)}|s_t, a_t^{1:n}) = p(s_{t+1}^{(n)}) \prod_{k=1}^{n} p(s_{t+1}^{(k-1)}|s_{t+1}^{(k)}, a_t^k, s_t), \tag{6}$$

where $s_{t+1}^{(n)}$ *is corrupted with the noise of maximum level* σ_n*, practically indistinguishable from pure Gaussian noise.*

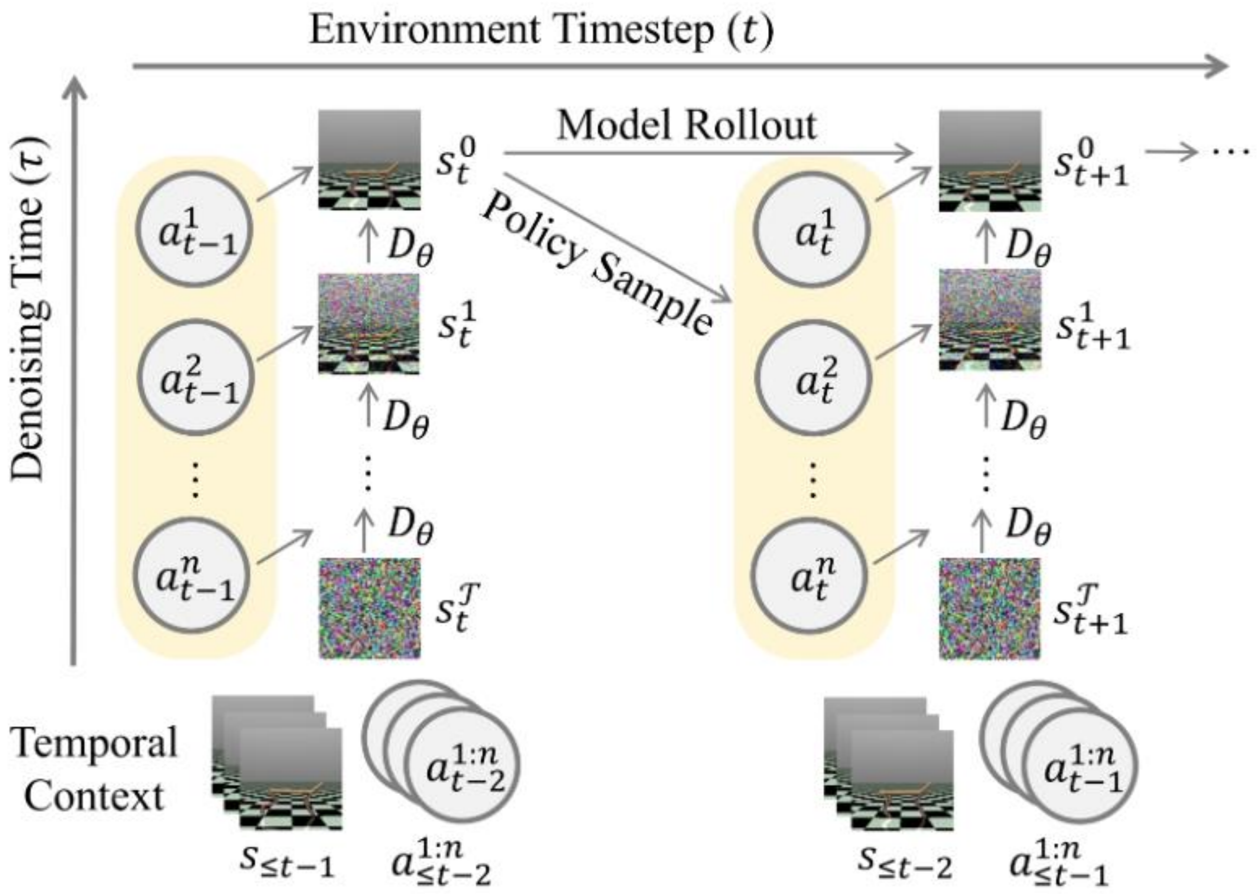
Theorem 2 (ELBO under the Diffusion-Inspired Formulation). *Under Assumption* 1*, the log-likelihood of the multi-agent global state transition (i.e., the evidence of the transition) is lower bounded as follows,*

$$\log P(s_{t+1}|s_t, a_t^{1:n}) \geq \underbrace{\mathbb{E}_{q(s_{t+1}^{(1)}|s_{t+1}^{(0)})}[\log p(s_{t+1}^{(0)}|s_{t+1}^{(1)}, a_t^1, s_t)]}_{\text{reconstruction term}} - \underbrace{\mathrm{D}_{\mathrm{KL}}(q(s_{t+1}^{(n)}|s_{t+1}^{(0)})\|p(s_{t+1}^{(n)}))}_{\text{prior matching term}}$$

$$- \sum_{k=2}^{n} \underbrace{\mathbb{E}_{q(s_{t+1}^{(k)}|s_{t+1}^{(0)})}\left[\mathrm{D}_{\mathrm{KL}}(q(s_{t+1}^{(k-1)}|s_{t+1}^{(k)}, s_{t+1}^{(0)})\|p(s_{t+1}^{(k-1)}|s_{t+1}^{(k)}, a_t^k, s_t))\right]}_{\text{denoising matching term}}. \tag{7}$$

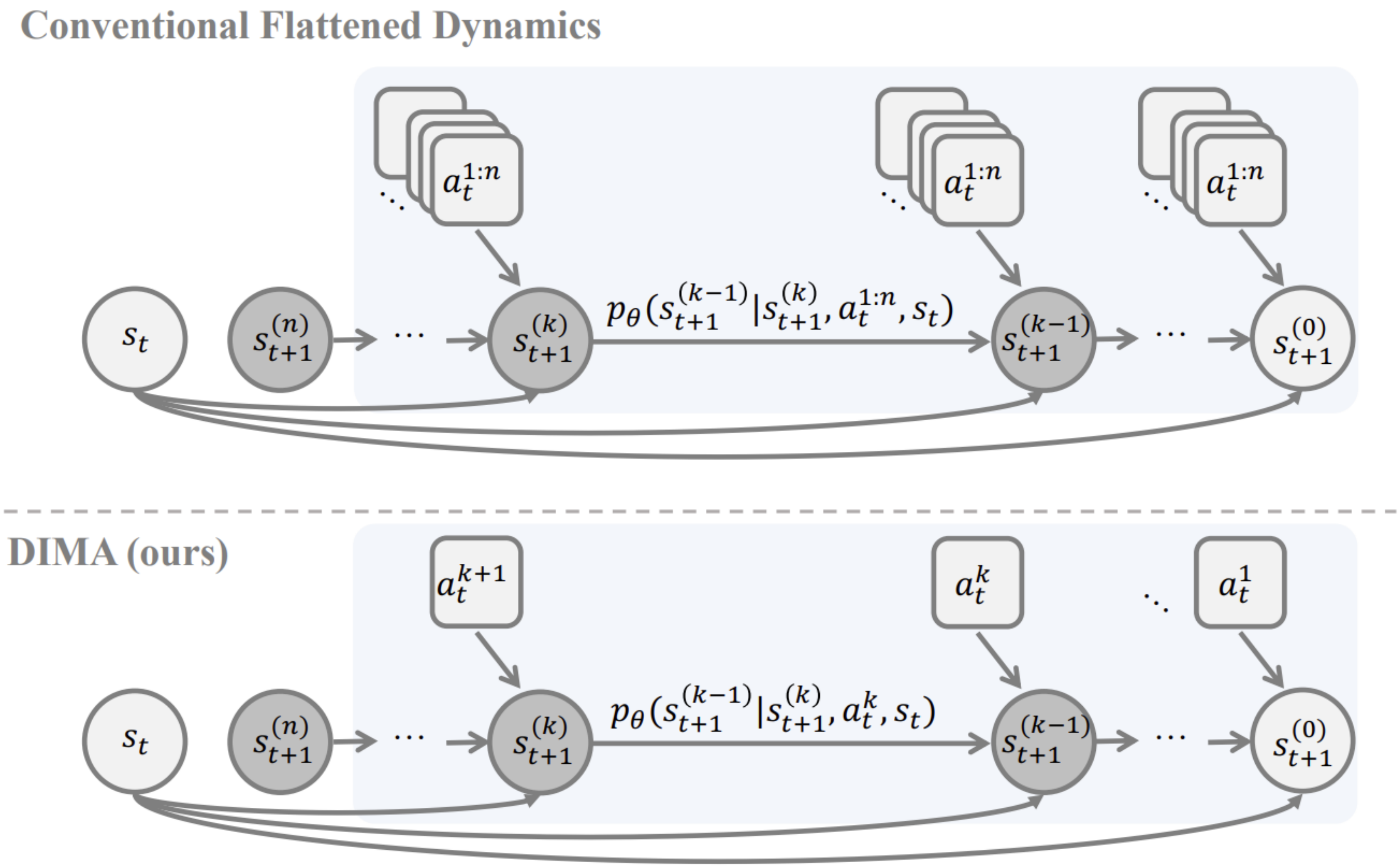
Practical Implementation

- Key implementation details:
  - **Permutation Invariance**: expectation over all possible agent orderings, making the model robust to arbitrary agent ordering.
  - **Condition-Independent Noising Process**: free to choose any noise levels for our sequential formulation.



$$\mathcal{L}(\theta) = \mathbb{E}_{\{\sigma_1,\dots,\sigma_n\}\sim\sigma(\tau)}\mathbb{E}_{\rho\sim\text{Perm}\{1,2,\dots,n\}}\left[\sum_{k=1}^{n}\|D_\theta(s_{t+1}^{(k)};\sigma_k, s_t, a_t^{i_k}) - s_{t+1}\|^2\right]$$
$$= \mathbb{E}_\tau\mathbb{E}_{k\sim\text{Uniform}\{1,2,\dots,n\}}\left[\|D_\theta(s_{t+1}^{\tau};\sigma(\tau), s_t, a_t^k) - s_{t+1}\|^2\right],$$

Model Comparison

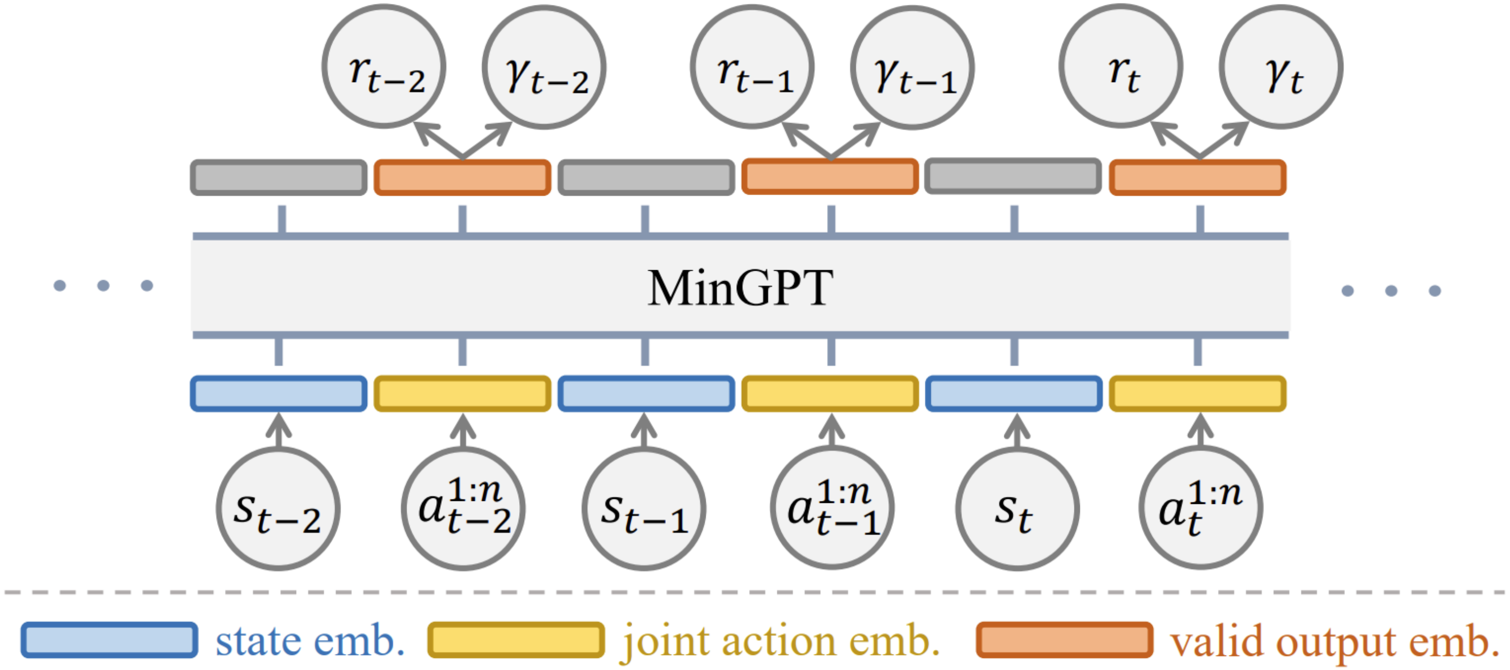


$|\mathcal{S}| \times |\mathcal{A}|^n \times |\mathcal{S}| \to |\mathcal{S}|$

$|\mathcal{S}| \times |\mathcal{A}| \times |\mathcal{S}| \to |\mathcal{S}|$

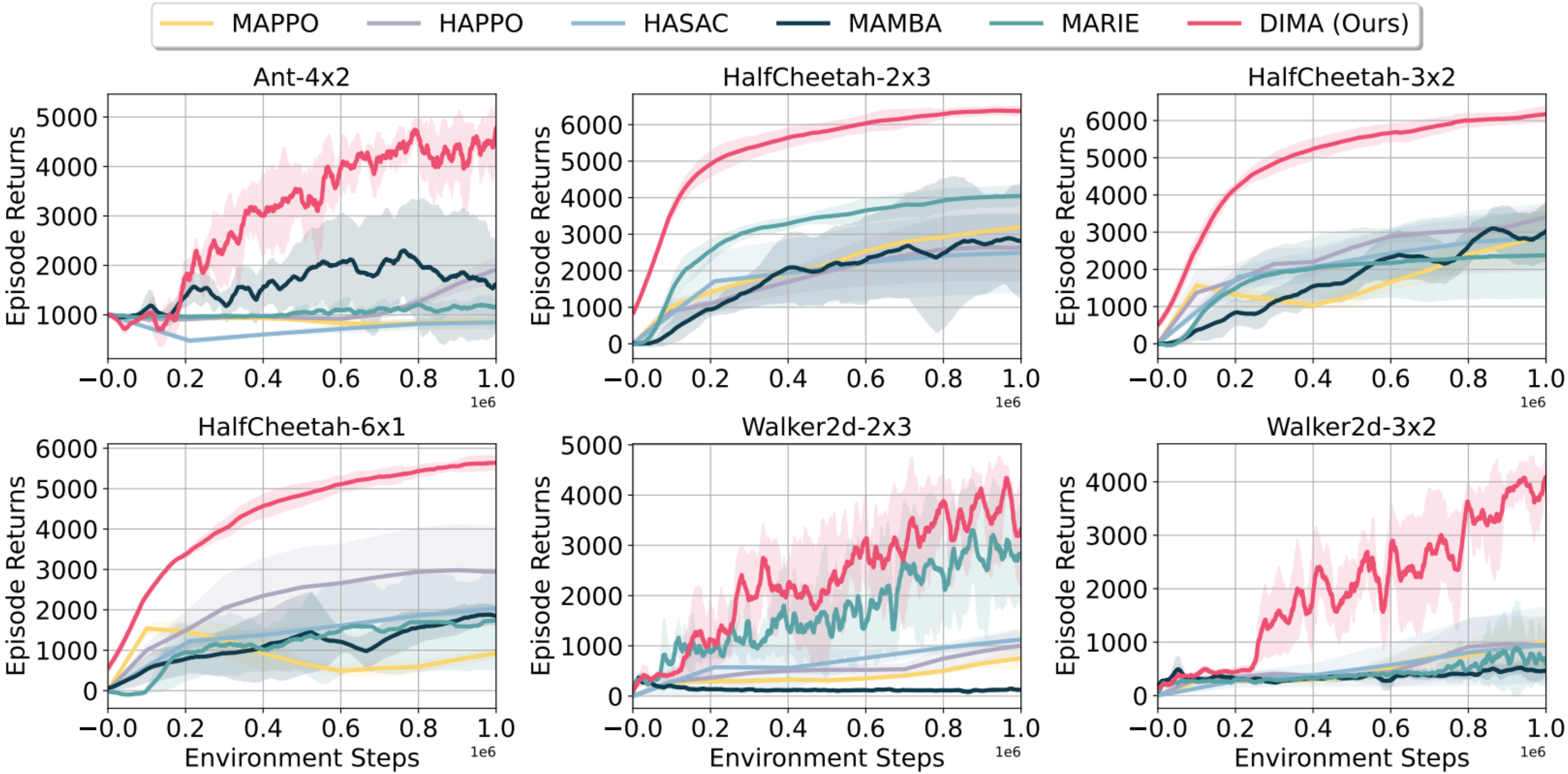
Learning in Imaginations

- For policy learning, DIMA integrates with a learning-in-imagination paradigm using:
  - A Transformer-based reward and termination model
  - A VQ-VAE state decoder for converting global states to local observations
  - MAPPO for policy optimization using Centralized Training with Decentralized Execution (CTDE)



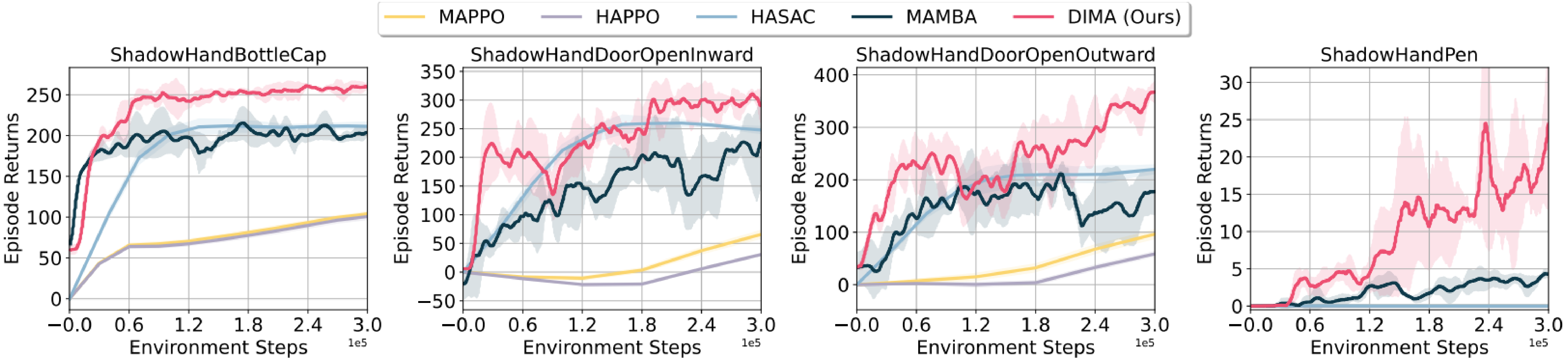
Experiment

- MAMuJoCo



Experiment

- Bi-DexHands

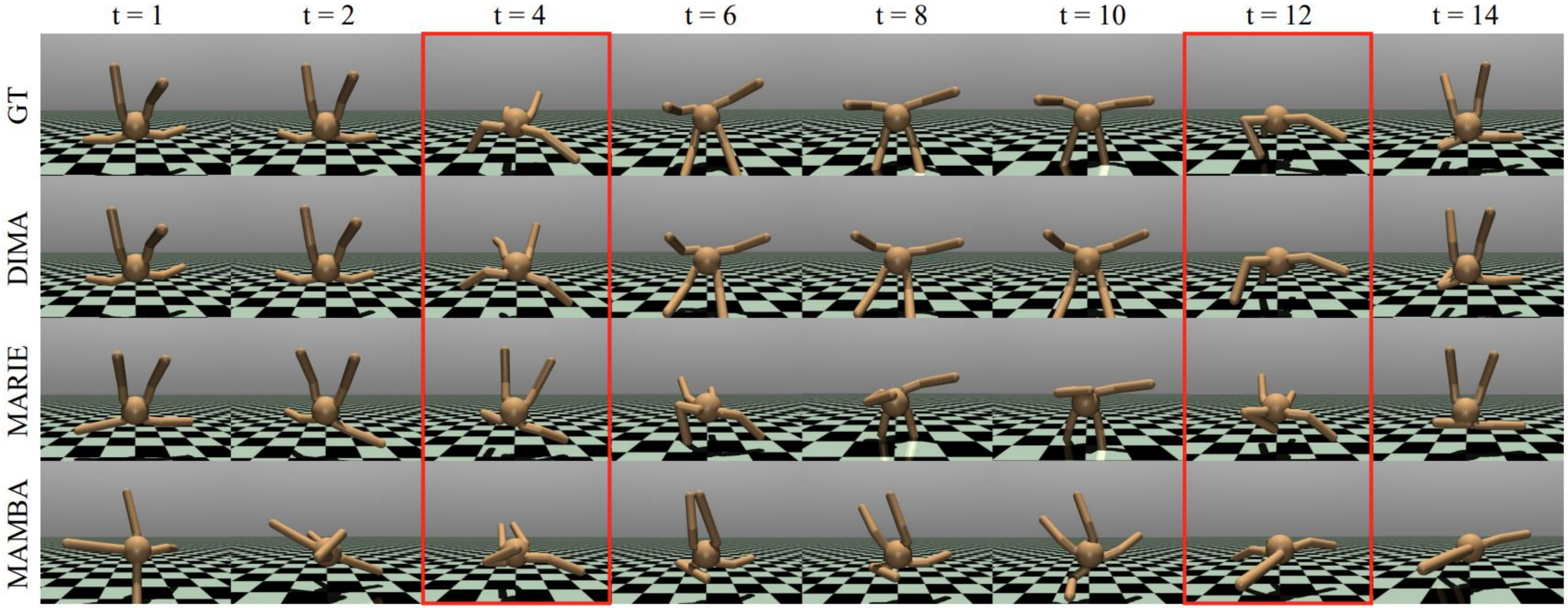


Experiment

Tasks	Steps	Methods					
		DIMA (Ours)	MARIE	MAMBA	HASAC	HAPPO	MAPPO
MAMuJoCo							
Ant-2x4	1M	$\mathbf{4881}_{\pm 756}$	$\underline{4471}_{\pm 553}$	$1314_{\pm 756}$	$1344_{\pm 282}$	$1716_{\pm 449}$	$859_{\pm 47}$
Ant-4x2		$\mathbf{4766}_{\pm 450}$	$1173_{\pm 136}$	$1618_{\pm 931}$	$850_{\pm 126}$	$\underline{1917}_{\pm 253}$	$854_{\pm 41}$
HalfCheetah-2x3		$\mathbf{6370}_{\pm 121}$	$\underline{4045}_{\pm 275}$	$2813_{\pm 1580}$	$2499_{\pm 1081}$	$2628_{\pm 893}$	$3196_{\pm 75}$
HalfCheetah-3x2		$\mathbf{6175}_{\pm 212}$	$2380_{\pm 1145}$	$3029_{\pm 798}$	$2872_{\pm 890}$	$\underline{3402}_{\pm 317}$	$2936_{\pm 766}$
HalfCheetah-6x1		$\mathbf{5643}_{\pm 163}$	$1738_{\pm 1213}$	$1848_{\pm 220}$	$2044_{\pm 110}$	$\underline{2939}_{\pm 1113}$	$925_{\pm 121}$
Walker2d-2x3		$\mathbf{3329}_{\pm 1056}$	$\underline{2822}_{\pm 997}$	$124_{\pm 19}$	$1135_{\pm 210}$	$1007_{\pm 282}$	$752_{\pm 216}$
Walker2d-3x2		$\mathbf{4084}_{\pm 357}$	$604_{\pm 349}$	$466_{\pm 103}$	$958_{\pm 715}$	$932_{\pm 513}$	$\underline{1004}_{\pm 480}$
Bi-DexHands							
BottleCap	300K	$\mathbf{259.9}_{\pm 4.1}$	-	$203.8_{\pm 5.2}$	$\underline{210.9}_{\pm 6.1}$	$100.7_{\pm 3.8}$	$104.0_{\pm 2.3}$
DoorOpenInward		$\mathbf{290.4}_{\pm 29.0}$	-	$225.0_{\pm 79.4}$	$\underline{246.3}_{\pm 7.0}$	$30.7_{\pm 2.5}$	$65.8_{\pm 6.9}$
DoorOpenOutward		$\mathbf{367.1}_{\pm 19.4}$	-	$177.4_{\pm 43.1}$	$\underline{221.9}_{\pm 7.3}$	$58.8_{\pm 4.6}$	$96.4_{\pm 8.5}$
BottleCap		$\mathbf{24.4}_{\pm 11.4}$	-	$\underline{4.3}_{\pm 0.4}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$	$0.0_{\pm 0.0}$

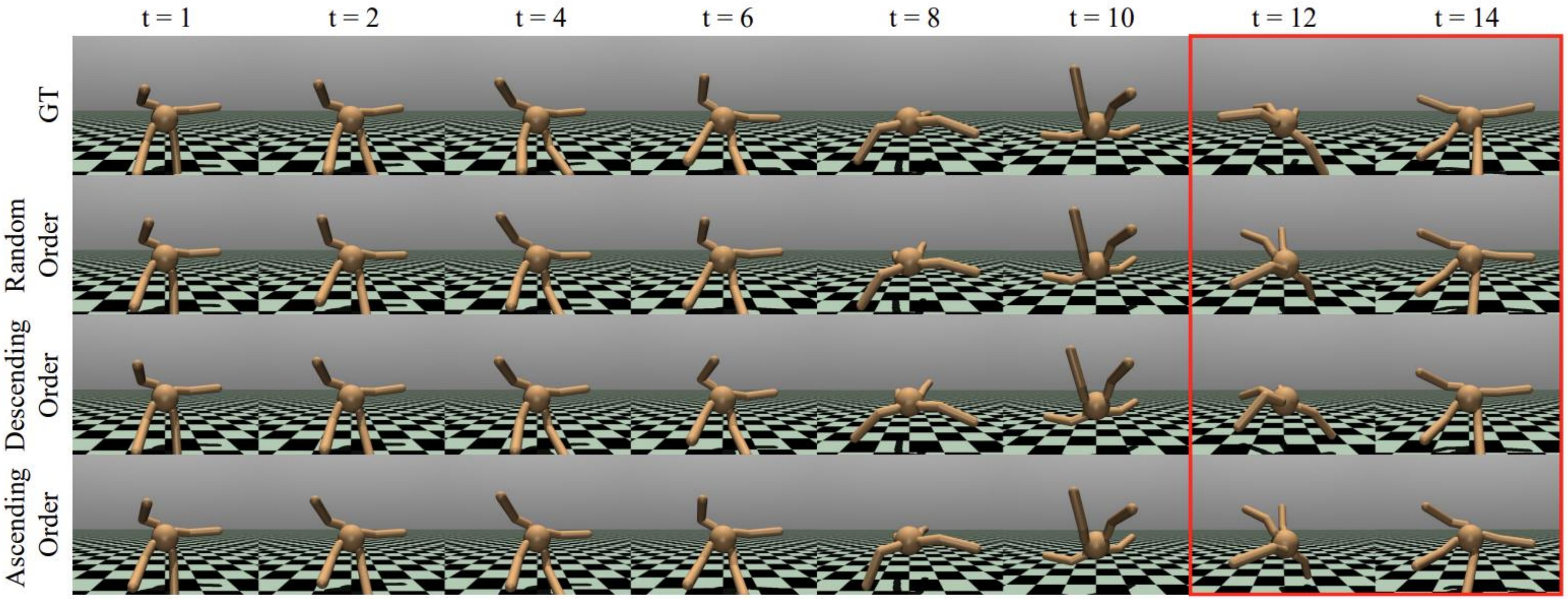
Qualitative Analysis

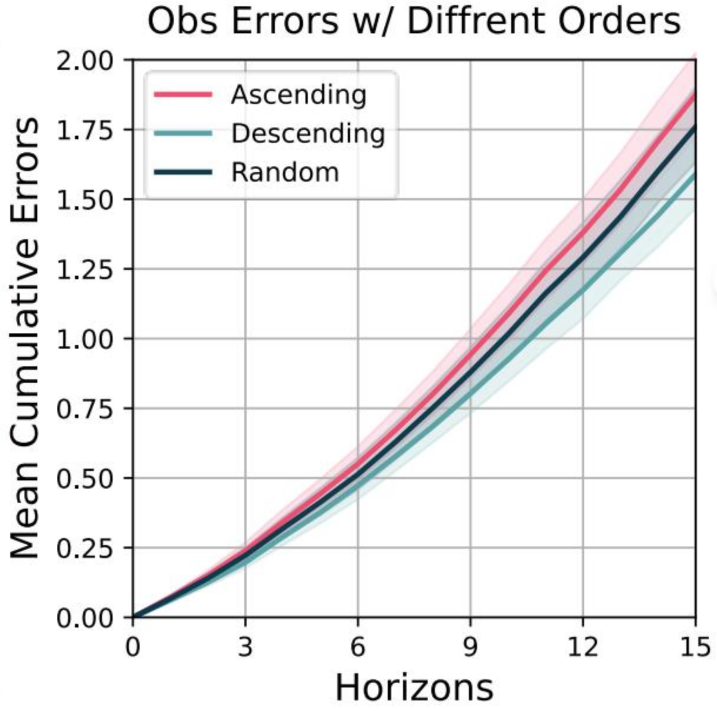
- DIMA demonstrates substantially more accurate and stable long-horizon predictions than existing multi-agent world models



Qualitative Analysis

- DIMA effectively preserves permutation invariance over a long horizon





Ablation Study

- Our proposed formulation improves sample efficiency in lower-data regimes

Table 2: Ablation study on **ShadowHandBottleCap** comparing sequential (DIMA) vs. joint modeling under varying data budgets (8 runs). Sequential modeling shows superior performance and lower variance in lower-data regimes.

Method	100K Steps	150K Steps	200K Steps	250K Steps	300K Steps
Joint	$234.1_{\pm 20.6}$	$238.6_{\pm 22.9}$	$246.7_{\pm 10.9}$	$243.7_{\pm 18.2}$	$255.2_{\pm 7.0}$
Sequential (Ours)	$\mathbf{251.8}_{\pm 17.3}$	$\mathbf{248.2}_{\pm 11.6}$	$246.3_{\pm 14.6}$	$\mathbf{251.9}_{\pm 12.7}$	$249.2_{\pm 10.7}$

Table 3: Ablation study on complex Bi-DexHands tasks at 300k steps (8 runs). The advantage of sequential modeling persists in more challenging environments.

Method	DoorOpenOutward @ 300K steps	DoorOpenInward @ 300K steps
Joint	$302.5_{\pm 76.9}$	$235.1_{\pm 68.1}$
Sequential (Ours)	$\mathbf{352.4}_{\pm 40.5}$	$\mathbf{290.3}_{\pm 30.4}$

Ablation Study

- Sequential Modeling Retains Full Predictive Accuracy with Reduced Complexity

Table 8: Ablation study comparing the **cumulative L1 observation errors** of sequential vs. joint modeling. Models were trained on 500k transitions and evaluated on a 500k held-out set. Sequential modeling achieves statistically indistinguishable prediction accuracy, validating its design.

Task	**Method**	**Obs L1 Error @ $H = 15$**	**Obs L1 Error @ $H = 20$**
DoorOpenOutward	Sequential (Ours)	$\mathbf{5.333}_{\pm 0.273}$	$\mathbf{7.081}_{\pm 0.325}$
	Joint	$5.345_{\pm 0.267}$	$7.092_{\pm 0.324}$
DoorOpenInward	Sequential (Ours)	$\mathbf{5.563}_{\pm 0.326}$	$\mathbf{7.447}_{\pm 0.393}$
	Joint	$5.565_{\pm 0.322}$	$7.453_{\pm 0.386}$
Pen	Sequential (Ours)	$\mathbf{6.667}_{\pm 1.764}$	$\mathbf{8.936}_{\pm 2.328}$
	Joint	$6.676_{\pm 1.762}$	$8.947_{\pm 2.322}$

Contact

Yanami Anna

北京 海淀

扫一扫上面的二维码图案，加我为朋友。

Thanks for listening