NEURAL INFORMATION PROCESSING SYSTEMS

# Structural Information-based Hierarchical Diffusion for Offline Reinforcement Learning

**Xianghua Zeng**[1], Hao Peng[1], Yicheng Pan[1], Angsheng Li[1,2], Guanlin Wu[3]

[1]State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
[2]Zhongguancun Laboratory, Beijing, China
[3]National University of Defense Technology, Changsha, China

**Email:** zengxianghua@buaa.edu.cn

**Code:** https://github.com/SELGroup/SIHD

北京航空航天大學 BEIHANG UNIVERSITY

# Hierarchical Diffusion-Based Offline RL

**Offline Reinforcement Learning:**

- policy learning without any environmental interaction
- extrapolation error caused by out-of-distribution states and actions
- mode-covering challenge from diverse hierarchical trajectories

**Diffusion Models for Offline RL:**

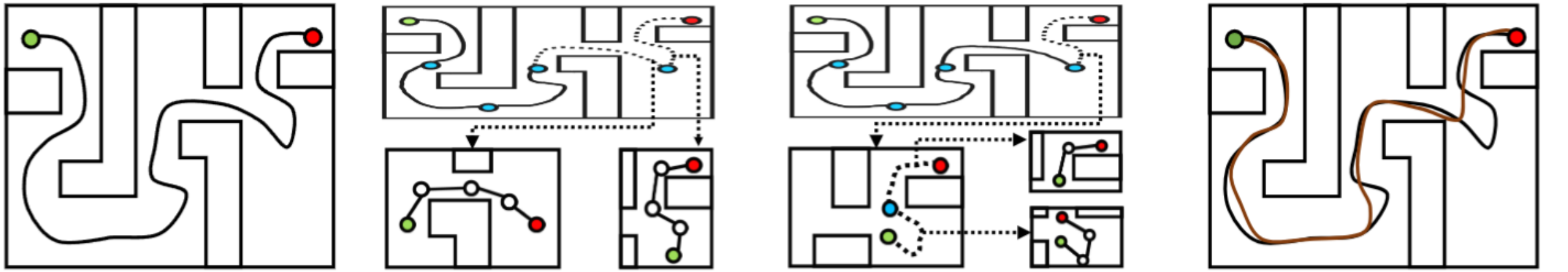
- building diffusion-based policies with strong generative capacity
- reframing decision-making as offline trajectory generation
- exponential increase in estimate variance and high computational cost

**Hierarchical Diffusion for Offline RL:**

- task decomposition into manageable subproblems
- introducing manually predefined skills into diffusion
- trajectory segmentation for subgoal-directed diffusion

# Hierarchical Diffusion-Based Offline RL

## Illustrative Example For Hierarchical Trajectory Generation:



a) offline trajectory b) 2-layer rigid diffusion c) multi-scale hierarchy d) regularized exploration

## Summary:

- single predefined temporal scale to segment offline trajectories
- fixed two-layer diffusion hierarchy composed of subgoal and action layers
- hindering adaptability to varying temporal patterns and task-specific complexities

# Structural Information Principles

The structural entropy measures the uncertainty of a graph under a strategy of hierarchical partitioning, which is called an "encoding tree".

The encoding tree of $G$ is defined as a rooted tree $T$ with the following properties:

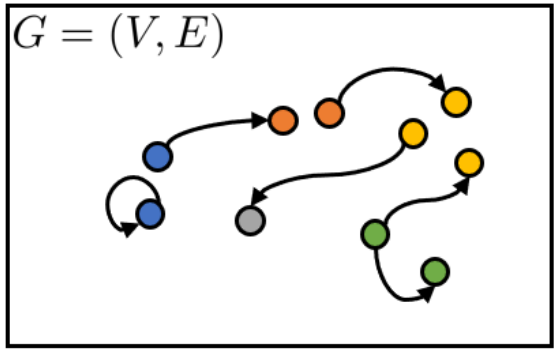
1) For each node $\alpha \in T$, there is a vertex subset in $G$ corresponding with $\alpha$, $T_\alpha \subseteq V$.

2) For the root node $\lambda$, we set $T_\lambda = V$.

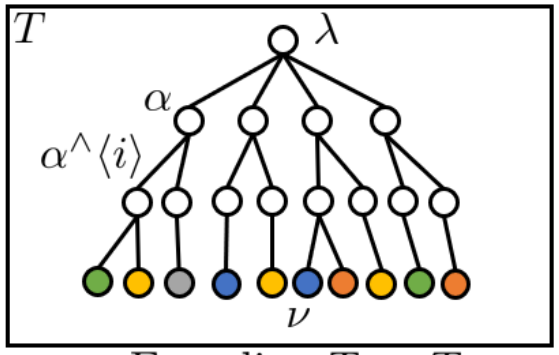
3) For each node $\alpha \in T$, its children nodes are marked as $\alpha^\wedge\langle i\rangle$ ordered from left to right as $i$ increases, and $\alpha^\wedge\langle i\rangle^- = \alpha$.

4) For each node $\alpha \in T$, we suppose that $L$ is the number of its children nodes; then all vertex subsets $T_{\alpha^\wedge\langle i\rangle}$ are disjointed, and $T_\alpha = \bigcup_{i=1}^{L} T_{\alpha^\wedge\langle i\rangle}$.

5) For each leaf node $\nu$, $T_\nu$ is a singleton containing a single vertex in $V$.




Graph $G$

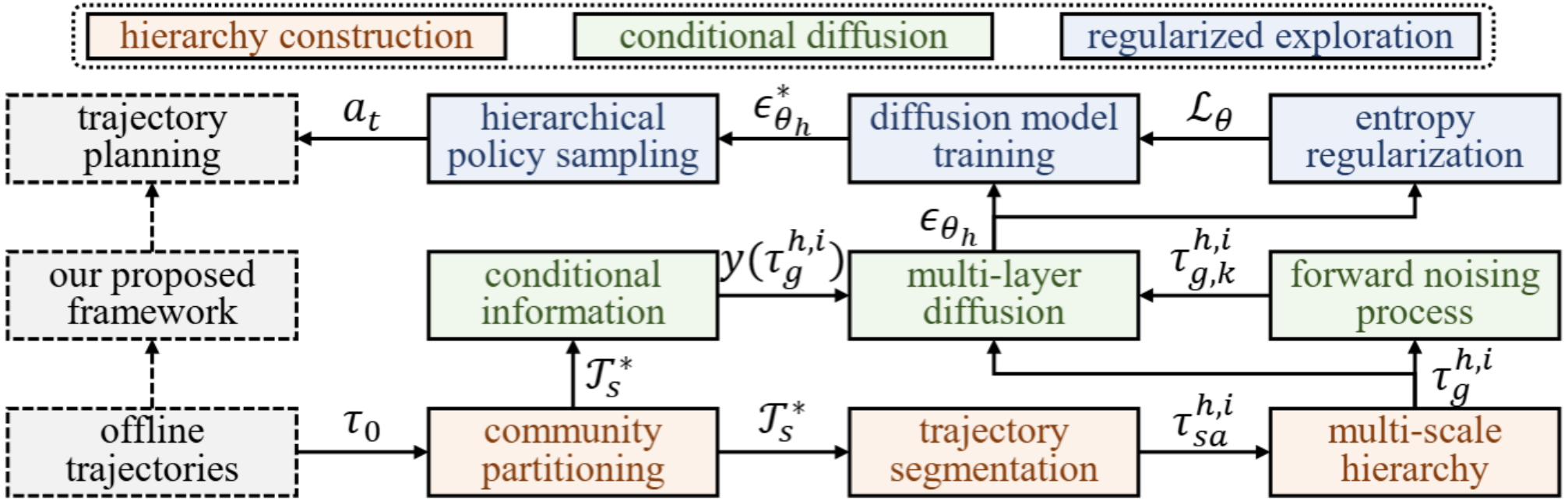



Encoding Tree $T$

# Our Proposed Adversarial Attack Framework

## Overall SIHD Framework:

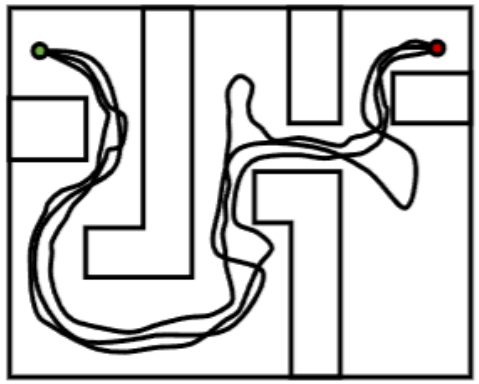
- structural information guided construction of adaptive multi-scale diffusion hierarchy
- conditional forward diffusion and reverse inference at multiple time scales
- encouraging regularized exploration for underrepresented offline states



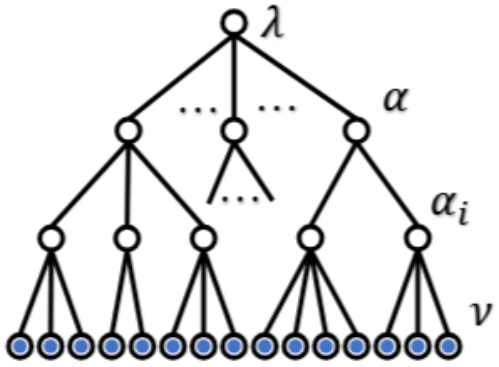
# Our Proposed Hierarchical Diffusion Framework

## Stage 1: Multi-Scale Diffusion Hierarchy

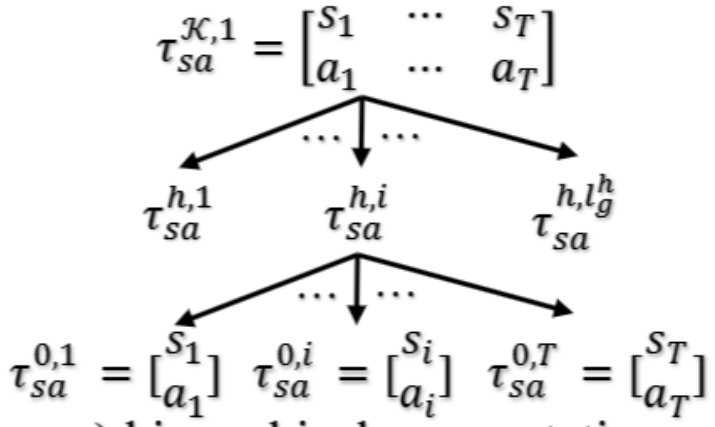
- leverage feature similarity to identify structural relationships among offline states
- derive a tree-structured partitioning of offline state communities
- construct the trajectory-adaptive multi-scale hierarchical diffusion instead of the rigid two-layer diffusion hierarchy operating at a single scale



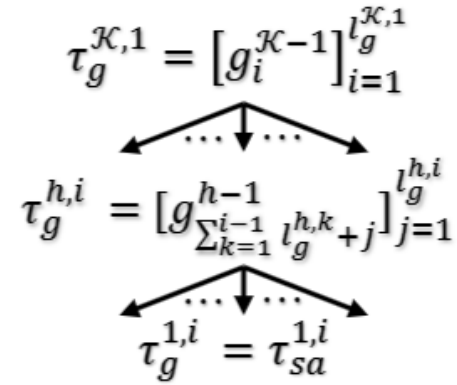
a) offline trajectories




b) community partitioning




c) hierarchical segmentation




d) multi-layer subgoals

# Our Proposed Hierarchical Diffusion Framework

## Stage 2: Conditional Diffusion Model

- generate a subgoal sequence conditioned on task rewards at the top of diffusion hierarchy
- generate subgoal sequences at intermediate levels of the hierarchy or state-action sequences at the base level, with each generation process conditioned on the parent subgoal
- compute the structural information gain as classifier-free guidance to enhance conditional hierarchical generation

## Stage 3: Structural Entropy Regularizer

- introduce a structural entropy-based exploration regularizer
- encourage the hierarchical policy to explore underrepresented regions of the state space
- limit extrapolation errors resulting from deviations from the offline behavioral policy

# Comparative Experiments on the D4RL Benchmark

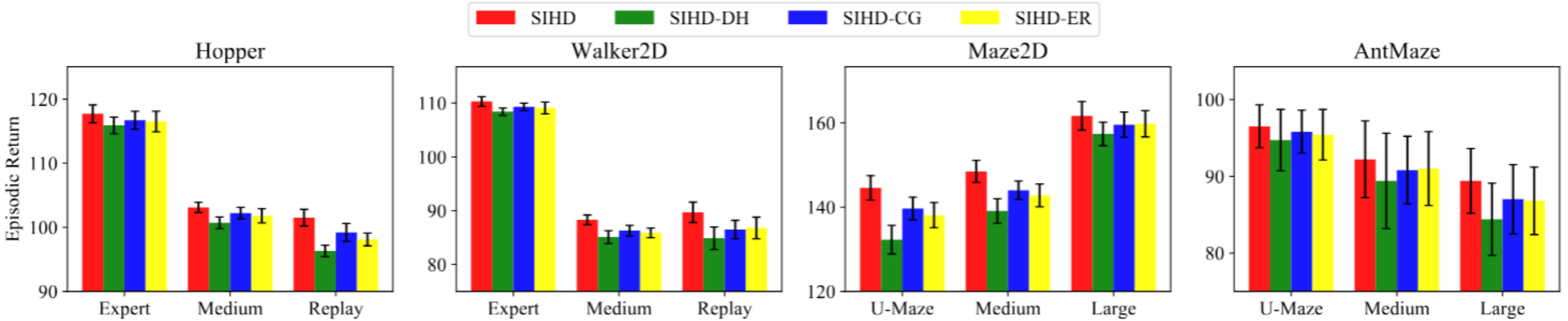
- **Datasets:** Gym-MuJoCo, Maze2D and AntMaze benchmark tasks
- **Baselines:** model-free methods (CQL, IQL, and Decision Transformer), model-based methods (MoReL and Trajectory Transformer), diffusion-based methods (Diffuser, HDMI, HD)

| Gym-MuJoCo | HalfCheetah | | | Hopper | | | Walker2D | | |
|---|---|---|---|---|---|---|---|---|---|
| Tasks | Expert | Medium | Replay | Expert | Medium | Replay | Expert | Medium | Replay |
| CQL | 91.6 | 44.0 | 45.5 | 105.4 | 58.5 | 95.0 | 108.8 | 72.5 | 77.2 |
| IQL | 86.7 | 47.4 | 44.2 | 91.5 | 66.3 | 94.7 | 109.6 | 78.3 | 73.9 |
| DT | 86.8 | 42.6 | 36.6 | 107.6 | 67.6 | 82.7 | 108.1 | 74.0 | 66.6 |
| MoReL | 53.3 | 42.1 | 40.2 | 108.7 | 95.4 | 93.6 | 95.6 | 77.8 | 49.8 |
| TT | 95.0 | 46.9 | 41.9 | 110.0 | 61.1 | 91.5 | 101.9 | 79.0 | 82.6 |
| Diffuser | 88.9 ± 0.3 | 42.8 ± 0.3 | 37.7 ± 0.5 | 103.3 ± 1.3 | 74.3 ± 1.4 | 93.6 ± 0.4 | 106.9 ± 0.2 | 79.6 ± 0.6 | 70.6 ± 1.6 |
| HDMI | 92.1 ± 1.4 | 48.0 ± 0.9 | 44.9 ± 2.0 | 113.5 ± 0.9 | 76.4 ± 2.6 | 99.6 ± 1.5 | 107.9 ± 1.2 | 79.9 ± 1.8 | 80.7 ± 2.1 |
| HD | 92.5 ± 0.3 | 46.7 ± 0.2 | 38.1 ± 0.7 | 115.3 ± 1.1 | 99.3 ± 0.3 | 94.7 ± 0.7 | 107.1 ± 0.1 | 84.0 ± 0.6 | 84.1 ± 2.2 |
| SIHD | **94.4** ± 0.5 | **48.7** ± 1.1 | **47.0** ± 0.4 | **117.7** ± 1.4 | **103.1** ± 0.8 | **101.5** ± 1.3 | **110.3** ± 0.9 | **88.3** ± 0.9 | **89.7** ± 1.9 |
| Abs.(%) Avg. | 1.9(2.1) | 0.7(1.5) | 1.5(3.3) | 2.4(2.1) | 3.8(3.8) | 1.9(1.9) | 0.7(0.6) | 4.3(5.1) | 5.6(6.7) |

| Gym-MuJoCo | Single-task Maze2D | | | Multi-task Maze2D | | | AntMaze | | |
|---|---|---|---|---|---|---|---|---|---|
| Tasks | U-Maze | Medium | Large | U-Maze | Medium | Large | U-Maze | Medium | Large |
| IQL | 47.4 | 34.9 | 58.6 | 24.8 | 12.1 | 13.9 | 62.2 | 70.0 | 47.5 |
| MPPI | 33.2 | 10.2 | 5.1 | 41.2 | 15.4 | 8.0 | - | - | - |
| Diffuser | 113.9 ± 3.1 | 121.5 ± 2.7 | 123.0 ± 6.4 | 128.9 ± 1.8 | 127.2 ± 3.4 | 132.1 ± 5.8 | 76.0 ± 7.6 | 31.9 ± 5.1 | - |
| IRIS | - | - | - | - | - | - | 89.4 ± 2.4 | 64.8 ± 2.6 | 43.7 ± 1.3 |
| HiGoC | - | - | - | - | - | - | 91.2 ± 1.9 | 79.3 ± 2.5 | 67.3 ± 3.1 |
| HDMI | 120.1 ± 2.5 | 121.8 ± 1.6 | 128.6 ± 2.9 | 131.3 ± 1.8 | 131.6 ± 1.9 | 135.4 ± 2.5 | - | - | - |
| HD | 128.4 ± 3.6 | 135.6 ± 3.0 | 155.8 ± 2.5 | 144.1 ± 1.2 | 140.2 ± 1.6 | 165.5 ± 0.6 | 94.0 ± 4.9 | 88.7 ± 8.1 | 83.6 ± 5.8 |
| SIHD | **144.6** ± 2.9 | **148.5** ± 2.6 | **161.7** ± 3.4 | **157.0** ± 0.6 | **156.8** ± 1.7 | **169.4** ± 2.7 | **96.5** ± 2.8 | **92.2** ± 5.0 | **89.4** ± 4.2 |
| Abs.(%) Avg. | 16.2(12.6) | 12.9(9.5) | 5.9(3.8) | 12.9(9.0) | 16.6(11.8) | 3.9(2.4) | 2.5(2.7) | 3.5(3.9) | 5.8(6.9) |

# Ablation Study on Primary SIHD Modules

- SIHD-DH: replacing the multi-scale hierarchy with a rigid two-layer single-scale diffuser
- SIHD-CG: removing the classifier-based guidance in conditional diffusion
- SIHD-ER: removing the structural entropy regularizer in policy learning
- highlighting the importance of each component and the critical role of the multi-scale diffusion hierarchy in SIHD
- becoming even more pronounced in long-horizon decision-making tasks



# Conclusion and Future Works

- This work proposes SIHD, a novel hierarchical diffusion framework that leverages structural information embedded in historical trajectories to construct an adaptive multi-scale diffusion hierarchy and promote exploration of underrepresented states in offline datasets, thereby enabling effective policy learning.
- Extensive evaluations on the challenging D4RL benchmark demonstrate the superior decision-making performance and generalization capabilities of SIHD across diverse offline RL tasks.
- In future work, we aim to refine the hierarchical diffusion framework by further exploring how to more effectively represent and integrate subgoal constraints as conditional information.
- We also plan to extend the SIHD framework to other offline RL environments and broader diffusion-based generative modeling domains.

NEURAL INFORMATION PROCESSING SYSTEMS

# Structural Information-based Hierarchical Diffusion for Offline Reinforcement Learning

**Xianghua Zeng**[1], Hao Peng[1], Yicheng Pan[1], Angsheng Li[1,2], Guanlin Wu[3]

[1]State Key Laboratory of Software Development Environment, Beihang University, Beijing, China
[2]Zhongguancun Laboratory, Beijing, China
[3]National University of Defense Technology, Changsha, China

**Email:** zengxianghua@buaa.edu.cn

**Code:** https://github.com/SELGroup/SIHD

北京航空航天大學 BEIHANG UNIVERSITY