COMPOSITIONAL DIFFUSION PLANNING WITH CHUNK SELECTION AND REACHABILITY MEASURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Compositional diffusion planners have demonstrated remarkable success on long-horizon diffusion planning by sampling short trajectories, namely chunks, and composing them into a long trajectory. However, averaging overlapping chunks can induce the mode-averaging problem, resulting in discontinuous trajectories. Several works aim to address this problem, whereas a non-negligible portion of plans remains infeasible. In this work, we propose *Selective Compositional Diffuser (SCD)*, handling this problem by measuring reachability. To this end, we present a self-supervised learning method for the reachability classifier. This is trained on offline data and generated chunks jointly to address potential distributional mismatch, while reducing the bias inherited from the generated data. Based on this, we construct a graph whose edge costs represent whether consecutive chunks are reachable, and traverse the graph to find effective yet efficient routes. Lastly, we generate bridging trajectories and refine the composed trajectory for smooth transitions. We evaluate our method on OGBench stitching tasks. It practically reduces incompatible mode compositions and the mode-averaging problem compared to the studied baselines. Across tasks, SCD improves success rates, reaching state-of-the-art results on HumanoidMaze-Large and Giant tasks.

1 INTRODUCTION

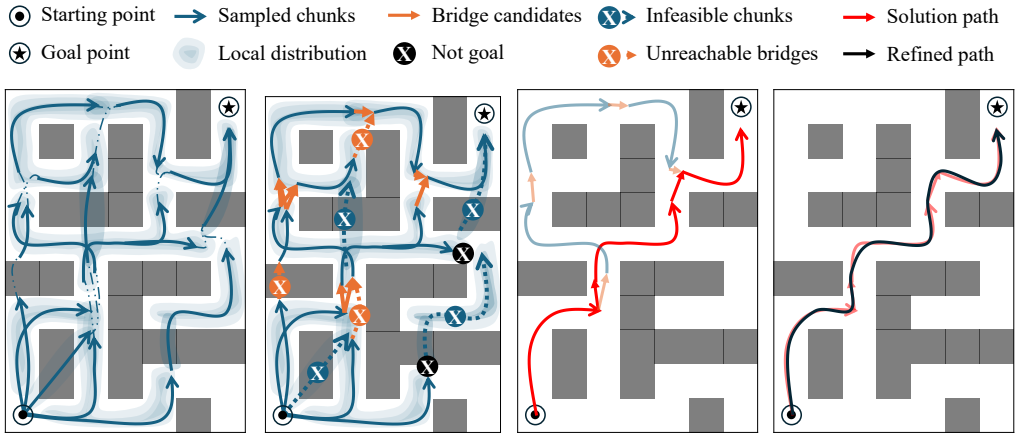
Diffusion models (Song & Ermon, 2019; Ho et al., 2020) have demonstrated remarkable results in robotic planning and decision-making (Ajay et al., 2023; Chi et al., 2023; Janner et al., 2022). These models abstract planning as a network-based generative process and are known to handle multimodality, flexible conditioning, and guidance-style control signals (Dhariwal & Nichol, 2021; Ho & Salimans, 2021). However, since diffusion models are inherently constrained by the training data, diffusion-based planners may require a combinatorial-scale dataset to model all possible plans.

To handle this problem, recent works (Luo et al., 2025; Lee et al., 2026; Sun et al., 2026; Mishra et al., 2026) try to stitch or compose short trajectories into a full trajectory for long-horizon tasks. To this end, they learn short-horizon trajectories, namely chunks, conditioned on boundary states and compose them at inference time. Since short trajectories are more practical to collect in the real world than full ones, compositional diffusion planners represent a promising and practical direction.

Previously, several works (Mishra et al., 2023; Luo et al., 2025) generate multiple chunks and compose overlapping regions by averaging. In particular, when the chunk distribution is multimodal, it is known to suffer from the mode-averaging problem, and it can result in discontinuous trajectories. Recently, Mishra et al. (2026); Sun et al. (2026); Lee et al. (2026) have addressed this problem with iterative resampling (Mishra et al., 2026), conservative score correction (Sun et al., 2026), and overlap consistency guidance (Lee et al., 2026). However, as shown in Fig. 2, *we observe that a non-negligible portion of plans remains infeasible, failing to arrange feasibly composable chunks.*

To handle this issue, *we propose Selective Compositional Diffuser, called SCD*. As shown in Fig. 1, we reformulate composition as a composable chunk selection problem over the candidate plans. To find a composable chunk sequence, we construct a graph using chunks as vertices. Edges are constructed between chunks and weighted by a score representing whether the next chunk could be *reachable* within a fixed time budget. By traversing the graph, we can find the minimum-cost path, which selects the sequence of chunks that are most likely to be reachable, thereby composable. We connect them with a fixed-length trajectory, often called a *bridge*.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

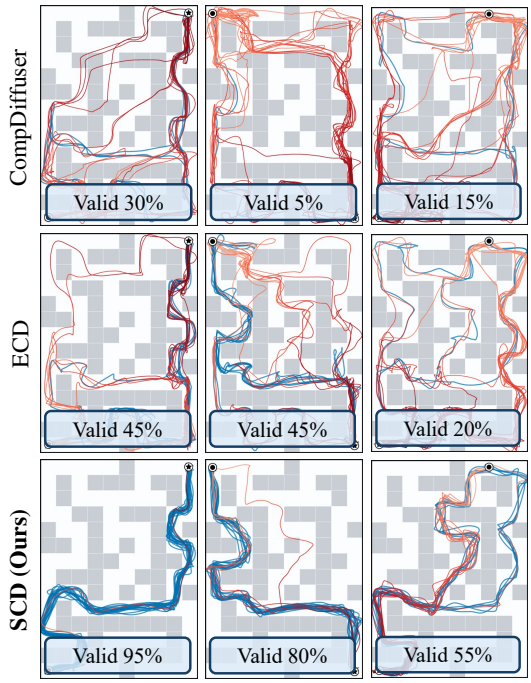


(a) Generated chunks (b) Measuring reachability (c) Graph-based search (d) Refining the solution

Figure 1: **Overview of the SCD pipeline.** Double-dotted lines in Fig. 1a represent the initially generated plans. To find the long-horizon executable plans, (i) we generate plans as in Fig. 1a, (ii) evaluate each state connection using our reachability classifier as in Fig. 1b, (iii) select a minimum-cost state sequence as in Fig. 1c, (iv) build the bridge and refine the generated trajectory as in Fig. 1d.

To measure reachability, we present a self-supervised method to train a lightweight MLP network, classifying whether one state can reach the other within a given time lag. Once we find the minimum-cost path with this learned reachability score, we generate bridges and finally stitch the chunks.

We evaluate our method on long-horizon planning tasks from OGBench (Park et al., 2025). We observe that our method reduces infeasible transitions between composed chunks. Our results outperform baseline diffusion planners and goal-conditioned reinforcement learning methods, particularly on the giant-sized HumanoidMaze task, improving the success rate from the previous state-of-the-art result of 67% to 83% in the studied settings.



(a) Task 01 (b) Task 02 (c) Task 03

Figure 2: **Failure cases in HumanoidMaze-Giant.** The red color denotes infeasible trajectories and the blue color denotes feasible ones. *Valid* refers to the percentage of feasible trajectories. More samples are provided in App. D, Fig. 12.

2 Related Work

Planning with Diffusion Models. Diffusion models (Song & Ermon, 2019; Ho et al., 2020) have been adopted for decision making (Janner et al., 2022; Ajay et al., 2023; Chi et al., 2023). In particular, diffusion planners (Janner et al., 2022) model complete state-action sequences, enabling inference-time constraints through conditioning and guidance. However, monolithic diffusion planners have been observed to degrade when generating trajectories longer than those learned during training (Luo et al., 2025). This is relevant when the available data are primarily short and local trajectories.

Compositional Diffusion Planning. To address this limitation, compositional diffusion planners (Mishra et al., 2023; Luo et al., 2025) represent a long trajectory through multiple local tra-

jectories. Generative Skill Chaining (Mishra et al., 2023) composes skill-specific diffusion models through shared boundary variables, while CompDiffuser (Luo et al., 2025) generates overlapping trajectory segments with bidirectional neighbor conditioning. However, as local trajectory distributions are multimodal, neighboring segments may converge to incompatible modes. Combining these predictions can therefore produce a low-density transition that is supported by neither segment.

Recent methods aim to address this mode-averaging failure. RCD (Lee et al., 2026) guides sampling using reconstruction-based density estimation with overlap consistency, ECD (Sun et al., 2026) derives a global energy correction, and CDGS (Mishra et al., 2026) searches over combinations of local modes through iterative resampling and pruning. However, we observe a non-negligible portion of infeasible plans in Fig. 2, failing to find feasibly composable chunk sequences.

Graph Search for Long-Horizon Planning. Graph search has been used to convert locally reliable transitions into long-horizon plans. SoRB (Eysenbach et al., 2019) constructs a graph over states from the replay buffer and uses a goal-conditioned value function as a cost. In offline settings, GAS (Baek et al., 2025) and TTGS (Opryshko et al., 2026) construct subgoal graphs from offline datasets, and locally plan the searched subgoal sequence using a goal-conditioned policy.

Recently, XDiffuser (Hassidof et al., 2026) constructs a temporal-distance graph from offline states and searches waypoints to guide a compositional diffusion model. Using offline data for graph construction may become incompatible with the current planning problem, and thus our method aims to construct graphs with currently generated trajectories to reduce the offline-online gap (Sec. 3.3).

Reachability and Temporal-Distance Estimation. Goal-conditioned reinforcement learning often measures how states are connected. C-Learning (Eysenbach et al., 2021) measures the state relation by distinguishing future states from independently sampled states with the recursive classifier. Contrastive RL (Eysenbach et al., 2022) also learns this relational metric by contrastive learning, while QRL (Wang et al., 2023) models the goal-reaching cost as a quasimetric value function. Primarily, they do not aim to directly measure whether two states are reachable in a given time budget.

More directly, Savinov et al. (2019) predicts whether two observations occur within a fixed number of environment steps, and Mezghani et al. (2022) adopts it for offline RL. They use observed trajectories for training datasets, and may suffer from the distributional mismatch over generated samples. This motivates us to use generative samples jointly for our reachability classifier in Sec. 3.2.

3 METHOD

In this section, we introduce our method for composing the generated chunks by leveraging the learned reachability score. We reformulate composition as a chunk selection problem over candidates, as described in Sec. 3.1. We propose the self-supervised method for a lightweight reachability network in Sec. 3.2, measuring the feasibility of a chunk connection. Then we present the graph-based composition method for reachable and feasible plans in Secs. 3.3 and 3.4.

3.1 SELECTION OF COMPOSITION MODE

Given the start state q_s and the goal state q_g, the compositional diffuser generates N candidate plans $\mathcal{C}_{1:N}$, each consisting of K overlapping chunks τ:

$$\mathcal{C}_n = \tau_{1:K}^{(n)} = (\tau_1^{(n)}, \dots, \tau_K^{(n)}), \quad \tau_k^{(n)} = s_{1:H}^{(n,k)} = (s_1^{(n,k)}, \dots, s_H^{(n,k)}), \tag{1}$$

where $s_i^{(n,k)}$ is a state vector. Typically, adjacent chunks are merged into a single trajectory. This operation is based on the assumption that the adjacent chunks on the same route are continuous and composable. However, when they commit to different route modes, merging overlapping regions by averaging cannot achieve a globally feasible plan. As shown in Fig. 2, it either interpolates the infeasible region or leaves the mode mismatch to the boundary chunks.

The observed failure arises from chunk arrangement, rather than their local generation. To handle this problem, we retain the chunks as a pool of candidates $\tau_i \in \{\tau_1^{(1)}, \dots, \tau_K^{(N)}\}$ for $\tau_i = s_{1:H}^{(i)}$, and reformulate the composition as a composable chunk selection problem, $\{\tau_i\}_{i=1}^{T}$. Let an assembly A_i be a subpart of the chunk:

$$A_i = \tau_i[\ell_i{:}r_i] = s_{\ell_i:r_i}^{(i)}. \tag{2}$$

We aim to select assemblies which are most likely to be executable when connected, while penalizing an unnecessarily long horizon:

$$A^* = \underset{A \in \mathcal{A}(\mathcal{C}; q_s, q_g)}{\arg\max} \; p(E|A)\gamma^{T(A)}, \quad 0 < \gamma \leq 1, \tag{3}$$

where $\mathcal{A}(\mathcal{C}; q_s, q_g)$ is a set of assembly chains, $A = (A_1, \ldots, A_T)$ for $s^{(1)}_{\ell_1} = q_s$ and $s^{(T)}_{r_T} = q_g$, $T(A)$ is the length of the plan generated by the chain A, and $p(E|A)$ denotes the executability of the given chain A. We use a learned reachability score to estimate local executability as described in Sec. 3.2, and propose the optimal selection strategy in Sec. 3.3.

3.2 Measuring Reachability

To be an executable trajectory, (i) we suppose that adjacent boundary states of consecutive chunks should be mutually reachable, and (ii) we generate bridging trajectories for composition rather than overlap averaging. Previous work (Mezghani et al., 2022) trains a classifier to represent reachability by using temporally close states as positive pairs and independently sampled or temporally distant states as negative pairs. Both pairs are sampled from the offline dataset; however, our classifier is required to evaluate generated endpoints, which may lie in sparse or unseen regions of offline trajectory distributions. Predictions on these endpoints can therefore be overconfident, or confuse an infeasible shortcut as a reachable path, as shown in Fig. 6 and Tab. 4.

To address this distribution mismatch, we construct a negative dataset as a union of the offline data and the generated data, given by the same generator used in the planning. In this way, the classifier is fit where it will be queried. However, this approach can lead to a second issue: the generated states are not uniformly distributed, and the reachability score can be biased by the generative density (Eq. 7). Therefore, we approximately estimate this generative density to reweight the negative term in Eq. (5), leaving the score depending on the offline data distribution while reducing the bias effect.

For a randomly sampled start state $x_{\text{obs}} = s_t$ and lag L, we sample the reached state $y_{\text{obs}} \sim \mathcal{U}\{s_{t+l}\}_{l=1}^{L}$ and sample negative states $x_{\text{neg}}, y_{\text{neg}}$ independently from a negative density p_{neg}. Let $p_{\text{obs}}(x, y|L)$ denote the joint density of the reachable state pair under the behavior policy for L steps. We train a classifier network f_ϕ to distinguish observed L-step transitions from negative ones:

$$D_\phi(x, y, L) = \sigma(f_\phi(x, y, L)), \quad f_\phi(x, y, L) = Q_\psi(x, \text{PE}(L))^\top \frac{K_\theta(y)}{\|K_\theta(y)\|_2}, \quad \phi = \{\psi, \theta\}, \tag{4}$$

where σ is a sigmoid function and PE is a sinusoidal positional encoding. In practice, we use a 3-layer MLP network for the query encoder Q_ψ and key encoder K_θ.

We consider a weighted binary cross-entropy objective,

$$\begin{aligned} &\mathcal{L}((x_{\text{obs}}, y_{\text{obs}}), (x_{\text{neg}}, y_{\text{neg}}), L) \\ &= -\log D_\phi(x_{\text{obs}}, y_{\text{obs}}, L) - w(x_{\text{neg}})w(y_{\text{neg}}) \log(1 - D_\phi(x_{\text{neg}}, y_{\text{neg}}, L)), \end{aligned} \tag{5}$$

for a scalar weight $w(x) > 0$. Minimizing this loss gives the optimal classifier:

$$D^*(x, y, L) = \frac{p_{\text{obs}}(x, y|L)}{p_{\text{obs}}(x, y|L) + w(x)w(y)p_{\text{neg}}(x)p_{\text{neg}}(y)}, \tag{6}$$

$$f^*(x, y, L) = \log p_{\text{obs}}(x, y|L) - \log(w(x)w(y)p_{\text{neg}}(x)p_{\text{neg}}(y)). \tag{7}$$

Without reweighting, the logit contains the term including $p_{\text{neg}}(x)$ and $p_{\text{neg}}(y)$, biasing the model toward the distribution of the negative dataset. To reduce this bias, we estimate the density of the negative dataset with a k-NN density estimator:

$$\hat{p}_{\text{neg}}(x) = \frac{k}{N_{\text{neg}} V_d r_k(x)^d} \propto \frac{1}{r_k(x)^d}, \tag{8}$$

where N_{neg} is the dataset size and V_d is the volume of the unit ball in d dimensional state space, and $r_k(x)$ is the distance from x to its k-th nearest neighbor in the negative state dataset.

Then, we use this estimated density for the weight $w(x)$:

$$w(x) = \text{clip}\left(r_k(x)^d / \overline{r_k^d},\ w_{\min},\ w_{\max}\right), \qquad \overline{r_k^d} = \mathbb{E}_{x \sim p_{\text{neg}}}\left[r_k(x)^d\right], \tag{9}$$

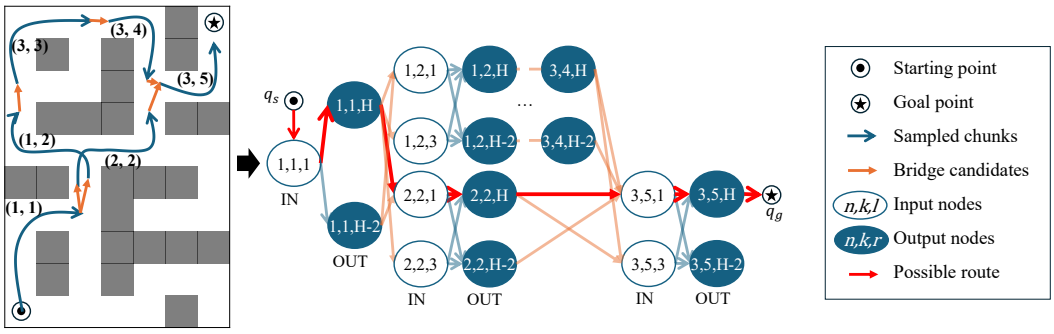


Figure 3: An example of graph construction. This example follows the setting of Fig. 1b.

where the clipping bounds the weight on unreliably estimated densities. Then $w(x) \propto 1/\hat{p}_{\text{neg}}(x)$ for reliably estimated supports and the weighted density is approximately constant in this space. Thus,

$$f^*(x, y, L) \approx \log p_{\text{obs}}(x) + \log p_{\text{obs}}(y|x, L) - \log C, \tag{10}$$

where C is a constant. This pushes the classifier to predict reachability only with offline distribution, rather than with generative distribution. The score can jointly measure two densities, whether the start state is observed in the offline distribution, $\log p_{\text{obs}}(x)$, and whether the transition is reliable, $\log p_{\text{obs}}(y|x, L)$. We use it as the reachability score when constructing the graph in Sec. 3.3.

We can also interpret this score as a lower bound on local executability. Let E be the event that (x, y) is executable within L, and X be the event that the pair is an observed L-or-fewer-step pair. A pair from the offline data is obviously executable, hence $X \subseteq E$. Since D^* is the posterior of X, the reachability score now can be viewed as a lower bound on executability E:

$$D^*(x, y, L) = p(X|x, y, L) \leq p(E|x, y, L). \tag{11}$$

3.3 State Chunk Graph Construction

To find an optimal assembly chain A^* in Eq. (3), we construct a directed graph to select composable chunks. Candidate boundaries ℓ_i and r_i in Eq. (2) include chunk endpoints $\ell_i = 1$ and $r_i = H$ following Yoon et al. (2025), and we extend them to interior points $\ell_i \geq 1$ and $r_i \leq H$ for more flexible chunk selection. Each selected segment $\tau_i[\ell_i{:}r_i]$ is preserved as generated, while the connections between the segments, namely bridges, will be generated after the graph search.

As shown in Fig. 3, we set the candidate boundaries at a fixed stride, such as $\ell \in \{1, 3, 5, \ldots\}$. We represent boundary states of each segment $\tau_k^{(n)}[\ell{:}r]$ with two vertices, $\text{IN}(n, k, \ell)$ and $\text{OUT}(n, k, r)$. For generated chunks, an edge between these vertices represents the segment:

$$\tau_k^{(n)}[\ell{:}r] : \text{IN}(n, k, \ell) \longrightarrow \text{OUT}(n, k, r), \quad r - \ell \geq M, \tag{12}$$

where M is the chunk length excluding overlapping regions on both sides. We add additional candidate edges for bridges, connecting:

$$\text{OUT}(n, k, r) \longrightarrow \text{IN}(n', k', \ell') \tag{13}$$

for $(n, k) \neq (n', k')$. We can evaluate the edge between two consecutive vertices v and v' by computing the reachability score of adjacent boundary states, regularized by the length penalty:

$$c(v, v') = -\log(D_\phi(v, v', h_o)) - h_o \log \gamma, \tag{14}$$

where $\gamma \in (0, 1]$ is a discount factor and h_o is a time budget between two vertices. We use $h_o = r - \ell$ for the segment and a preset constant for the bridge. The reachability score of the segment in Eq. (12) is $D_\phi(s_\ell^{(n,k)}, s_r^{(n,k)}, r - \ell)$, and the score of the bridge in Eq. (13) is $D_\phi(s_r^{(n,k)}, s_{\ell'}^{(n',k')}, h_o)$.

Then we can reorganize an assembly as a sequence of vertices $A = (v_0 = q_s, v_1, \ldots, v_{J-1}, v_J = q_g)$ that form a directed path from q_s to q_g. We can interpret the executability of A as a product of the executability on each element:

$$p(E|A) = \prod_{j=0}^{J-1} p(E|v_j, v_{j+1}, h_o(v_j, v_{j+1})) \geq \prod_{j=0}^{J-1} D_\phi(v_j, v_{j+1}, h_o(v_j, v_{j+1})), \tag{15}$$

where $p(E|v_j, v_{j+1}, h_o)$ is a likelihood of the executability between two states v_j and v_{j+1} within a time budget h_o and $h_o(v_j, v_{j+1})$ is a time budget between two vertices. If we assume that the planner can generate a feasible local plan between reachable state pairs within a given time budget, the executability of the full assembly chain A can be decomposed, thereby inducing the lower bound.

As we prefer short paths, the length-regularized executability becomes:

$$\gamma^{T(A)} \cdot p(E|A) \geq \prod_{j=0}^{J-1} \gamma^{h_o(v_j, v_{j+1})} D_\phi(v_j, v_{j+1}, h_o(v_j, v_{j+1})), \tag{16}$$

which coincides with the objective in Eq. (3). Moreover, this is consistent with our cost function in log-scale. Therefore, we maximize this decomposed lower bound by minimizing the cost:

$$A^* = \arg\min_A \sum_{j=0}^{J-1} c(v_j, v_{j+1}). \tag{17}$$

Since the cost is non-negative, we can use Dijkstra's algorithm to find the minimum-cost path even if the graph contains cycles. We report the wall-clock times of the studied baselines in Tab. 3.

3.4 Bridge Generation and Refinement

The solution path from Eq. (17) specifies which segments to retain and to connect. For each bridge β_i, we provide the adjacent segments as the left and right references including $s_{r_i}^{(i)}$ and $s_{\ell_{i+1}}^{(i+1)}$ for the diffusion planner, then generate the trajectory with length h_o. We concatenate the bridges with the segments to form a full trajectory $\tilde{\tau}$.

Although graph selection identifies reachable connections, it does not guarantee stable motion across the segments. Therefore, we apply a refinement to jointly adjust the segments and bridges. We perturb the full trajectory $\tilde{\tau}$ with small noise and re-denoise the remaining steps, while inpainting the start and goal states. The trajectory is covered with partially overlapping windows, and we denoise each window in parallel, merging the final outputs into a single trajectory. Within the overlapping band, neighboring predictions are combined by a linear interpolation:

$$\hat{s}_i = \left(1 - \frac{i}{h_o - 1}\right) \hat{s}_i^{\text{left}} + \left(\frac{i}{h_o - 1}\right) \hat{s}_i^{\text{right}}, \quad i = 0, \ldots, h_o - 1 \tag{18}$$

where $\hat{s}_i^{\text{left}}$ and $\hat{s}_i^{\text{right}}$ are the refined windows for the same overlap position. For stable generation, we follow the sizes of the window and the overlapping band in Luo et al. (2025).

4 Experiments

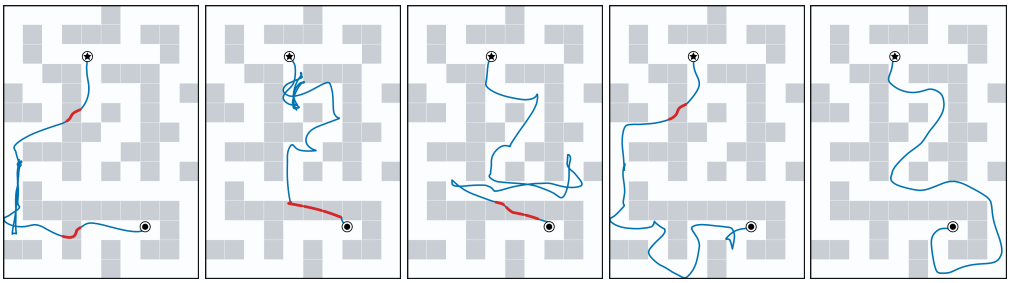
In this section, we evaluate our method on long-horizon composition planning tasks from OGBench (Park et al., 2025). Our experiments focus on three aspects: (1) whether our method reduces infeasible transitions between composed trajectory segments, (2) whether improved plan feasibility leads to higher task success rate as the planning horizon and dynamic complexity increase, and (3) how each component accounts for the observed improvements.

Datasets and Environments. We consider the stitch datasets of PointMaze, AntMaze, and HumanoidMaze, each evaluated on the Medium, Large, and Giant maze layouts. The stitch datasets consist of short trajectory segments that do not individually connect the start and goal states of evaluation settings. Therefore, solving these tasks requires composing locally observed behaviors into much longer trajectories.

Baselines. We compare against offline goal-conditioned RL (GCRL) and diffusion-based compositional planning methods. The offline GCRL baselines include goal-conditioned behavioral cloning (GCBC, Lynch et al., 2020), goal-conditioned implicit V-learning and Q-learning (GCIVL, GCIQL, Kostrikov et al., 2022), Quasimetric RL (QRL, Wang et al., 2023), Contrastive RL (CRL, Eysenbach et al., 2022), and Hierarchical Implicit Q-Learning (HIQL, Park et al., 2023). The compositional planning baselines consist of Generative Skill Chaining (GSC, Mishra et al., 2023), CompDiffuser (CD, Luo et al., 2025), CDGS (Mishra et al., 2026), ECD (Sun et al., 2026), and RCD (Lee et al., 2026). We use the same pretrained compositional diffusion model and evaluate SCD on OGBench, following the same evaluation protocol for all diffusion-based methods.

Table 1: **Success rates (%) on OGBench Stitch maze tasks.** Each environment provides 5 goal-conditioned tasks and we evaluate 20 episodes per task for three random seeds.

Family	Method	PointMaze			AntMaze			HumanoidMaze			Avg.
		Medium	Large	Giant	Medium	Large	Giant	Medium	Large	Giant	
Offline GCRL	GCBC	23±18	7±5	0±0	45±11	3±3	0±0	29±5	6±3	0±0	12.56
	GCIVL	70±14	12±6	0±0	44±6	18±2	0±0	12±2	1±1	0±0	17.44
	GCIQL	21±9	31±2	0±0	29±6	7±2	0±0	12±3	0±0	0±0	11.11
	QRL	80±12	84±15	50±8	59±7	18±2	0±0	18±2	3±1	0±0	34.67
	CRL	0±1	0±0	0±0	53±6	11±2	0±0	71±3	6±1	0±0	15.67
	HIQL	74±6	13±6	0±0	94±1	67±5	21±2	**96±4**	31±3	12±2	45.33
Diffusion planners	GSC	**100±0**	**100±0**	29±3	**97±2**	66±2	20±1	92±1	70±3	5±1	64.33
	CD	**100±0**	**100±0**	68±3	96±2	86±2	65±3	91±1	72±3	67±4	82.78
	CDGS	**100±0**	**100±0**	87±3	**97±1**	88±2	85±3	93±1	74±2	55±3	86.56
	ECD	**100±0**	**100±0**	84±2	**97±1**	82±1	82±6	92±2	64±4	49±1	83.33
	RCD	**100±0**	**100±0**	**100±0**	**97±2**	91±2	**89±2**	93±2	79±3	62±3	90.11
	Ours	**100±0**	**100±0**	91±1	95±1	**92±1**	85±2	95±1	**89±2**	**83±3**	**92.22**



(a) CompDiffuser (b) CDGS (c) ECD (d) RCD (e) **SCD (Ours)**

Figure 4: **Qualitative results of long-horizon planning in HumanoidMaze-Giant.** SCD successfully finds feasible connections among the chunk candidates. See Fig. 13 for more results.

4.1 Evaluating Plan Feasibility

First, we examine whether our method produces feasible trajectories. Following Lee et al. (2026), we sample 20 plans for the five tasks of OGBench HumanoidMaze-Giant-Stitch, and count the plans staying outside the wall without making any shortcuts. This environment requires at least four-times longer horizons than PointMaze and AntMaze, and it is comparatively hard to retain mode consistency. As shown in Tab. 2 and the qualitative result Fig. 4, our method consistently improves the feasibility, and this gain holds for all reported tasks.

Table 2: **The ratio of geometrically feasible plans.** For each HumanoidMaze-Giant task, we sample 20 plans and report the fraction whose trajectories are geometrically feasible, not passing through walls. The numerical entries in the header row denote the task IDs.

Method	01	02	03	04	05	Avg.
CD	30	5	15	10	40	20
CDGS	0	30	0	0	45	15
ECD	45	45	20	25	50	37
RCD	25	30	40	10	60	33
SCD (Ours)	**95**	**80**	**55**	**75**	**75**	**76**

4.2 Solving Long-Horizon Tasks

We examine whether our method remains effective as the required composition horizon increases. We consider OGBench stitch tasks of PointMaze, AntMaze, and HumanoidMaze across Medium, Large, and Giant maze sizes. Larger mazes require longer paths and more local segments to be composed. This may increase the chance of generating an infeasible connection.

In Tab. 1, offline GCRL methods show degraded performance as the maze size increases. Diffusion-based planners perform better across most tasks, nearly saturated on PointMaze-Medium and Large. However, their performance declines on longer compositions, particularly in HumanoidMaze-Giant.

Our method remains competitive in the saturated settings while improving certain long-horizon tasks. It achieves success rates of 91%, 85%, and 83% in Giant PointMaze, AntMaze, and Hu-

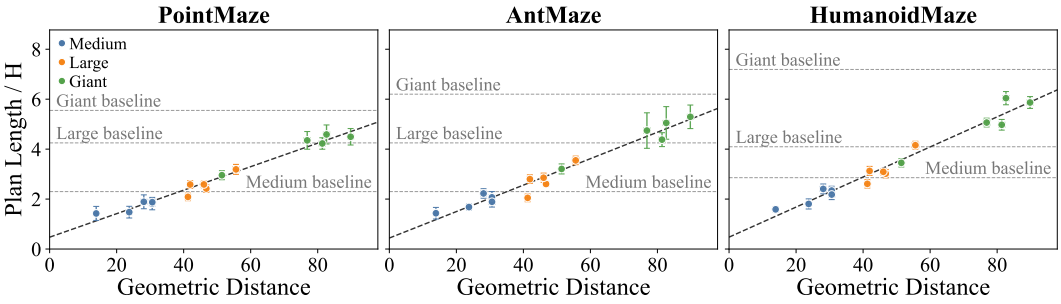


Figure 5: **Adaptive plan length of SCD.** Each point represents one evaluation task, and five tasks are evaluated for each maze size. The x-axis represents a geometric distance of the shortest path and the y-axis represents the length of the generated plan divided by the planner horizon (H). Horizontal dashed lines indicate the fixed number of chunks used by compositional planners. Error bars show the standard deviation over 20 plans.

manoidMaze tasks, respectively. These results support that our reachability score reliably distinguishes composable segment pairs and that graph-based assembly produces plans with more likely connections.

4.3 Adaptive Planning Length and Planning Cost

The appropriate trajectory length can vary across start and goal state pairs depending on their distances and difficulty. Our method supports adaptive planning by selecting the chunks based on the executability and the trajectory length, which can dynamically alter the number of selected chunks.

As shown in Fig. 5, the resulting trajectory lengths closely follow the distance of the shortest path between the start and goal states, regardless of the size of the environment. The length of plans depends on the distance rather than a predefined number of chunks, and SCD can find an appropriate length for each task to execute.

Beyond sampling candidate trajectories, our method evaluates candidate connections, searches for the minimum-cost path, and rebuilds the bridges. To quantify the overhead, we measure and compare wall-clock planning time on AntMaze-Giant and HumanoidMaze-Giant with a single RTX 4090. We use the same pretrained compositional model for every method in Tab. 3. We observe that connection scoring and graph search require a small additional time cost: each candidate is encoded once, and the reachability model evaluates all state pairs through a single matrix multiplication. Only the selected bridges require additional denoiser calls, and they cost much less than generating whole candidates. Since we produce candidates in parallel, SCD runs faster than the listed methods except for CD (parallel).

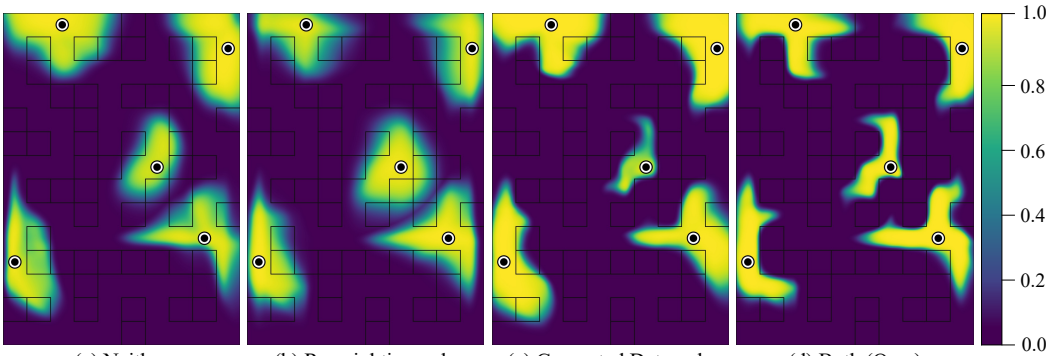
Table 3: **Planning cost comparison.** Wall-clock planning times and success rates on AntMaze-Giant and HumanoidMaze-Giant. We report mean time ± std across 10 plans on a single RTX 4090.

Method	AntMaze		HumanoidMaze	
	Succ.	**Time(s)**	**Succ.**	**Time(s)**
CD (parallel)	48±1	3.09±0.02	57±3	7.58±0.02
CD (interleave)	65±3	6.31±0.01	67±4	13.08±0.03
CDGS	85±3	15.83±0.03	55±3	20.33±0.03
ECD	82±6	9.24±0.01	49±1	18.71±0.03
RCD	**89±2**	4.52±0.01	62±3	11.07±0.04
SCD (Ours)	85±2	**4.24±0.01**	**83±3**	**9.08±0.04**

4.4 Ablation Study

Since a single infeasible connection may ruin the entire trajectory, avoiding false-positive edges is important for executable routes. We therefore evaluate whether the classifier retains reachable states and rejects unreachable ones.

We examine three design choices: using generated states, density reweighting, and the neighborhood size k used by the k-NN density estimator. We evaluate each classifier on a controlled test set. Further details on the evaluations are provided in App. B.3.3.



(a) Neither (b) Reweighting only (c) Generated Data only (d) Both (Ours)

Figure 6: **Reachability heatmap comparison.** $D_\phi(x, y, h_o)$ is evaluated for five anchor states (circles) in AntMaze Giant, under four combinations of the ablation study in Tab. 4. Including generated states reduces overconfident predictions in regions that are unseen in the offline dataset. Density reweighting removes the start state bias induced by nonuniform sampling. Together, the two components produce reachability maps that better respect walls and traversable corridors.

Tab. 4 shows that using generated states and reweighting improve both acceptance of feasible pairs and rejection of infeasible pairs. Without our approach, the score is affected by the sampling density of the generated pool, which is concentrated in the interior of the maze. The classifier then becomes overconfident near the corner and underconfident near the center. Fig. 6 shows this effect, and high-probability regions past walls are also observed without reweighting. Results were insensitive to k over the range we tested.

Table 4: **Conditional reachability benchmark.** Ranking acceptance rates between feasible (positive) and infeasible (negative) pair scores: positives above the negative 95th percentile and negatives above the positive 95th percentile. We report mean ± std across 3 seeds.

Generation	Reweight	k	Feas. Accept. ↑	Infeas. Accept. ↓
✗	✗	–	31±1	0.49±0.14
✗	✓	50	40±2	0.07±0.02
✓	✗	–	47±3	0.05±0.05
✓	✓	10	73±1	**0.00±0.00**
✓	✓	50	**74±0**	**0.00±0.00**
✓	✓	200	**74±1**	**0.00±0.00**
✓	✓	1000	**74±0**	**0.00±0.00**

5 Conclusion

We presented a Selective Compositional Diffuser (SCD), which selects composable chunks to form an executable trajectory. The method reformulates the composition problem of compositional diffusion planners as a chunk selection problem over generated chunks. We train a lightweight MLP network in a self-supervised manner to measure reachability. We use generated chunks jointly when training the reachability classifier to reduce the train-test mismatch, and propose the reweighting term for debiased estimation. After generating the candidate plans from diffusion models, we construct a graph based on the generated chunks and add edges weighted with the reachability scores. We find the minimum-cost chunk sequences by traversing the graph, generate bridging trajectories, and refine the full trajectory. We observe that this solution reduces incompatible mode compositions and mode-averaging issues. Across OGBench stitching tasks, SCD improves performance, reaching state-of-the-art success rates on HumanoidMaze-Large and Giant tasks.

Limitations and future work. Our method depends on the coverage of candidate chunks and the accuracy of the reachability classifier. If the pool contains no start-to-goal route connectable within the bridge budget, selection cannot produce a feasible plan. The reachability classifier is trained with k-NN density reweighting, whose errors may introduce false-positive edges that graph search then selects, particularly in high-dimensional state spaces. Pairwise edge evaluation also scales quadratically with the number of candidate vertices, limiting the size of the pool. Future work could improve coverage by generating new chunks around low-score connections, improve the reachability classifier with learned density estimates, and reduce the edge evaluation cost with approximate nearest-neighbor search.

AI USE STATEMENT

In this work, we used generative AI tools to polish writing. We have not used generative AI tools for proving mathematical claims, generating synthetic data, drafting sections of the paper, research ideation or execution, or retrieval and discovery. We have reviewed all AI-assisted work. In particular, the authors reviewed and revised all AI-polished writing. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

The contribution focuses on reformulating the mode-averaging problem as a chunk selection problem, with a modified reachability classification objective. The ethical and social implications are therefore aligned with those of existing generative diffusion planning research. These include concerns about the safety of executing generated plans on physical robots and the potential misuse of autonomous planning systems. All experiments use only publicly available datasets (Park et al., 2025) and codebases (Luo et al., 2025), credited by citation and used under their licenses in App. E. We collect no new data and conduct no human-subject research.

REPRODUCIBILITY STATEMENT

Sec. 3 describes SCD, and Alg. 1 gives the inference procedure in pseudocode. Apps. A and B provide the implementation details, hyperparameters, evaluation protocols, and computational resources. The sweeps used to select the hyperparameters are included in App. C.

REFERENCES

Anurag Ajay, Yilun Du, Abhi Gupta, Joshua B. Tenenbaum, Tommi S. Jaakkola, and Pulkit Agrawal. Is conditional generative modeling all you need for decision making? In *The Eleventh International Conference on Learning Representations*, 2023.

Seungho Baek, Taegeon Park, Jongchan Park, Seungjun Oh, and Yusung Kim. Graph-assisted stitching for offline hierarchical reinforcement learning. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 2391–2408. PMLR, 13–19 Jul 2025.

Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin CM Burchfiel, and Shuran Song. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion. In *Proceedings of Robotics: Science and Systems*, Daegu, Republic of Korea, July 2023. doi: 10.15607/RSS.2023.XIX.026.

Prafulla Dhariwal and Alexander Nichol. Diffusion models beat gans on image synthesis. In *Advances in Neural Information Processing Systems*, volume 34, pp. 8780–8794. Curran Associates, Inc., 2021.

Ben Eysenbach, Russ R Salakhutdinov, and Sergey Levine. Search on the replay buffer: Bridging planning and reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Benjamin Eysenbach, Ruslan Salakhutdinov, and Sergey Levine. C-learning: Learning to achieve goals via recursive classification. In *International Conference on Learning Representations*, 2021.

Benjamin Eysenbach, Tianjun Zhang, Sergey Levine, and Ruslan Salakhutdinov. Contrastive learning as goal-conditioned reinforcement learning. In *Advances in Neural Information Processing Systems*, 2022.

Yaniv Hassidof, Adir Morgan, Yilun Du, and Kiril Solovey. Plan first, diffuse later: Extrinsic graph guidance for long-horizon diffusion planning, 2026.

Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc., 2020.

Michael Janner, Yilun Du, Joshua Tenenbaum, and Sergey Levine. Planning with diffusion for flexible behavior synthesis. In *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9902–9915. PMLR, 17–23 Jul 2022.

Ilya Kostrikov, Ashvin Nair, and Sergey Levine. Offline reinforcement learning with implicit q-learning. In *International Conference on Learning Representations*, 2022.

Kyowoon Lee, Yunhao Luo, Anh Tong, and Jaesik Choi. Refining compositional diffusion for reliable long-horizon planning, 2026.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.

Yunhao Luo, Utkarsh Aashu Mishra, Yilun Du, and Danfei Xu. Generative trajectory stitching through diffusion composition. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

Corey Lynch, Mohi Khansari, Ted Xiao, Vikash Kumar, Jonathan Tompson, Sergey Levine, and Pierre Sermanet. Learning latent plans from play. In *Conference on robot learning*, pp. 1113–1132. Pmlr, 2020.

Lina Mezghani, Sainbayar Sukhbaatar, Piotr Bojanowski, Alessandro Lazaric, and Karteek Alahari. Learning goal-conditioned policies offline with self-supervised reward shaping. In *6th Annual Conference on Robot Learning*, 2022.

Utkarsh Mishra, David He, Yongxin Chen, and Danfei Xu. Compositional diffusion with guided search for long-horizon planning. In *International Conference on Learning Representations*, volume 2026, pp. 68308–68322, 2026.

Utkarsh Aashu Mishra, Shangjie Xue, Yongxin Chen, and Danfei Xu. Generative skill chaining: Long-horizon skill planning with diffusion models. In *Proceedings of The 7th Conference on Robot Learning*, volume 229 of *Proceedings of Machine Learning Research*, pp. 2905–2925. PMLR, 06–09 Nov 2023.

Evgenii Opryshko, Junwei Quan, Claas A Voelcker, Yilun Du, and Igor Gilitschenski. Test-time graph search for goal-conditioned reinforcement learning. In *Forty-third International Conference on Machine Learning*, 2026.

Seohong Park, Dibya Ghosh, Benjamin Eysenbach, and Sergey Levine. HIQL: Offline goal-conditioned RL with latent states as actions. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Seohong Park, Kevin Frans, Benjamin Eysenbach, and Sergey Levine. OGBench: Benchmarking offline goal-conditioned RL. In *The Thirteenth International Conference on Learning Representations*, 2025.

Nikolay Savinov, Anton Raichuk, Damien Vincent, Raphael Marinier, Marc Pollefeys, Timothy Lillicrap, and Sylvain Gelly. Episodic curiosity through reachability. In *International Conference on Learning Representations*, 2019.

Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2021.

Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019.

Tao Sun, Utkarsh Aashu Mishra, Jiaxin Lu, Danfei Xu, and Iro Armeni. Energy-based compositional diffusion planning. In *Forty-third International Conference on Machine Learning*, 2026.

Tongzhou Wang, Antonio Torralba, Phillip Isola, and Amy Zhang. Optimal goal-reaching reinforcement learning via quasimetric learning. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 36411–36430. PMLR, 23–29 Jul 2023.

Jaesik Yoon, Hyeonseo Cho, and Sungjin Ahn. Compositional monte carlo tree diffusion for extendable planning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025.

A Implementation Details

A.1 Algorithm

SCD constructs a reachability graph over the candidate chunks (Stages 1 and 2), finds a minimum-cost path (Stage 3), and converts the path into the final trajectory (Stages 4 and 5). ParallelSample samples N plans of K chunks in parallel with CompDiffuser, while q_s and q_g are inpainted. Bridge generates the h_o intermediate states while conditioning M neighboring states of each adjacent chunk by inpainting. Noise noises a trajectory to the noise level κ. Denoise applies the last j-th DDIM step in a windowed manner and blends the window overlaps linearly while keeping q_s and q_g fixed. The procedure is repeated at every replan (App. B.3.2).

Algorithm 1 Selective Compositional Diffuser

Require: The start and goal states q_s, q_g, the classifier D_ϕ with the key and query encoders K_θ, Q_ψ, the number of plans N, the number of chunks in each plan K, the boundary stride S, the minimum segment length M, the discount factor γ, the maximum lag $L_{\max}$, the bridge budget h_o, the number of bridge candidates N_β, and the refinement level κ.

Stage 1: Candidate sampling

$\{\tau_{1:K}^{(n)}\}_{n=1}^{N} \leftarrow \textsc{ParallelSample}(q_s, q_g; N, K)$ ▷ Inpaint $s_1^{(n,1)} = q_s$, $s_H^{(n,K)} = q_g$

$\tau_i \leftarrow \tau_k^{(n)} = s_{1:H}^{(i)} \quad \forall i = (n,k) \in \mathcal{I} = [1,N] \times [1,K]$

Stage 2: Graph construction

$\mathcal{T} \leftarrow \{1, 1+S, 1+2S, \dots\} \cup \{H\}$ ▷ Boundary stride S

$\mathcal{V} \leftarrow \{v_0, v_g\} \cup \{\mathrm{IN}(i,\ell) : i \in \mathcal{I},\ \ell \in \mathcal{T}\} \cup \{\mathrm{OUT}(i,r) : i \in \mathcal{I},\ r \in \mathcal{T}\}$ ▷ Vertices

$c \leftarrow \infty$ ▷ Initialize all costs

$c\big(v_0, \mathrm{IN}(i,1)\big) \leftarrow 0 \quad \forall i = (n,1)$ ▷ For the root node v_0

$c\big(\mathrm{OUT}(i,H), v_g\big) \leftarrow 0 \quad \forall i = (n,K)$ ▷ For the goal node v_g

for $i \in \mathcal{I}, \ell, r \in \mathcal{T}, r - \ell \geq M$ **do** ▷ For the segment edges

  $\ell = t_0 < t_1 < \cdots < t_P = r$ s.t. $t_j - t_{j-1} \leq L_{\max}$

  $c\big(\mathrm{IN}(i,\ell), \mathrm{OUT}(i,r)\big) \leftarrow -\sum_{j=1}^{P} \log D_\phi\big(s_{t_{j-1}}^{(i)}, s_{t_j}^{(i)}, t_j - t_{j-1}\big) - (r-\ell)\log\gamma$

end for

$\mathbf{Z}_{\mathrm{q}} \leftarrow \big[Q_\psi\big(s_r^{(i)}, \mathrm{PE}(h_o)\big)\big]_{i \in \mathcal{I},\, r \in \mathcal{T}}$ ▷ Cached queries

$\mathbf{Z}_{\mathrm{k}} \leftarrow \big[K_\theta(s_\ell^{(i)}) / \|K_\theta(s_\ell^{(i)})\|_2\big]_{i \in \mathcal{I},\, \ell \in \mathcal{T}}$ ▷ Cached keys

$\mathbf{F} \leftarrow \mathbf{Z}_{\mathrm{q}} \mathbf{Z}_{\mathrm{k}}^{\top}$ ▷ Logits of all possible bridges

for $i, i' \in \mathcal{I}$ s.t. $i \neq i', r, \ell' \in \mathcal{T}$ **do** ▷ Bridge edges

  $c\big(\mathrm{OUT}(i,r), \mathrm{IN}(i',\ell')\big) \leftarrow -\log\sigma\big(\mathbf{F}_{(i,r),(i',\ell')}\big) - h_o \log\gamma$

end for

$\mathcal{G} \leftarrow (\mathcal{V}, c)$ ▷ Reachability graph with non-negative edges

Stage 3: Path search

$(v_0, v_1, \dots, v_J) \leftarrow \textsc{Dijkstra}(\mathcal{G}; v_0, v_g)$ ▷ Find the path of $\min \sum_{j=0}^{J-1} c(v_j, v_{j+1})$

$A_t \leftarrow s_{\ell_t:r_t}^{(i_t)}$ for $(v_{2t-1}, v_{2t}) = \big(\mathrm{IN}(i_t,\ell_t), \mathrm{OUT}(i_t,r_t)\big)$ ▷ $T = (J-1)/2$

Stage 4: Bridge generation

for $t = 1, \dots, T-1$ **do**

  $\{\beta_t^{(b)}\}_{b=1}^{N_\beta} \leftarrow \textsc{Bridge}\big(s_{r_t-M+1:r_t}^{(i_t)},\ s_{\ell_{t+1}:\ell_{t+1}+M-1}^{(i_{t+1})};\ h_o\big)$

  $\beta_t \leftarrow \arg\min_b \mathcal{E}\big(\beta_t^{(b)}\big)$ ▷ Eq. (19)

end for

$\tilde{\tau} \leftarrow (A_1, \beta_1, A_2, \dots, \beta_{T-1}, A_T)$

Stage 5: Refinement

$\tilde{\tau} \leftarrow \textsc{Noise}(\tilde{\tau}, \kappa)$

for $j = \kappa, \dots, 1$ **do**

  $\tilde{\tau} \leftarrow \textsc{Denoise}(\tilde{\tau}, j;\ q_s, q_g)$ ▷ Windowed refinement in Sec. 3.4

end for

return $\tilde{\tau}$

Table 5: **Hyperparameters of the released CompDiffuser.**

Component	PointMaze	AntMaze	HumanoidMaze
Diffusion model			
Planner dim.	2D	2D	2D
U-Net base dim.	128	128	128
U-Net dim. multipliers	$(1, 2, 4, 8)$	$(1, 2, 4, 8)$	$(1, 2, 4, 8)$
Boundary encoder	CNN	CNN	CNN
Boundary encoding dim.	256	256	256
Controller			
Type	PD controller	MLP inverse dynamics	MLP inverse dynamics
PD gain	5.0	–	–
MLP hidden sizes	–	$(512, 512, 512)$	$(512, 1024, 1024, 512, 256)$

A.2 Method Implementation

Pretrained Components. For each environment, we use the pretrained compositional diffusion models and controllers released by CompDiffuser (Luo et al., 2025). All diffusion-based methods in our experiments share these checkpoints (App. B.2). The diffusion model is a 1D U-Net that denoises trajectories over xy states with separate CNN encoders for boundary conditions. These conditions come from either the start and goal conditions or the overlapping parts of neighboring chunks. We sample it with DDIM using 50 steps. For execution, PointMaze uses a PD controller with PD gain 5.0. AntMaze and HumanoidMaze use MLP inverse-dynamics models that map the full observation and the current xy subgoal to an action. Tab. 5 summarizes the architectures.

Generated State Bank. For classifier training, we build a state bank with states from generated chunks. We use the same pretrained diffusion model for this state bank. Following RCD (Lee et al., 2026), we compute the self-reconstruction error of each generated sample to keep low-density samples over the offline distribution. We noise each generated sample to a probe level, obtain two independent noised samples, denoise each once, compute the squared error between the original sample and reconstructions, and use the average of two error values. We keep generated samples when this self-reconstruction error exceeds the 90th percentile of the errors, computed over the offline validation data. For the probe level, we use 35% of the denoising steps of the training setting, $179 = \lfloor 512 \times 0.35 \rfloor$ for AntMaze and PointMaze, $350 = \lfloor 1000 \times 0.35 \rfloor$ for HumanoidMaze, following the RCD setting. We note that this filtering aims to make chunks more distinguishable over offline and generated chunks for better classifier training, and we do not use this strategy for SCD at inference time. See Fig. 11 for samples of generated trajectories.

Table 6: **Training settings** for the reachability classifier.

Setting	Value
Query encoder Q_ψ	3-layer MLP, width 512, SiLU
Key encoder K_θ	3-layer MLP, width 512, SiLU
Lag encoding	Sinusoidal positional encoding, 64-dim.
Maximum lag $L_{\max}$	56 by default / 128 for HumanoidMaze
Lag sampling	Uniform $\mathcal{U}[1, L_{\max})$ for 75%, $L_{\max}$ for 25%
k-NN reweighting	$k = 50$, $w_{\min} = 10^{-8}$, $w_{\max} = 20$
Optimizer	AdamW (Loshchilov & Hutter, 2019)
Learning rate η	3×10^{-4}
Momentum parameters (β_1, β_2)	(0.9, 0.999)
Weight decay λ	0.01
Epsilon ϵ	10^{-8}
Learning rate scheduler	Cosine decay without warmup ($T_{\max} = 100$K)
Batch size	8192 for the positive batch, 8192 for the additional negative batch
Training steps	100K
Data split (train / valid / test)	80 / 10 / 10

Reachability Classifier. We train a reachability classifier for each environment. The settings follow Tab. 6, which are shared across environments except $L_{\max}$. The query encoder Q_ψ takes x with a

sinusoidal encoding of L, and the key encoder K_θ takes y. To build an observed pair, we first sample a lag $L \in [1, L_{\max}]$, where the bridge budget $h_o = L_{\max}$ in Sec. 3.3. We then pick the target y_{obs} uniformly from the next L states of x_{obs} in the same offline trajectory. The two states of a negative pair are sampled independently from the bank, and the weights use the k-NN distances.

For k-NN estimation, we normalize xy into the range $[-1, 1]$ with min-max normalization. We use FAISS's `IndexIVFFlat` with `IndexFlatL2` for efficient approximate nearest-neighbor search.

Table 7: **Default settings** for candidate generation, graph search, and trajectory refinement. We present environmental settings such as the chunk length H and the bridge budget h_o in Tab. 8.

Setting	Value
Baseline	CompDiffuser, sample chunks in parallel (without interleaving)
Scheduler	50-step DDIM (Song et al., 2021)
Minimum segment length M	48 by default / 80 for HumanoidMaze (i.e., $M = H - 2h_o$)
Boundary stride S	24 by default / 40 for HumanoidMaze
Discount factor γ	0.999
Shortest-path solver	Dijkstra's algorithm (SciPy, CPU)
Bridge candidates N_β	5
Refinement noise level κ	Last five DDIM steps

Planning. We sample N candidate plans of K chunks with the parallel sampler of CompDiffuser, which denoises all chunks of each plan together. The first state of the first chunk and the last state of the last chunk are inpainted with q_s and q_g. We set candidate boundaries at a fixed stride of S, such as $\ell \in \{1, 25, 49, 73, \ldots\}$ (see Fig. 3 in Sec. 3.3). If a segment is longer than $L_{\max}$, we partition the segment into subparts, equal to or shorter than $L_{\max}$, and accumulate costs of each subpart. We set a root vertex as the starting point q_s, connecting to each first chunk with zero cost, and run Dijkstra's algorithm to find the minimum-cost path from the root vertex to the goal state vertex.

We generate each bridge with DDIM (Song et al., 2021) for 50 sampling steps. We inpaint M states on each side with the neighboring states from the adjacent chunks, and denoise the h_o bridging states between them. If there are remaining states ($H > h_o + 2M$), we leave both ends untouched until denoising:

$$\underbrace{\big|\ \text{untouched}\ \big|\ M\ \text{inpainted}\ \big|\ h_o\ \text{mainly-denoised}\ \big|\ M\ \text{inpainted}\ \big|\ \text{untouched}\ \big|}_{H}.$$

We draw N_β candidate samples per bridge and measure the mismatch of each bridge candidate β as

$$\mathcal{E}(\beta) = \sum_{i=0}^{M-1} \frac{i}{M-1} \left\| \hat{s}_i^L - s_i^L \right\|^2 + \sum_{i=0}^{M-1} \frac{M-1-i}{M-1} \left\| \hat{s}_i^R - s_i^R \right\|^2, \tag{19}$$

where s_i^L and s_i^R are the i-th inpainted states, and $\hat{s}_i^L$ and $\hat{s}_i^R$ are their clean predictions at the last DDIM step. We select the candidate with the smallest $\mathcal{E}(\beta)$. Finally, we noise the assembled trajectory to the last five DDIM steps and denoise it with the same windows and overlaps as the initial plan generation. We blend the overlaps linearly while keeping q_s and q_g fixed, as introduced in Eq. (18). This planner outputs a sequence of waypoints, one per environment step, which the controllers execute as described in App. B.3.2. Detailed parameters follow Tab. 7.

B Experimental Details

B.1 Environment Details

We evaluate on the OGBench Stitch datasets for PointMaze, AntMaze, and HumanoidMaze, each at Medium, Large, and Giant maze sizes. Each trajectory covers only local transitions of the maze, whereas evaluation requires reaching start–goal pairs across longer distances. The planner must compose behaviors beyond the horizon of individual training trajectories. The dataset identifiers follow the pattern `<env>-<size>-stitch-v0` (e.g., `antmaze-giant-stitch-v0`), and all planners operate in the 2D xy space. Tab. 8 lists the episode limit of each environment. See Fig. 10 for samples of offline stitch data.

B.2 Baselines

The baseline scores in Tab. 1 are drawn from published results. Scores for GCBC, GCIVL, GCIQL, QRL, CRL, HIQL, GSC, and CD are taken from CompDiffuser (Luo et al., 2025), and scores from CDGS, ECD, and RCD are taken from their respective papers (Mishra et al., 2026; Sun et al., 2026; Lee et al., 2026). For the feasibility and planning cost analyses in Tabs. 2 and 3, we run the publicly released implementations of CDGS, ECD, and RCD with their default settings and the same pretrained local diffusion model, allowing us to compare their plans and runtimes under a shared model.

In Tab. 3, *CD (parallel)* denoises all chunks simultaneously, whereas *CD (interleave)* sequentially denoises chunks, allowing later chunks to condition on previously updated neighbors.

B.3 Evaluation Details

All experiments, including the planning-time measurements in Tab. 3, are conducted on a single NVIDIA RTX 4090 GPU. SCD uses the pretrained diffusion model of CompDiffuser (Luo et al., 2025) without further training, and only the reachability classifier is trained for our method.

B.3.1 Feasibility Evaluation

We measure the geometric feasibility of planned paths. For each of the five HumanoidMaze-Giant-Stitch tasks, we generate 20 plans per method, yielding 100 plans per method. Tab. 2 reports the percentage of feasible plans for each task and the mean across tasks.

We evaluate geometric feasibility by checking whether any line segment between consecutive planned positions crosses a maze wall. Each wall is represented by a 4×4 square, contracted inward by δ on each side to tolerate small boundary incursions. We use $\delta = 0.3$ for Tab. 2. To detect paths passing between adjacent walls, we also check whether they intersect the line segments connecting the walls' centers. We count a plan as feasible only if no segment between consecutive xy positions intersects a contracted wall or one of these lines. Fig. 7 reports the feasible-plan ratio as δ varies; SCD remains highest throughout the tested range.

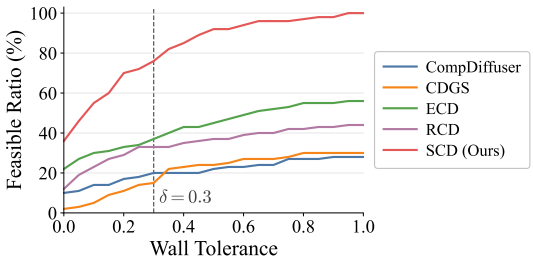


Figure 7: **Feasible ratio over wall tolerance.** Feasible ratio of the plans in Tab. 2 as a function of the tolerance δ. The dashed line marks $\delta = 0.3$ used in Tab. 2.

B.3.2 Main Evaluation

We evaluate task success on nine OGBench Stitch environments in Tab. 8. Each environment provides five evaluation tasks. Each task consists of 20 fixed start–goal pairs, and we run one episode per pair with each of three random seeds, yielding 300 rollouts per environment in total. The seeds control only the stochastic sampling during planning, and the start–goal pairs are shared across seeds. An episode is successful if the environment reports that the xy distance between the agent and the goal is at most ϵ within the episode limit. OGBench sets $\epsilon = 1.0$ for PointMaze and $\epsilon = 0.5$ for AntMaze and HumanoidMaze. We average the success rates of the five tasks within each seed and report the mean and the standard deviation across seeds in Tab. 1. Baseline results in that table are borrowed from the sources listed in App. B.2.

We execute planned trajectories with the controllers in App. A.2, which are shared by all diffusion-based methods. At each step, the controller takes the waypoint of the plan at the current time step as a subgoal and predicts an action from the current state and the subgoal. Following CompDiffuser, we replan from the current state to the original goal when the Euclidean distance to the subgoal exceeds a threshold, which differs by environment, or when the agent runs out of remaining subgoals. We also follow its schedule that reduces the number of chunks K at each replan. Tab. 8 lists the threshold and the maximum number of replans for each environment, together with other planner settings. Once the replanning budget is exhausted, the agent tracks the last waypoint until the episode limit.

Table 8: **Planning and evaluation settings per environment.** *Max steps* denotes the episode limit. H denotes the chunk horizon, h_o the overlap length of the pretrained planner, K the number of chunks per candidate plan, and N the number of candidate plans. *Replan thr.* and *Max replans* denote the tracking threshold and the maximum number of replans. All values follow the released configuration of CompDiffuser (Luo et al., 2025).

Environment	Size	Max steps	H	h_o	K	N	Replan thr.	Max replans
PointMaze	Medium	1000	160	56	3	40	1	0
	Large	1000	160	56	6	40	1	0
	Giant	1000	160	56	8	40	1	10
AntMaze	Medium	1000	160	56	3	40	4	10
	Large	2000	160	56	6	40	4	10
	Giant	2000	160	56	9	40	4	10
HumanoidMaze	Medium	5000	336	128	4	40	10	10
	Large	5000	336	128	6	40	10	10
	Giant	8000	336	128	11	40	10	10

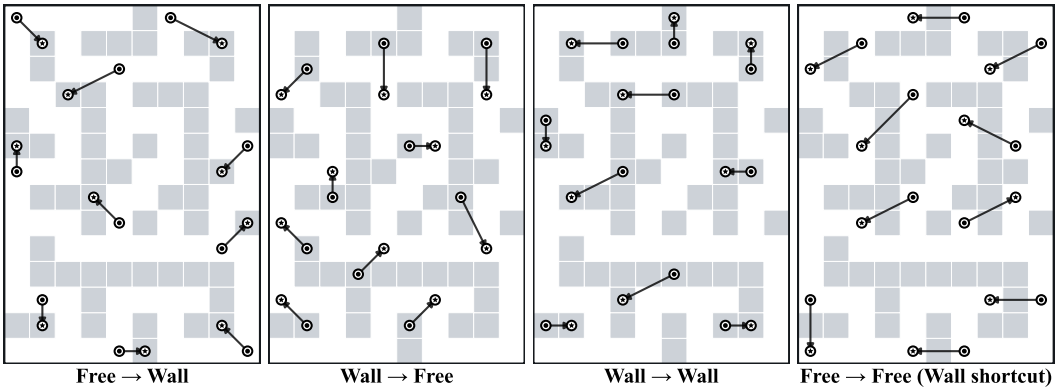


Figure 8: **Examples of the four infeasible pair types** in the reachability test set (AntMaze-Giant). Each arrow points from a source (circle) to a target (star), and gray cells denote walls.

For PointMaze-Giant, increasing the replanning threshold from 1 to 4 raises SCD's success rate from 91 ± 1 to 99 ± 1. However, we retain the threshold of 1 in Tab. 1 for fair comparison.

B.3.3 Reachability Evaluation

We evaluate the reachability classifier on a fixed test set in AntMaze-Giant with lags $L \leq 56$. The test set consists of 400 feasible and 400 infeasible base start-to-goal state pairs on the maze grid. Each feasible base pair takes the cells of an offline state pair (s_t, s_{t+L}) that is not used for training. The infeasible pairs consist of four types with 100 pairs each: free-to-wall, wall-to-free, wall-to-wall pairs, and free-to-free pairs. In particular, free-to-free pairs are close in Euclidean distance but far in geodesic distance. For each base pair, we sample the source and target 100 times uniformly, yielding 40,000 feasible and 40,000 infeasible pairs. Fig. 8 shows examples of the four infeasible types.

We compute the classifier logit $f(x, y, L)$ for all pairs and set two thresholds: t_{neg}, the 95th percentile of the infeasible scores, and t_{pos}, the 95th percentile of the feasible scores. *Feas. Accept.* in Tab. 4 is the percentage of feasible pairs scoring above t_{neg}, and *Infeas. Accept.* is the percentage of infeasible pairs scoring above t_{pos}.

Generation ✗ samples the negative endpoints only from the offline dataset states, whereas ✓ samples them from both the offline dataset states and the generated states. *Reweight* ✗ sets all reweighting terms to $w(x) = 1$, and ✓ uses the clipped inverse-density weights in Sec. 3.2 estimated with k-NN distances. All variants share the same architecture and training setup in App. A. We train each variant with three random seeds and report the mean and the standard deviation across seeds. For Fig. 6, we choose five start states $s_1, \ldots, s_5$ outside the walls that are far enough apart that no state

is reachable from more than one of them. We then compute $D_\phi(s_i, y, L)$ with $L = 56$ for every state y in the maze and show $\max_i D_\phi(s_i, y, L)$.

C Ablation Analysis

We conduct an ablation study of hyperparameters on AntMaze-Giant in Tab. 9. We follow the evaluation protocols in App. B.3.2 and use default settings for all other hyperparameters.

Table 9: **Ablation of hyperparameters on AntMaze-Giant.** Each subtable varies one hyperparameter. We report success rates (%) as mean ± std across three seeds.

(a) Bridge samples

N_β	Succ.
1	83.0±2.2
3	84.0±2.2
5	**85.3±1.7**

(b) Boundary stride

S	Succ.
12	82.0±3.6
24	**85.3±1.7**
Endpoints	85.0±2.8

(c) Refinement steps

κ	Succ.
0	84.3±2.1
5	**85.3±1.7**
10	81.3±1.3

(d) Candidates

N	Succ.
10	79.3±5.6
20	81.7±5.0
40	**85.3±1.7**
80	84.7±0.5

(e) Discount factor

γ	Succ.
0.99	69.0±2.9
0.999	**85.3±1.7**
0.9999	83.7±1.3
1	75.7±1.3

Changing the number of bridge samples, the boundary stride, the refinement steps or the number of candidate plans ($N \geq 20$) shifts the success rate by at most 4%. $N = 10$ produces a larger drop, consistent with the coverage limitation. The discount factor γ shows a stronger effect as it trades the reachability for the plan length. A Δ-step-shorter plan is preferred when its reachability score is larger than γ^Δ times the original score. This means that when $\gamma = 0.999$, a 200-step shorter trajectory would be selected when its reachability score is at least 82% ($\approx 0.999^{200}$) of that of the longer plan. This threshold falls to $0.99^{200} \approx 13\%$ at $\gamma = 0.99$, which may disrupt the reachability assessment. At $\gamma = 1$, there is no length penalty to discourage unnecessarily long paths, which can hit the time limitation of each task.

D Qualitative Results

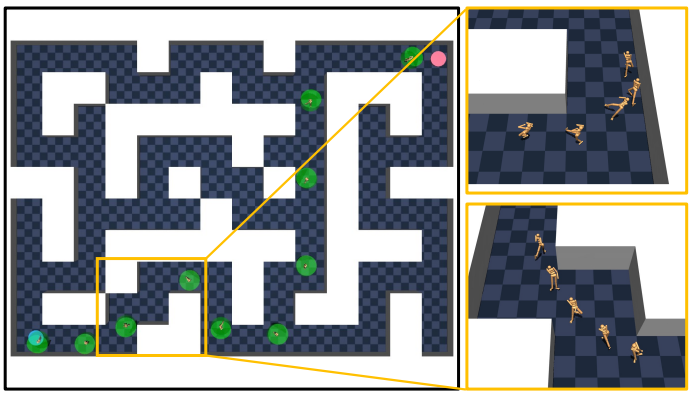


Figure 9: **Top-down plan overview.** SCD rollout in HumanoidMaze-Giant, where green circles mark the humanoid intervals along its path from the start state (cyan) to the goal state (pink). The insets magnify the boxed region, showing the humanoid following through the plan.

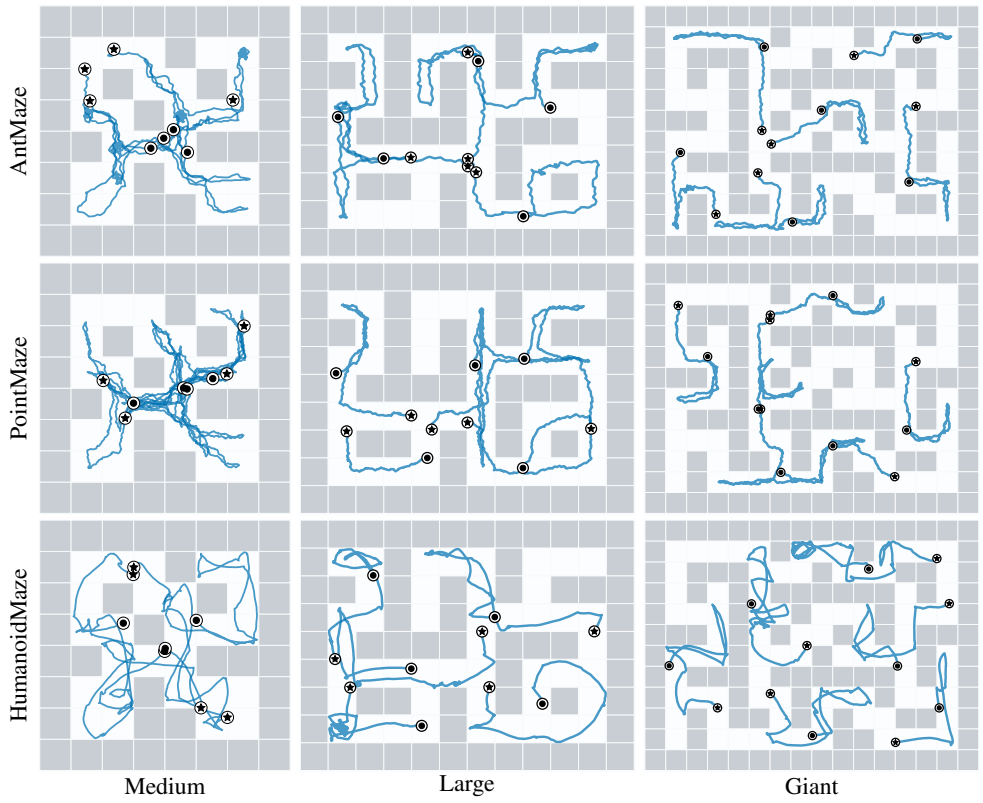


Figure 10: **Samples of offline trajectory data**, with each source (circle) connected to a target (star).

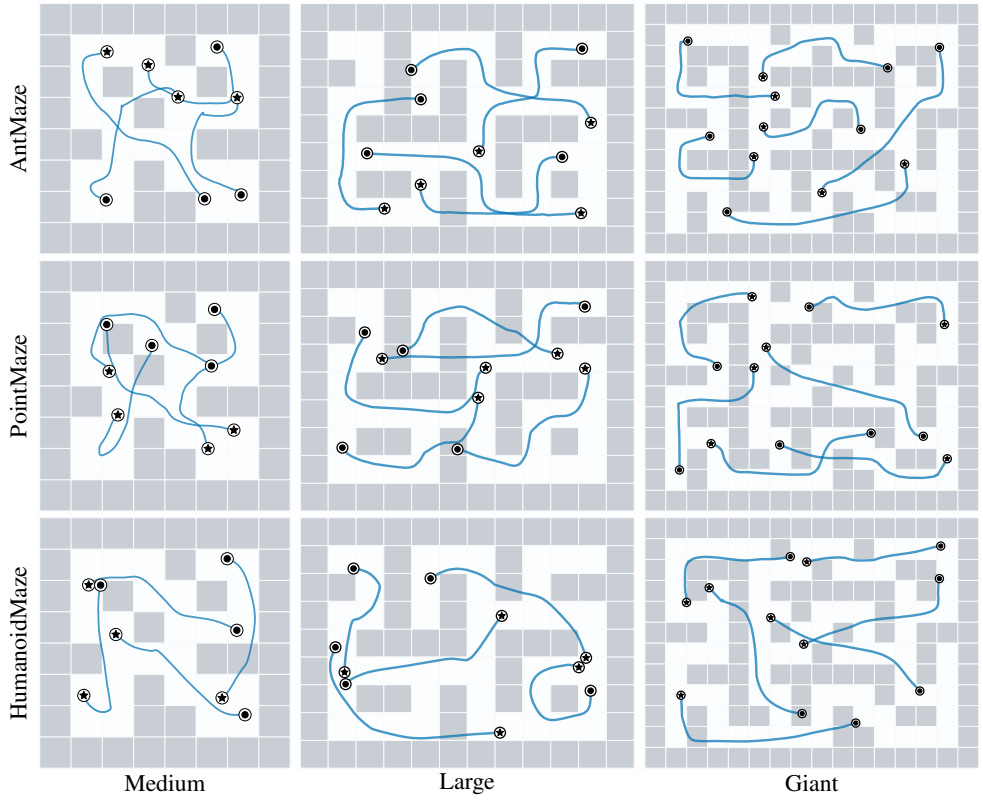


Figure 11: **Samples of generated trajectory data.**

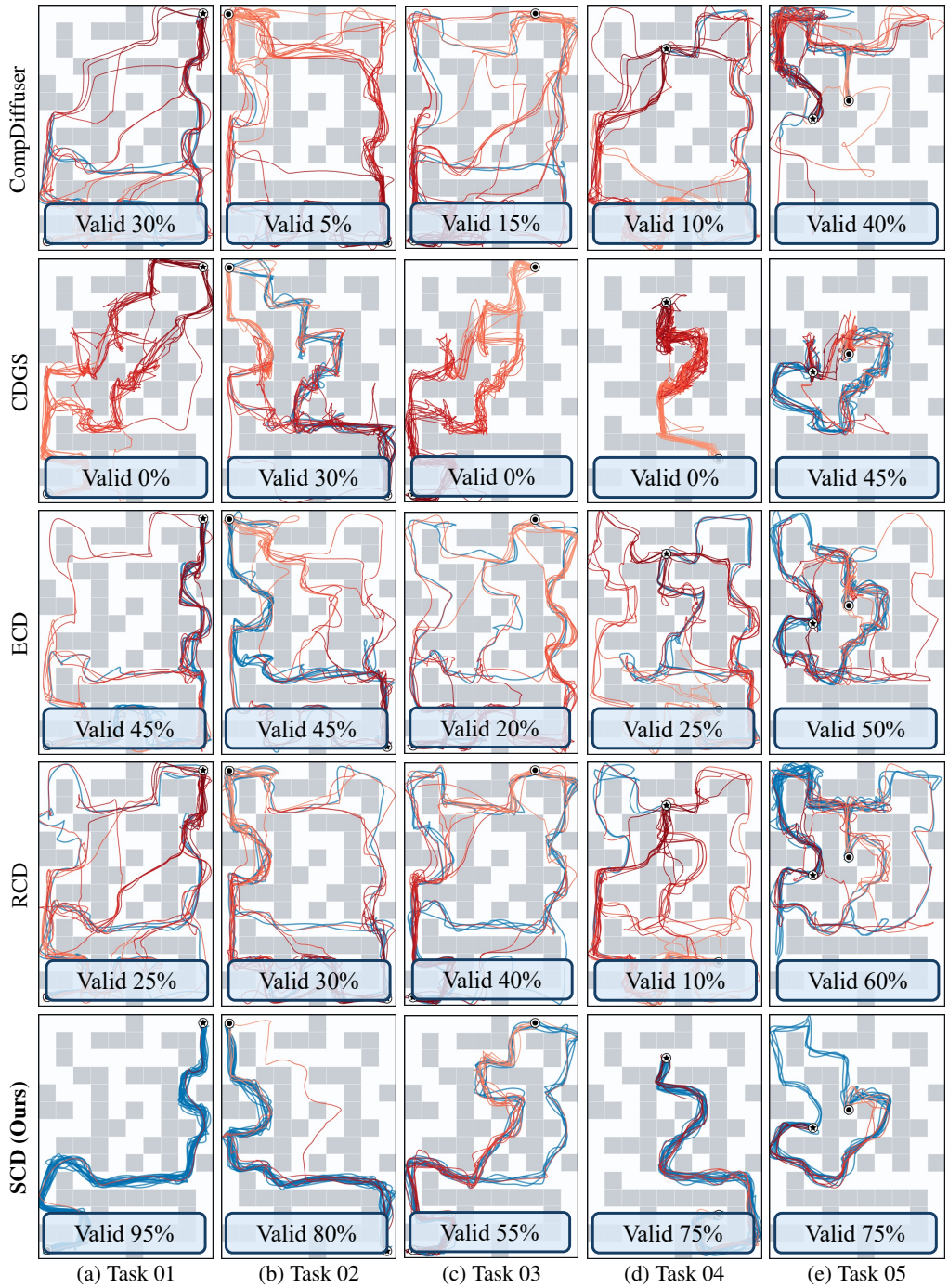


Figure 12: **Failure cases in HumanoidMaze-Giant.** The red color denotes infeasible trajectories and the blue color denotes feasible ones. *Valid* refers to a percentage of feasible trajectories.

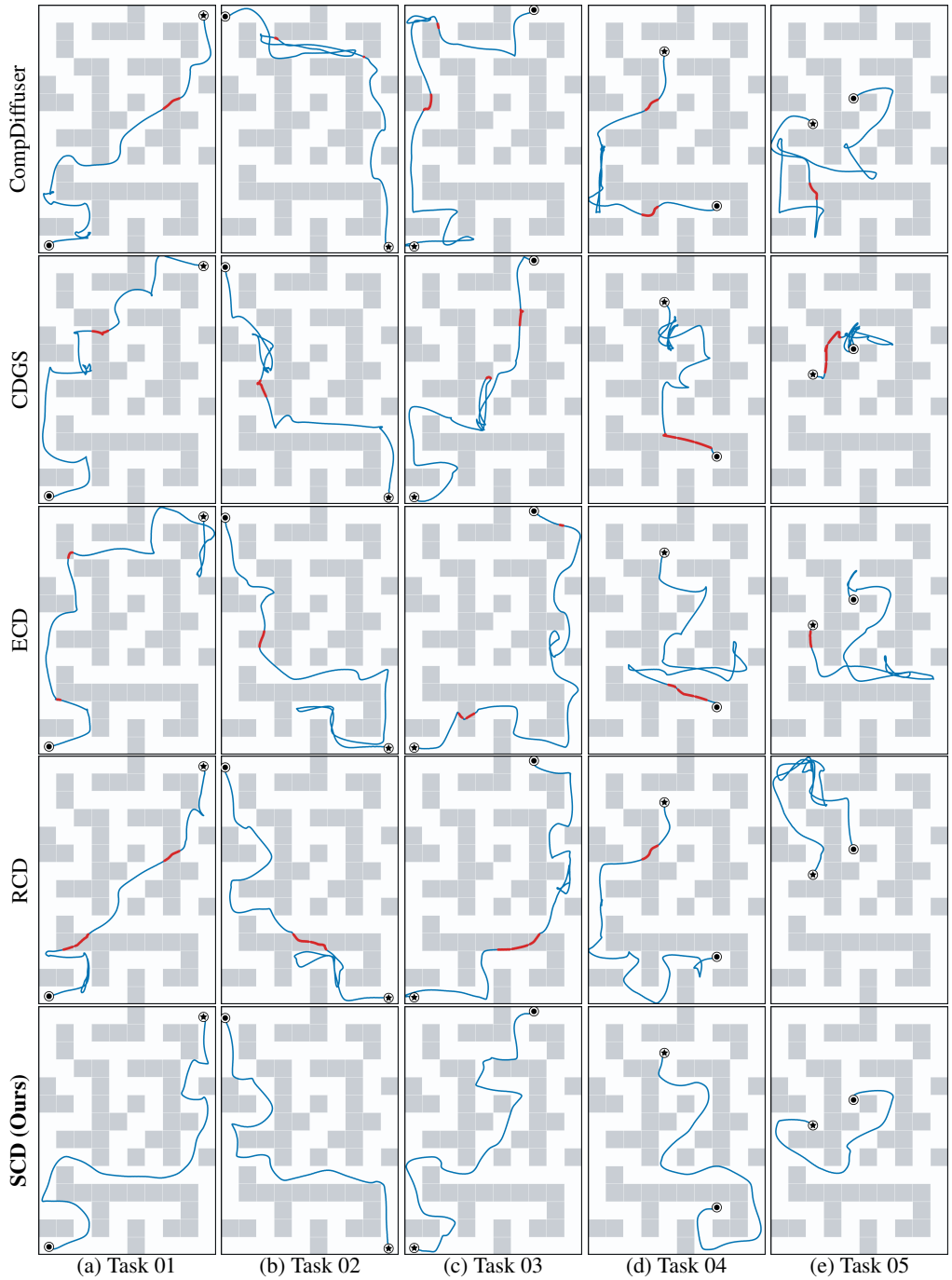


Figure 13: **Qualitative results of long-horizon planning in HumanoidMaze-Giant.** The red color denotes infeasible parts and the blue color denotes feasible parts.

E LICENSES

This work uses the following existing codebases, models, and datasets:

- OGBench (Park et al., 2025, `https://github.com/seohongpark/ogbench`): Used for the datasets and evaluation environments. Licensed under the MIT License.
- MuJoCo (`https://github.com/google-deepmind/mujoco`): Used as the physics simulator of OGBench. Licensed under the Apache-2.0 License.
- CompDiffuser (Luo et al., 2025, `https://github.com/devinluo27/comp_diffuser_release`): Used for the pretrained diffusion models and controllers, and as the base implementation of the planner. Licensed under the MIT License.
- CDGS (Mishra et al., 2026, `https://github.com/UtkarshMishra04/CDGS_ogbench`): Used as a reference implementation for the CDGS baseline. Licensed under the MIT License.
- ECD (Sun et al., 2026, `https://github.com/GradientSpaces/ECD`): Used as a reference implementation for the ECD baseline. Licensed under the Apache-2.0 License.
- RCD (Lee et al., 2026, `https://github.com/leekwoon/rcd`): Used as a reference implementation for the RCD baseline. Licensed under the MIT License.
- FAISS (`https://github.com/facebookresearch/faiss`): Used for the k-NN search in density reweighting. Licensed under the MIT License.