

Hyperspherical Contrastive Representations Unlock Planning and Exploration

Jonghae Park*, Sungil Son*, Jonghun Shin, Junseo Lee

H. Jin Kim, Benjamin Eysenbach

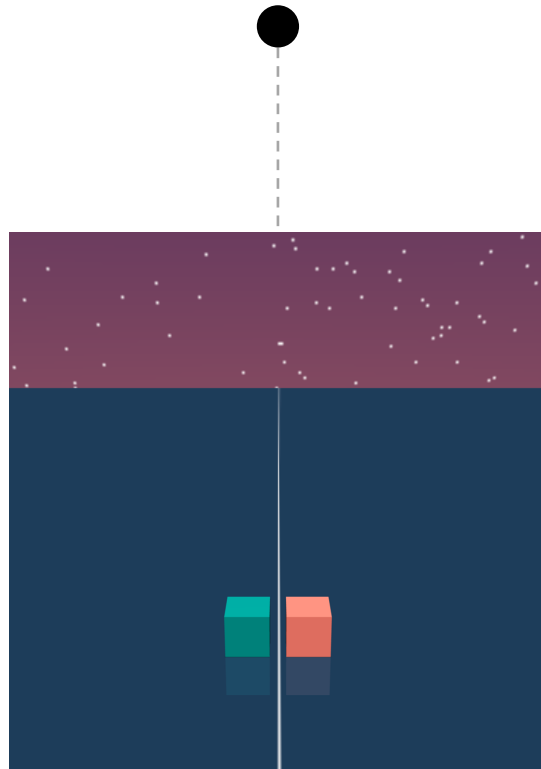


Consider a robot  asked to build a tower of blocks.

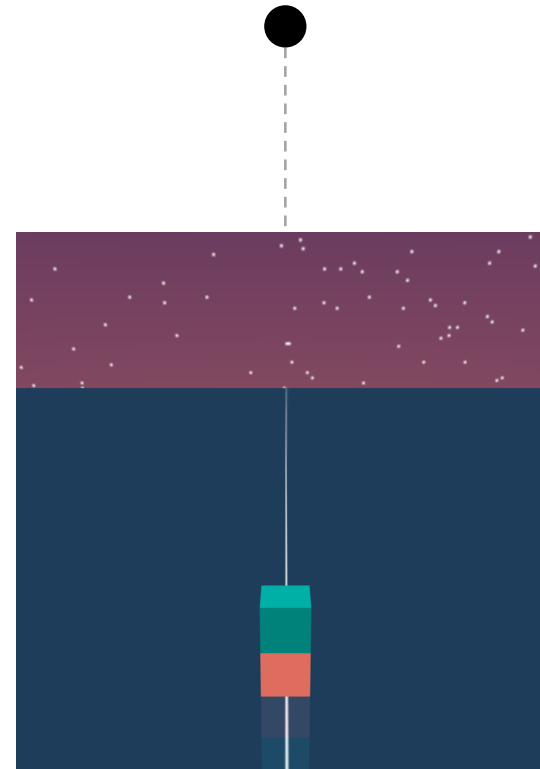
Consider a robot  asked to build a tower of blocks.

How many subgoals, *or waypoints*, does the robot need?

Stacking Two Blocks

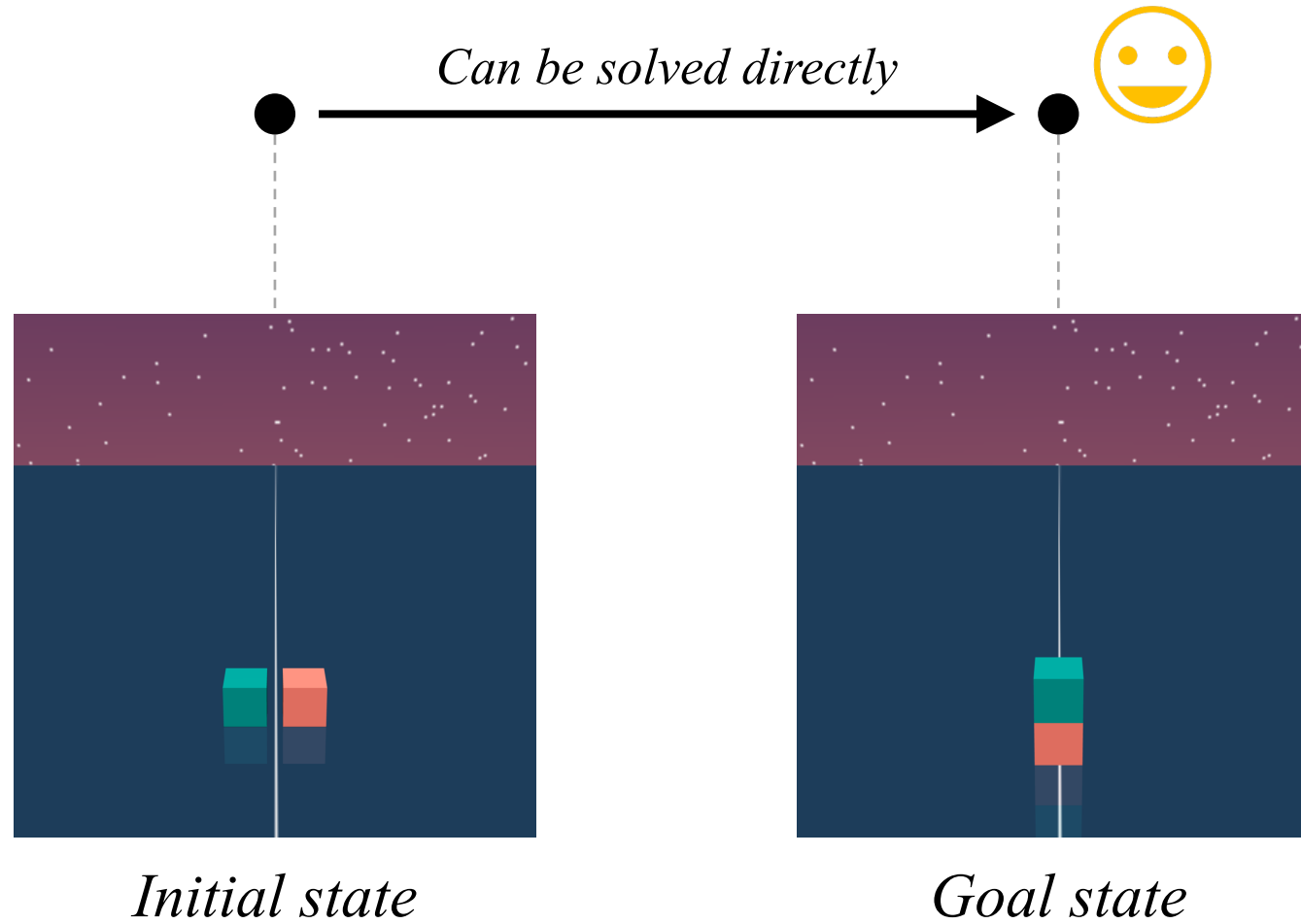


Initial state

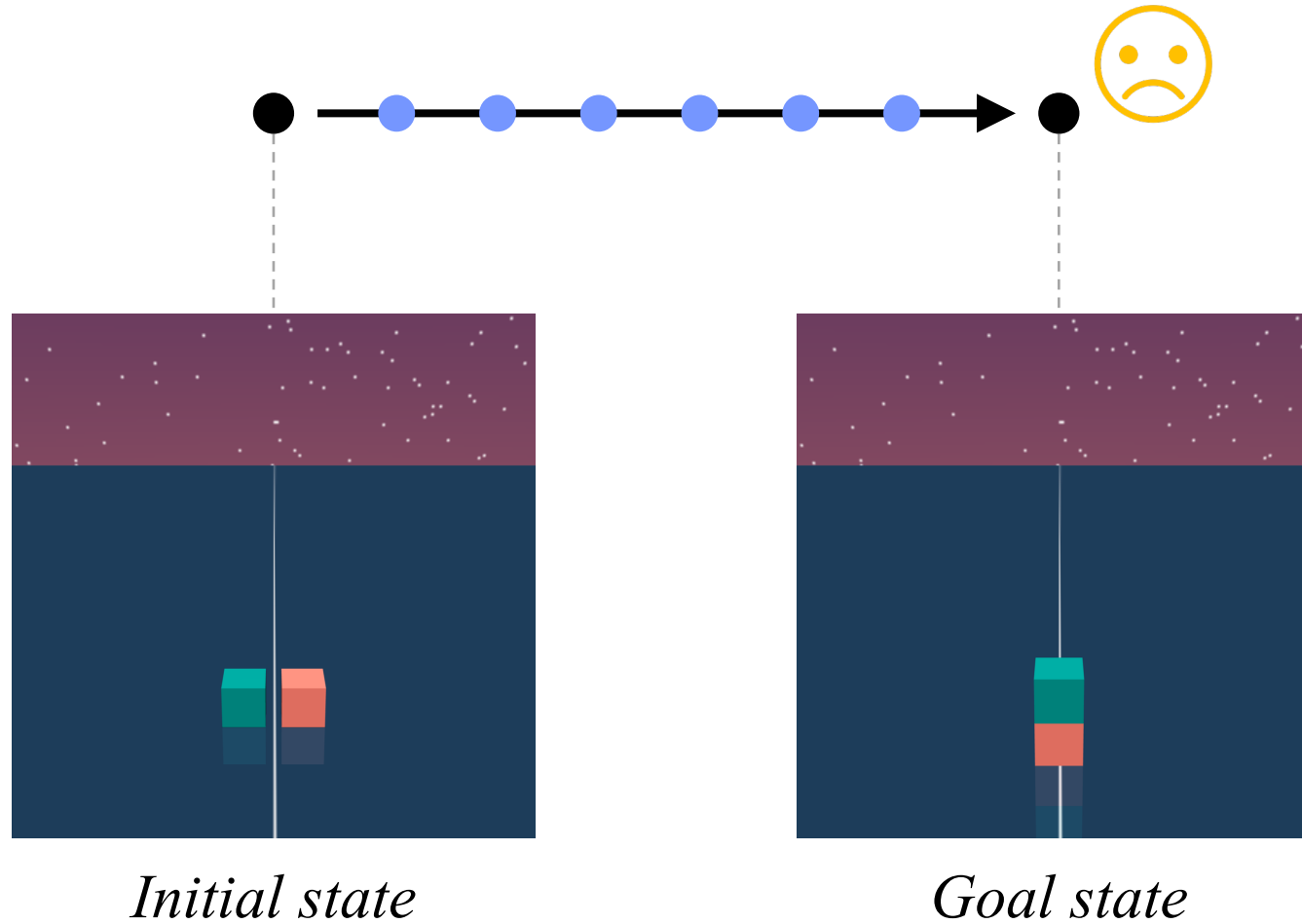


Goal state

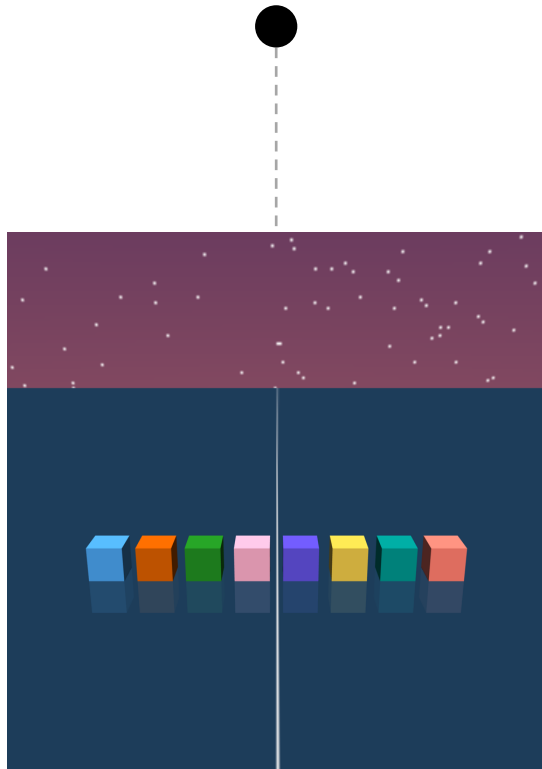
Stacking Two Blocks



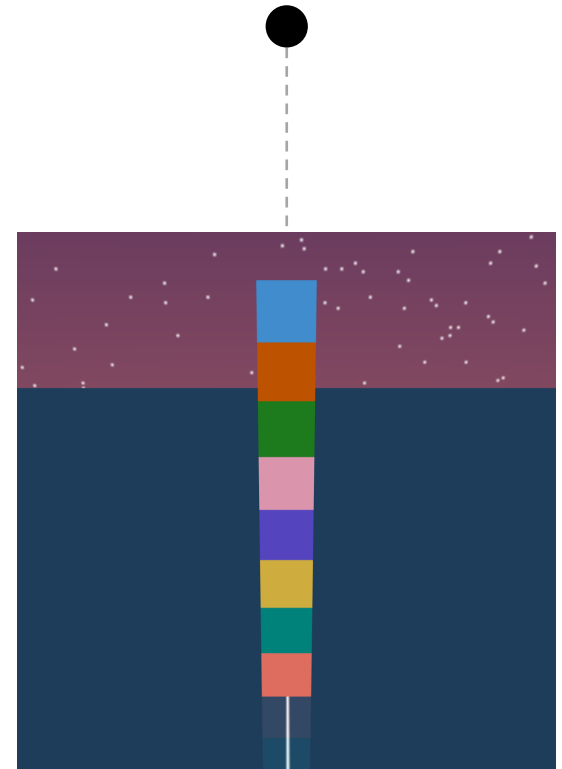
Stacking Two Blocks with too many planning stages



What about Eight Blocks?

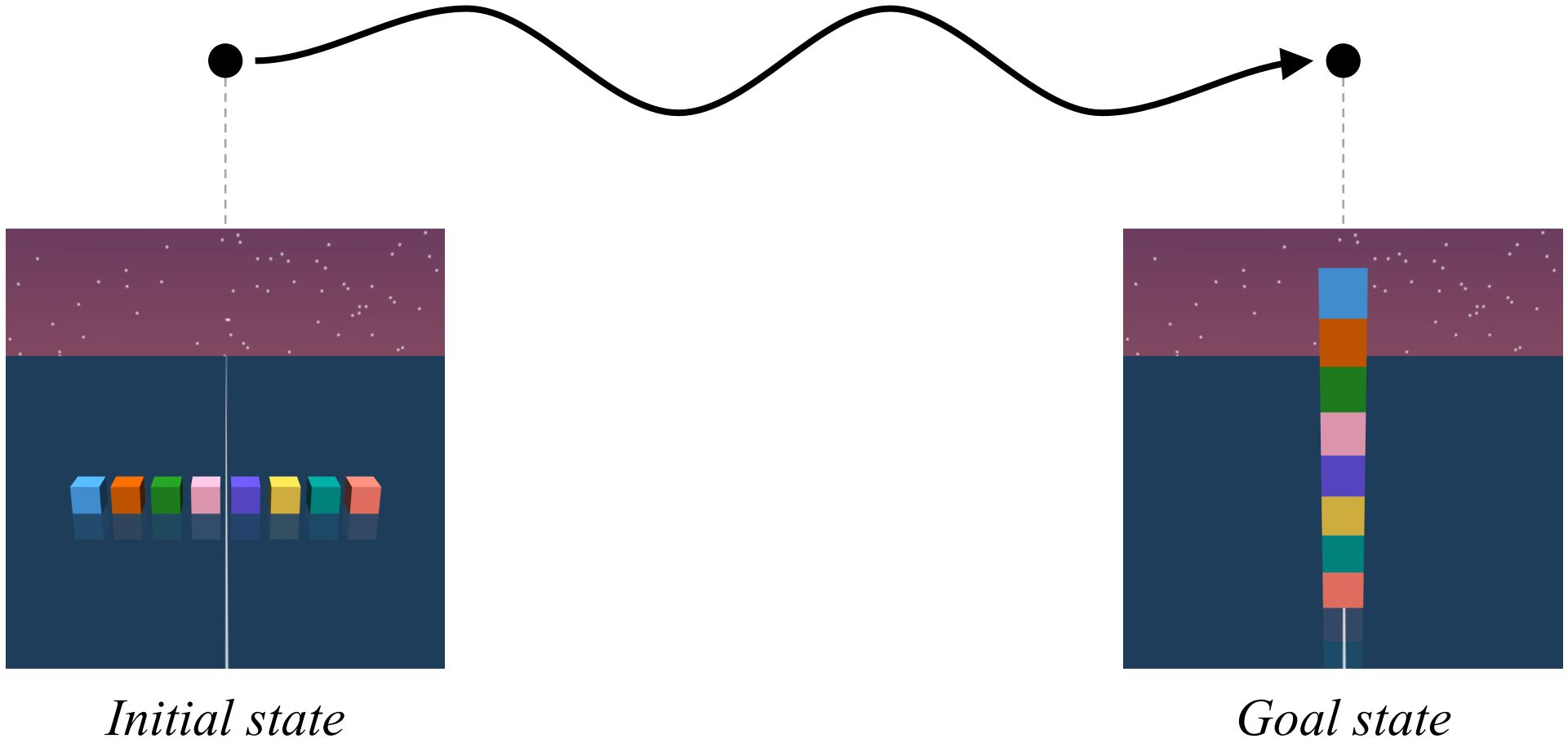


Initial state

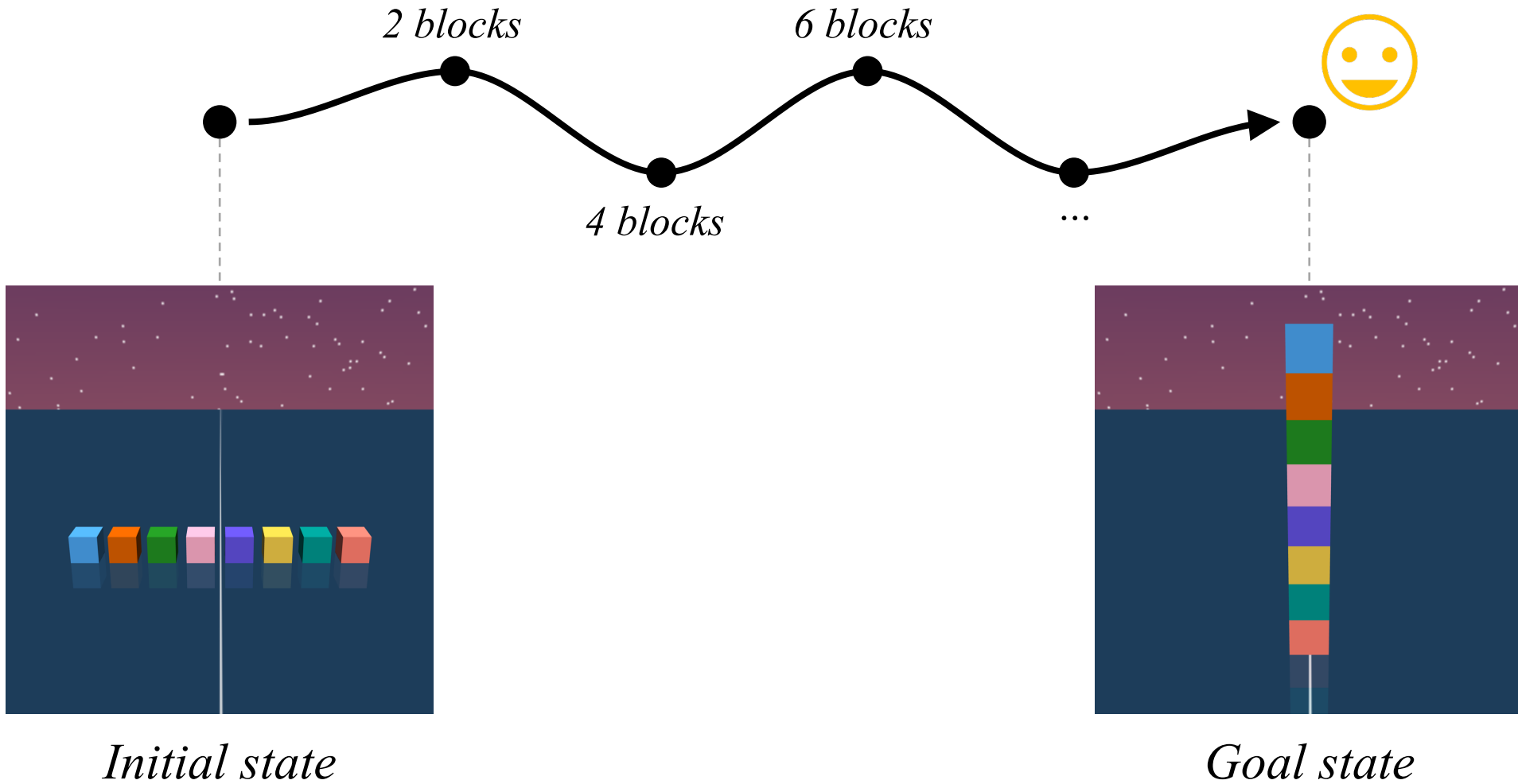


Goal state

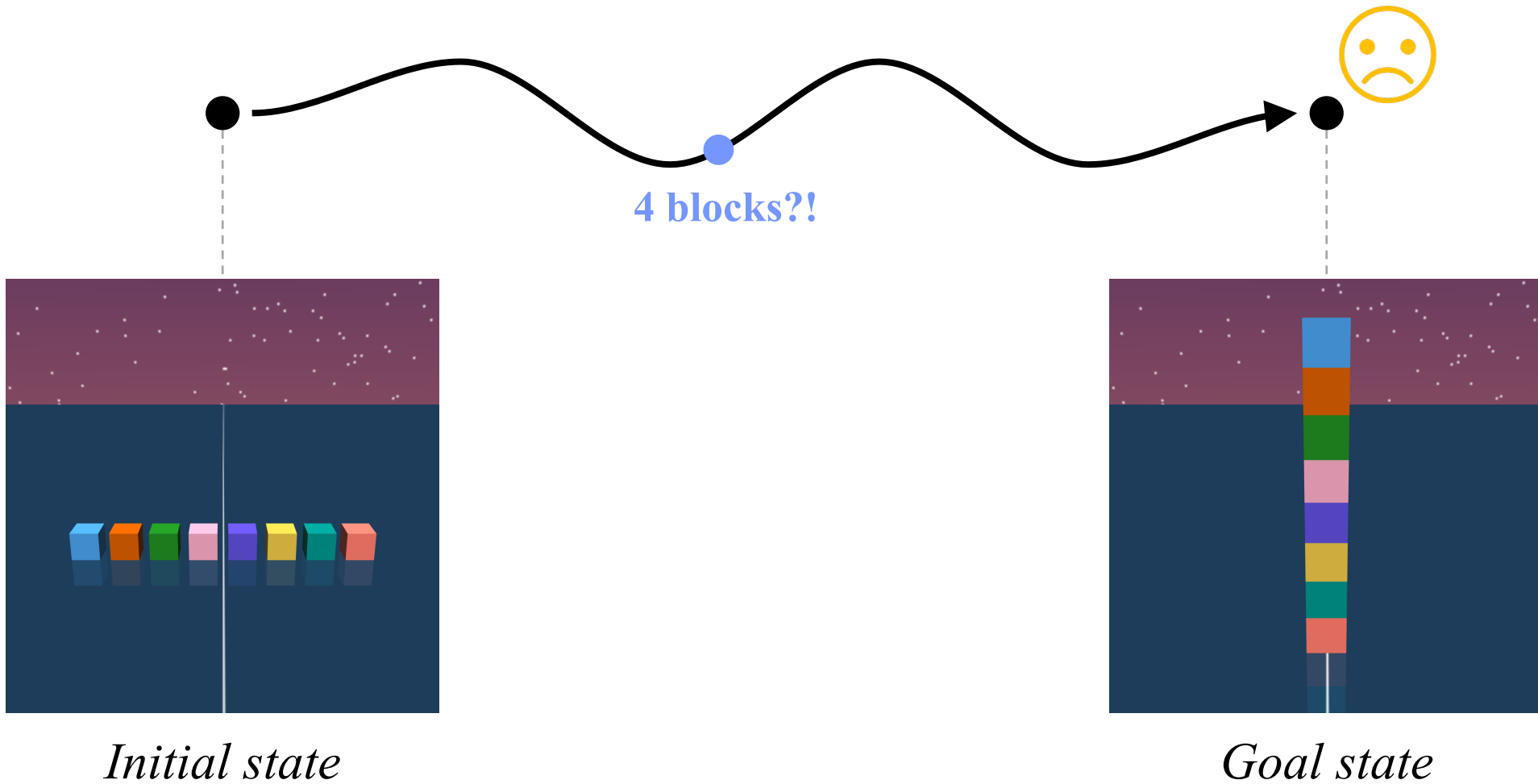
What about Eight Blocks?



What about Eight Blocks?



Stacking Eight Blocks with too few planning stages



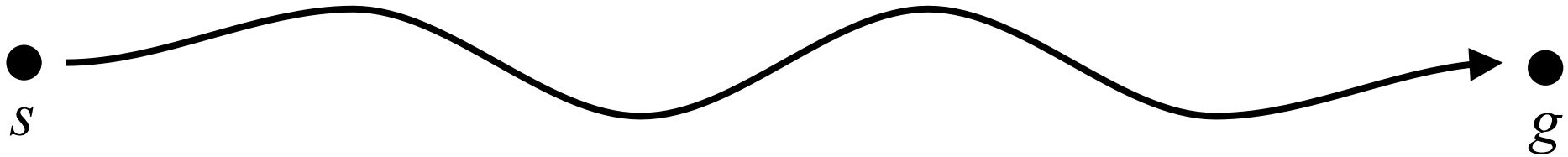
Desiderata for **Long-horizon** Planning and Exploration

Desiderata for **Long-horizon** Planning and Exploration

- (1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ *waypoints*)

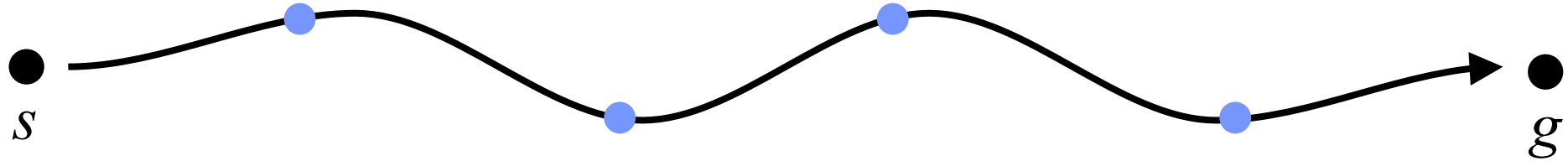
Desiderata for **Long-horizon** Planning and Exploration

- (1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ *waypoints*)



Desiderata for **Long-horizon** Planning and Exploration

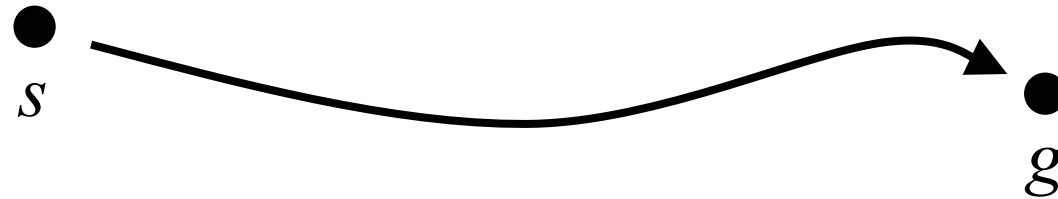
(1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ *waypoints*)



: **Many waypoints**

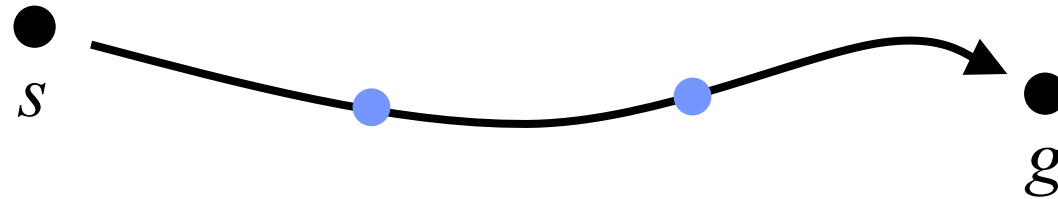
Desiderata for **Long-horizon** Planning and Exploration

- (1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ *waypoints*)



Desiderata for **Long-horizon** Planning and Exploration

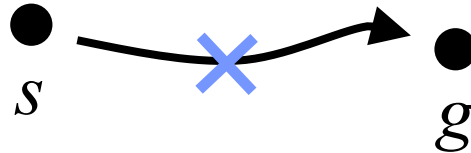
(1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ waypoints)



: Fewer waypoints

Desiderata for **Long-horizon** Planning and Exploration

(1) **The number of planning stages** should match the **difficulty** of the problem
($\approx$ waypoints)



: Or even **NONE**

Desiderata for **Long-horizon** Planning and Exploration

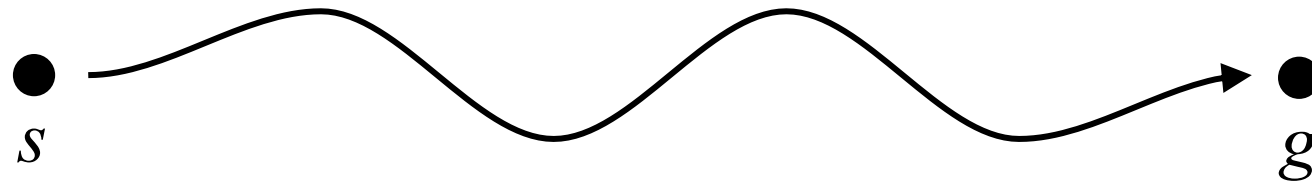
(1) The number of planning stages should match the difficulty of the problem

(2) Waypoints should be **directed toward the goal**.

Desiderata for **Long-horizon** Planning and Exploration

- (1) The number of **planning stages** should match the **difficulty** of the problem
- (2) Waypoints should be **directed toward the goal**.

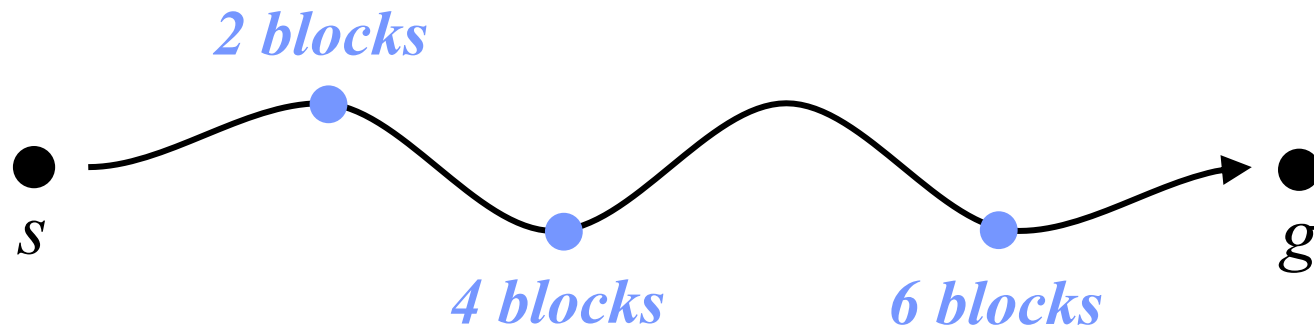
Consider the eight-block stacking task,



Desiderata for **Long-horizon** Planning and Exploration

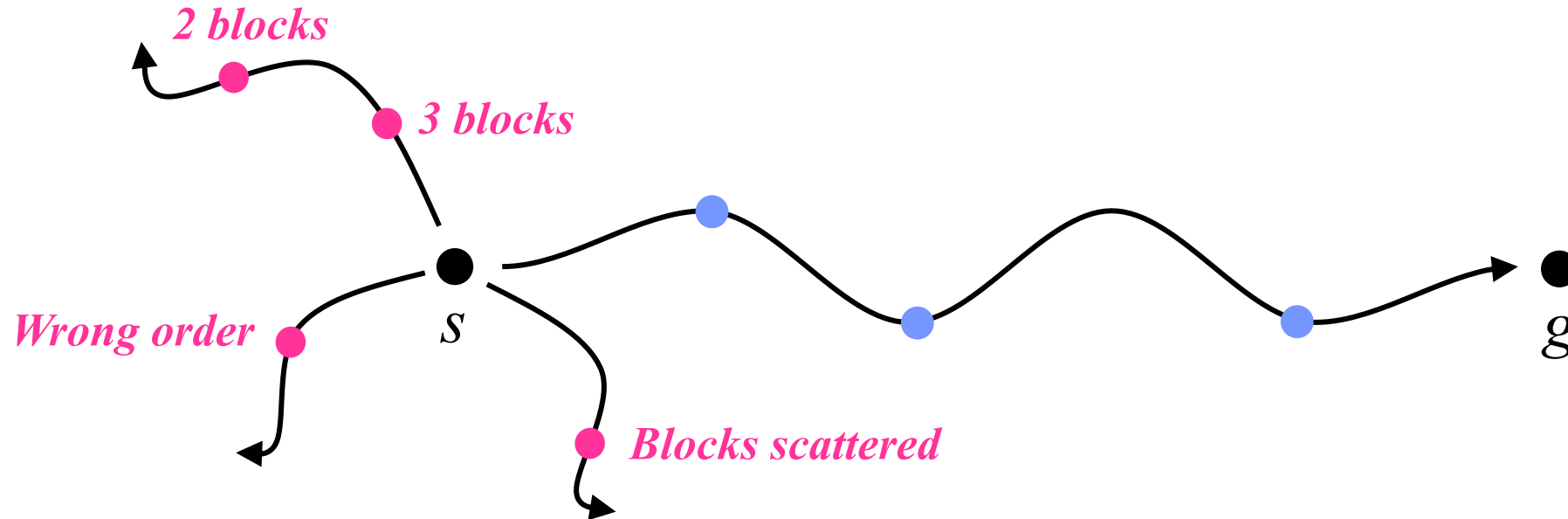
- (1) The number of planning stages should match the **difficulty** of the problem
- (2) Waypoints should be **directed toward the goal**.

Consider the eight-block stacking task,



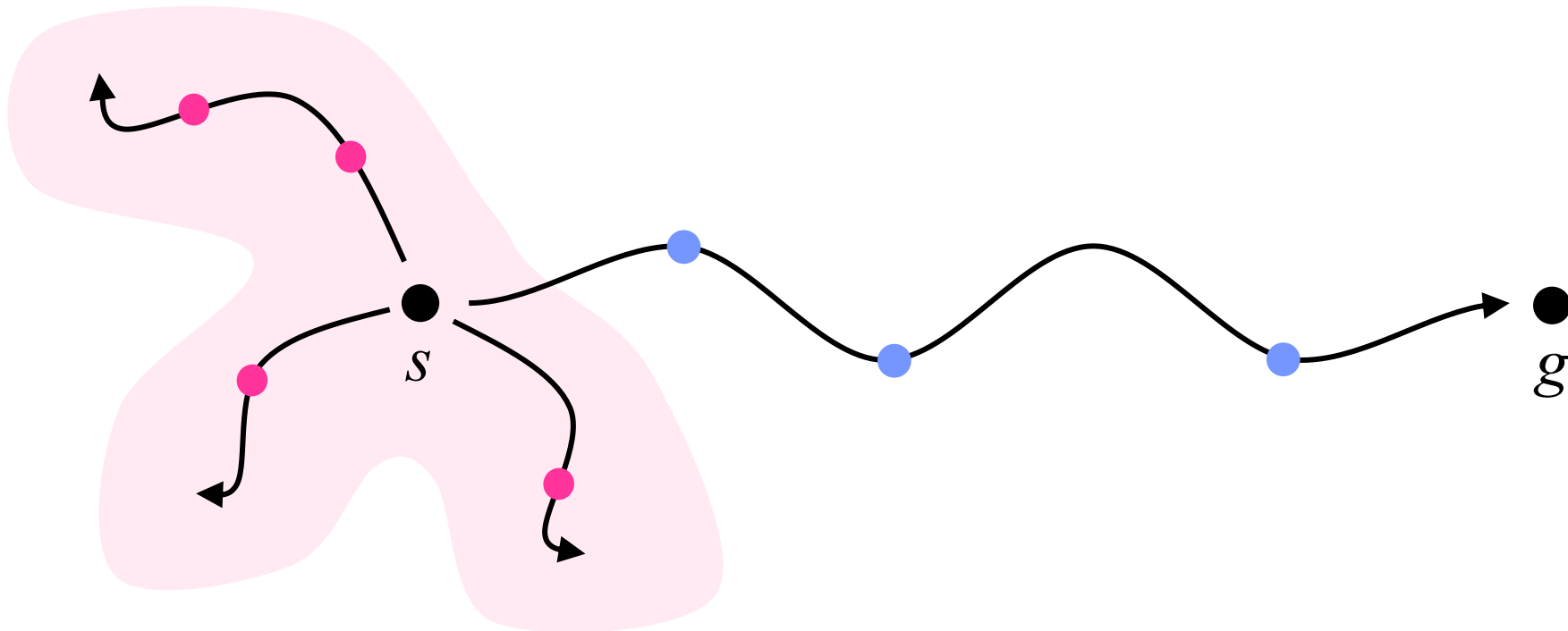
Desiderata for **Long-horizon** Planning and Exploration

- (1) The number of planning stages should match the **difficulty** of the problem
- (2) Waypoints should be **directed toward the goal**.



Desiderata for **Long-horizon** Planning and Exploration

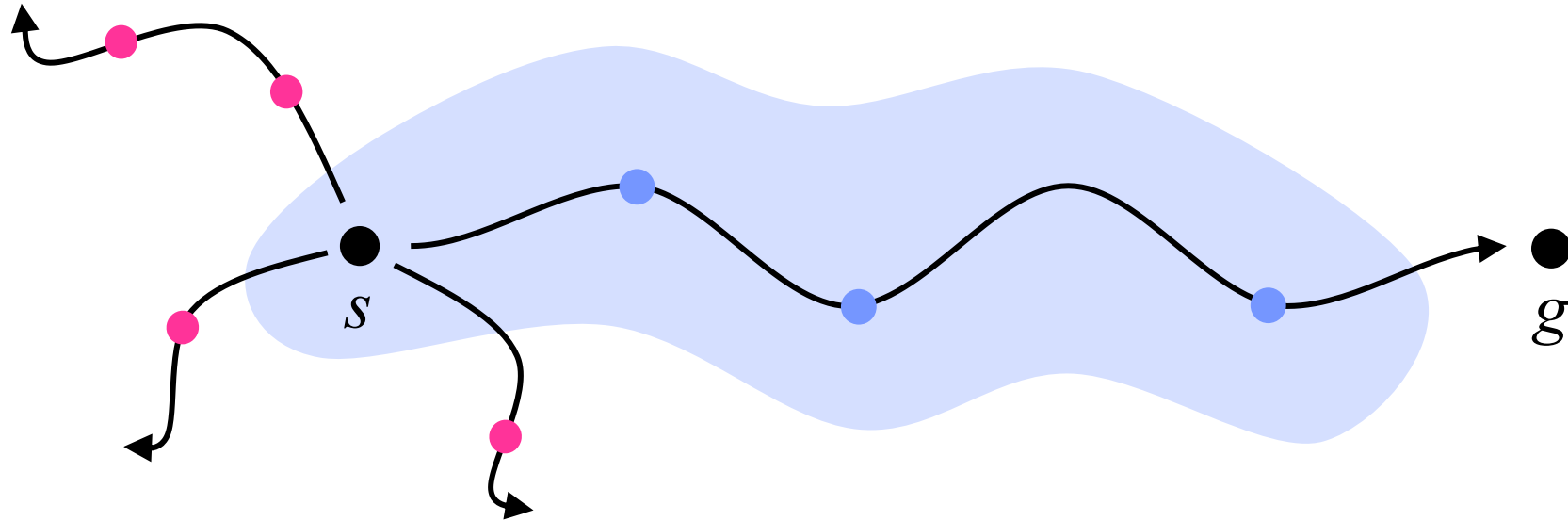
- (1) The number of planning stages should match the **difficulty** of the problem
- (2) Waypoints should be **directed toward the goal**.



Desiderata for **Long-horizon** Planning and Exploration

(1) The number of planning stages should match the **difficulty** of the problem

(2) Waypoints should be **directed toward the goal**



Problem Setting



Basic Setup : Data collection process

Basic Setup : Data collection process

High level planner

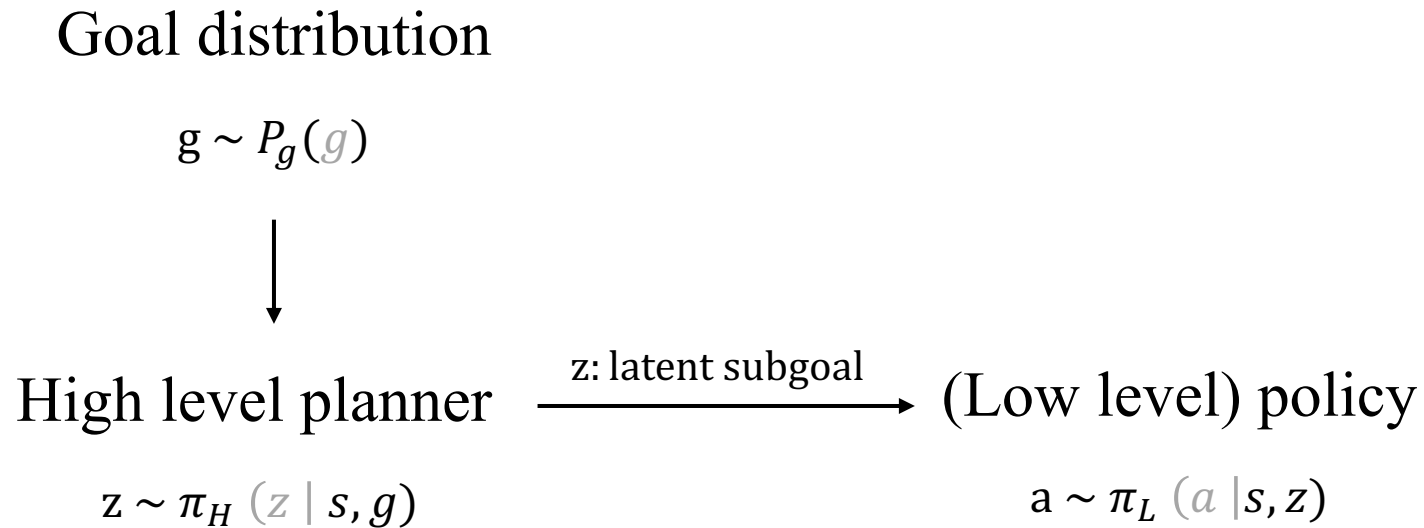
$$z \sim \pi_H (z \mid s, g)$$

(Low level) policy

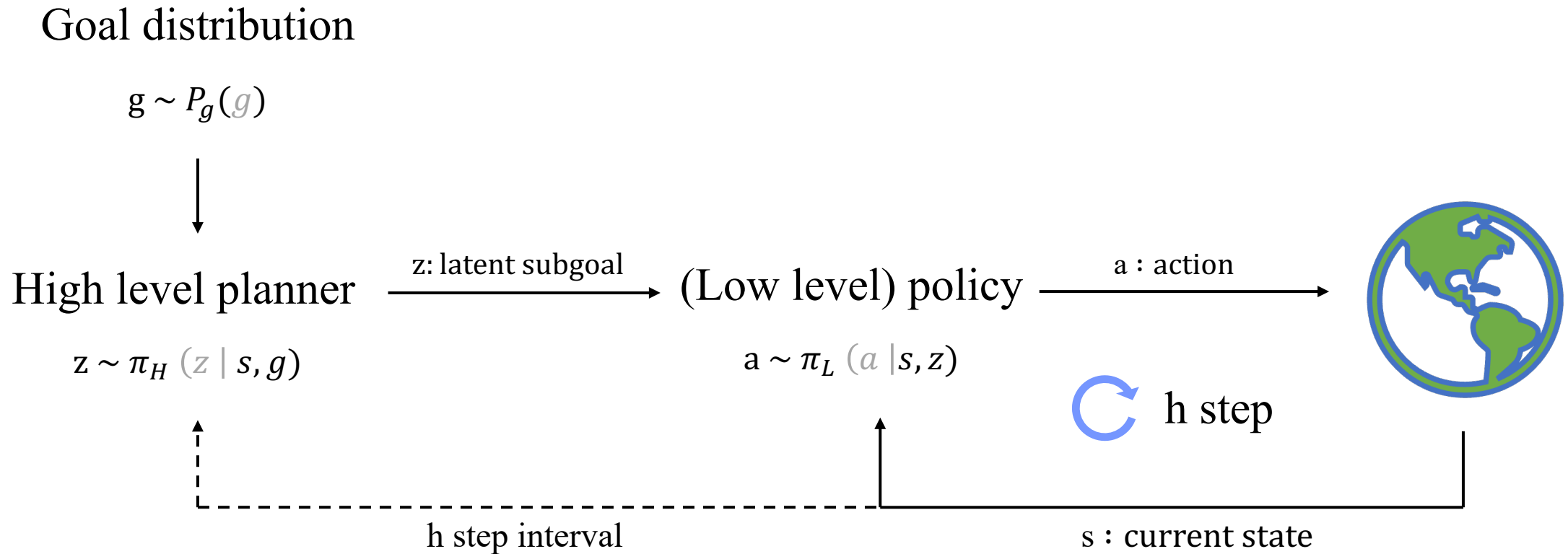
$$a \sim \pi_L (a \mid s, z)$$



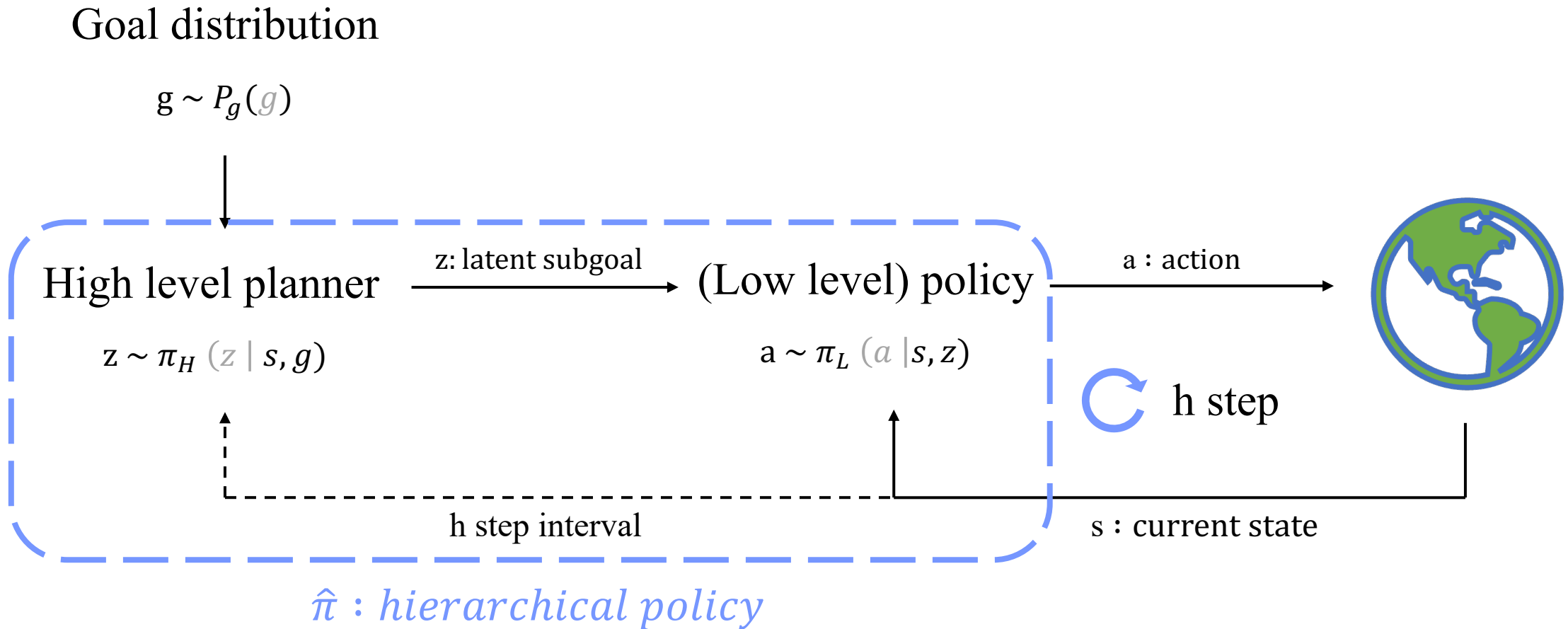
Basic Setup : Data collection process



Basic Setup : Data collection process



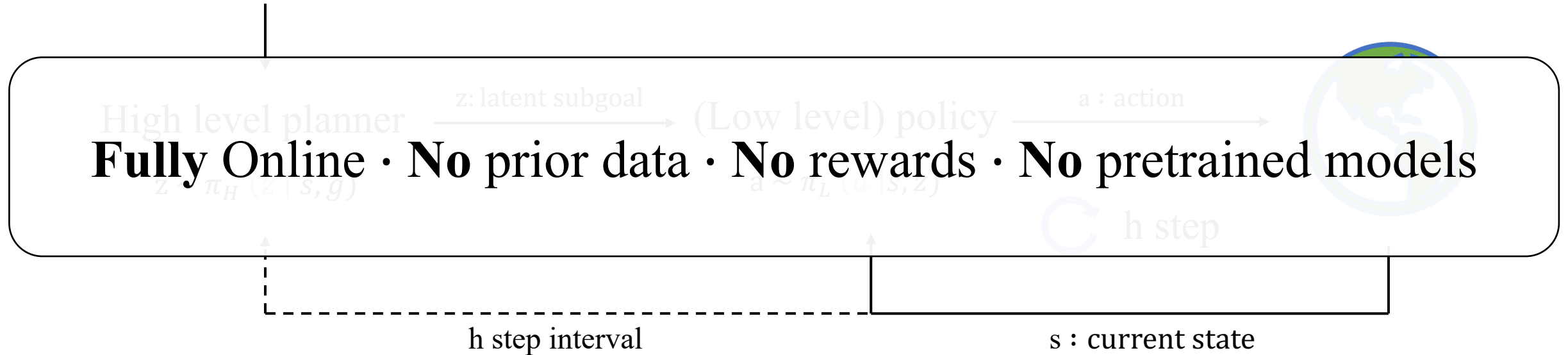
Basic Setup : Data collection process



Basic Setup : Data collection process

Goal distribution

$$g \sim P_g(g)$$



Basic Setup : A probabilistic approach to planning

Basic Setup : A probabilistic approach to planning

Given a state s and a goal g ,

Basic Setup : A probabilistic approach to planning

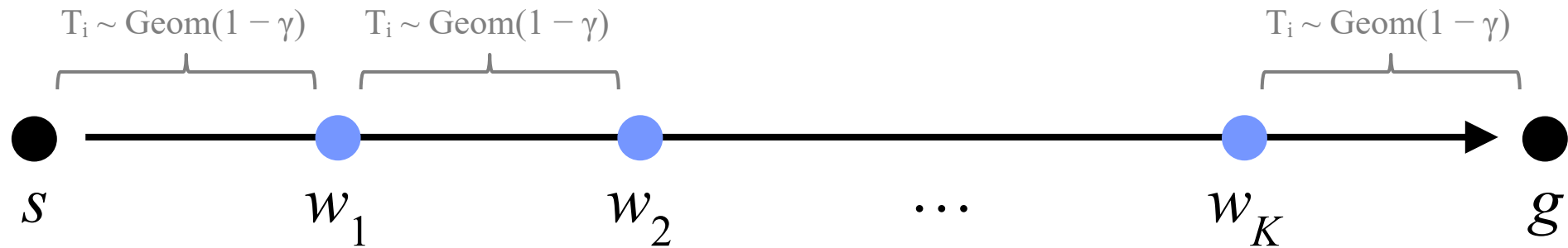
Given a state s and a goal g ,

assume **K waypoints** in between, with intervals $\mathbf{T_i} \sim \mathbf{Geom}(1 - \gamma)$.

Basic Setup : A probabilistic approach to planning

Given a state s and a goal g ,

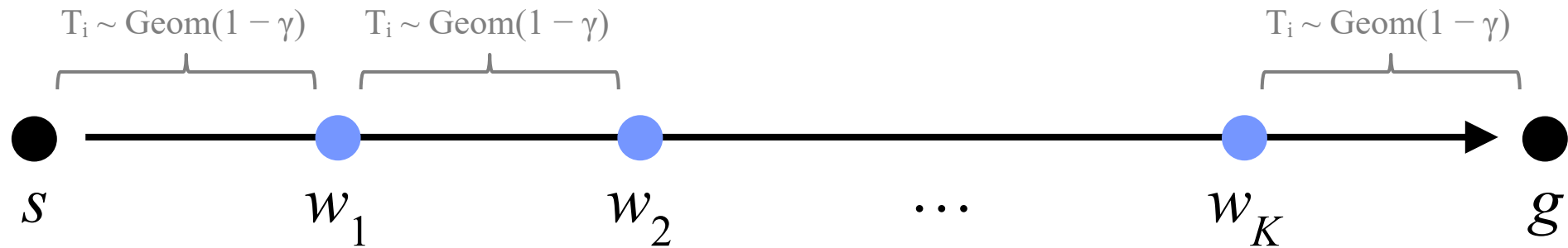
assume **K waypoints** in between, with intervals $T_i \sim \mathbf{Geom}(1 - \gamma)$.



Basic Setup : A probabilistic approach to planning

Given a state s and a goal g ,

assume **K waypoints** in between, with intervals $T_i \sim \mathbf{Geom}(1 - \gamma)$.

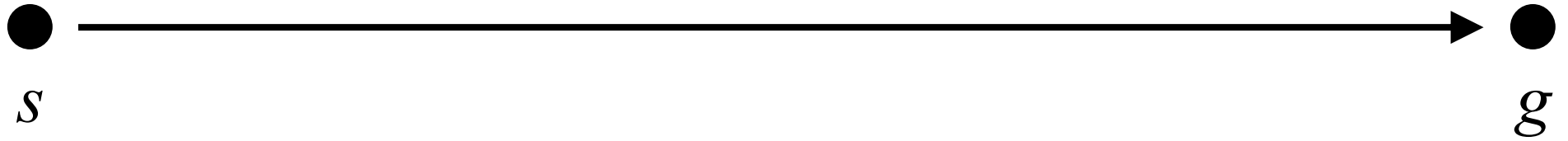


For notational consistency, we set $w_0 = s$ and $w_{K+1} = g$

The Objective

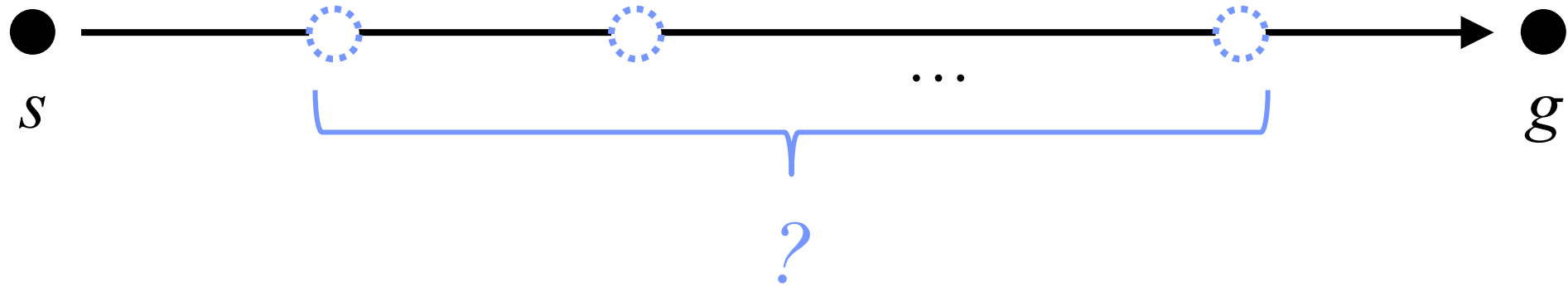
The Objective

Given (s, g) and number of waypoints K ,



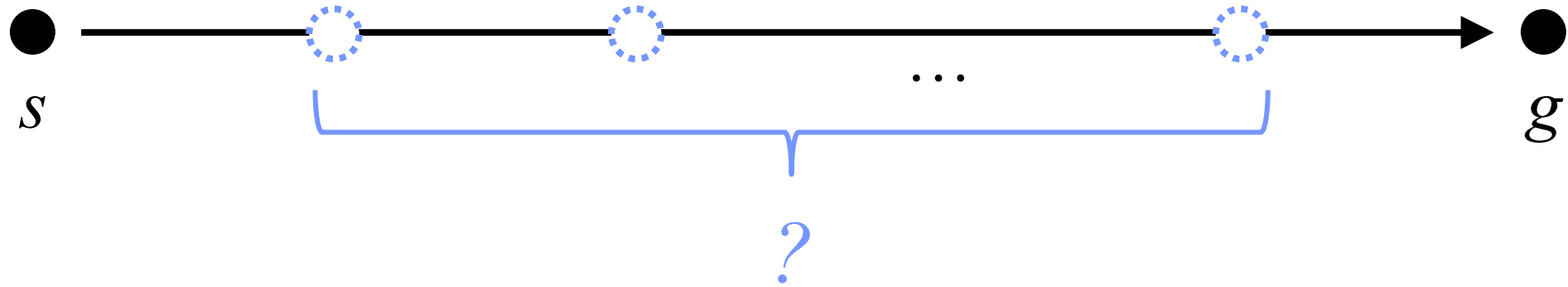
The Objective

Given (s, g) and number of waypoints K ,



The Objective

Given (s, g) and number of waypoints K ,



For now, K is given. We then derive a **criterion for selecting K adaptively**.

The Objective

Given (s, g) and number of waypoints K ,

$$q^{\star} = \arg \min_q D_{\text{KL}} \left(q(w_{1:K} \mid s, g) \parallel p_{\hat{\pi}}^{\gamma, K}(w_{1:K} \mid s, g) \right)$$

The Objective

Given (s, g) and number of waypoints K ,

$$q^{\star} = \arg \min_q D_{\text{KL}} \left(q(w_{1:K} \mid s, g) \parallel p_{\hat{\pi}}^{\gamma, K}(w_{1:K} \mid s, g) \right)$$

The distribution over the K intermediate waypoints
given s and g , under data collected by $\hat{\pi}$

The Objective

Given (s, g) and number of waypoints K ,

$$q^{\star} = \arg \min_q D_{\text{KL}} \left(\boxed{q(w_{1:K} \mid s, g)} \parallel p_{\hat{\pi}}^{\gamma, K}(w_{1:K} \mid s, g) \right)$$

A tractable **approximation** to the posterior $p_{\hat{\pi}}^{\gamma, K}(\cdot \mid s, g)$

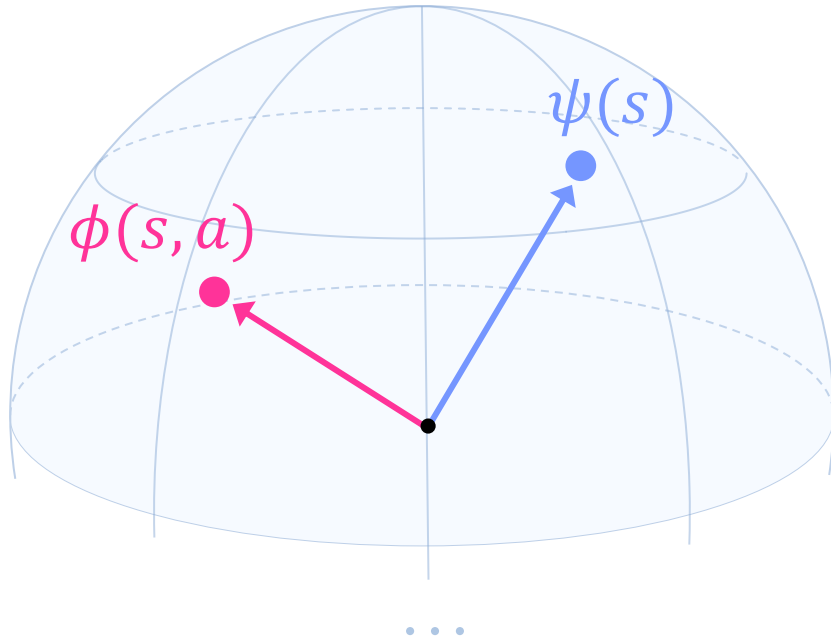
Hyperspherical Contrastive Representations

Hyperspherical Contrastive Representations

Two representations on the unit hypersphere S^{d-1} + One learnable **rotation matrix** R .

Hyperspherical Contrastive Representations

Two representations on the unit hypersphere S^{d-1} + One learnable rotation matrix R .



$\phi(s, a)$: state–action representation

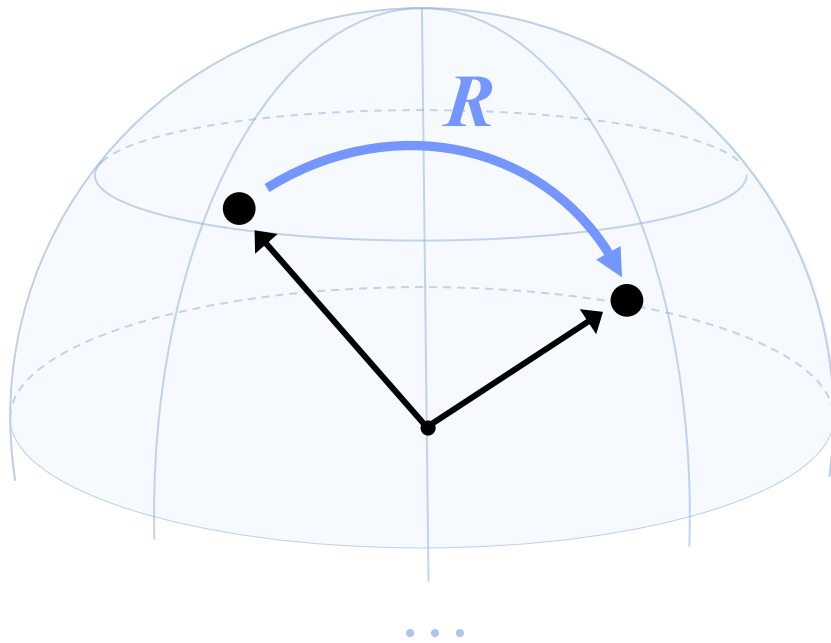
$\psi(s)$: state representation

$$\|\phi(s, a)\|_2 = \|\psi(s)\|_2 = 1$$

: lie on the unit hypersphere S^{d-1}

Hyperspherical Contrastive Representations

Two representations on the unit hypersphere S^{d-1} + One learnable **rotation matrix** R .



Cayley transform of a *learnable* matrix A :

$$\left. \begin{aligned} \mathbf{\Omega} &= A - A^T \\ \mathbf{R} &= (I + \mathbf{\Omega})^{-1}(I - \mathbf{\Omega}) \end{aligned} \right\} \Rightarrow R^T R = I, \|Rx\| = \|x\|$$

: preserves hyperspherical geometry of representations

Parameterization

Parameterization

1. Score for *planning*

$$f_H(s, s^+) = \psi(s)^T R_{learnable} \psi(s^+)$$

Parameterization

1. Score for *planning*

$$f_H(s, s^+) = \psi(s)^T R_{learnable} \psi(s^+)$$

1. *Why use the same encoder for s and s^+ ?*

To infer consecutive waypoints in a common latent space.

Parameterization

1. Score for *planning*

$$f_H(s, s^+) = \psi(s)^T R_{\text{learnable}} \psi(s^+)$$

1. *Why use the same encoder for s and s^+ ?*

To infer consecutive waypoints in a common latent space.

2. *Why introduce the **rotation matrix R** ?*

In general, dynamics are *asymmetric*,

but the inner product is *symmetric* : $\psi(s)^T \psi(s^+) = \psi(s^+)^T \psi(s)$

Parameterization

1. Score for *planning*

$$f_H(s, s^+) = \psi(s)^T R_{learnable} \psi(s^+)$$

2. Score for *training policy*

$$f_L(s, a, s^+) = \phi(s, a)^T \psi(s^+)$$

Representation Learning

Sample a batch $\mathcal{B} = \{(s_i, a_i, s_i^+)\}_{i=1}^B$ from the replay buffer, where s_i^+ is $\Delta \sim \text{Geom}(1 - \gamma)$ steps after s_i :

$$\mathcal{L}^{\text{symmetric}}(\mathbf{R}, \psi; \mathcal{B}) = -\frac{1}{B} \sum_{i=1}^B \left[\log \frac{\exp(f_H(s_i, s_i^+))}{\sum_{j=1}^B \exp(f_H(s_i, s_j^+))} + \log \frac{\exp(f_H(s_i, s_i^+))}{\sum_{j=1}^B \exp(f_H(s_j, s_i^+))} \right]$$

$$\mathcal{L}^{\text{backward}}(\phi, \psi; \mathcal{B}) = -\frac{1}{B} \sum_{i=1}^B \log \frac{\exp(f_L(s_i, a_i, s_i^+))}{\sum_{j=1}^B \exp(f_L(s_j, a_j, s_i^+))}$$

$$\mathcal{L}_{\text{repr}} = E_{\mathcal{B}} \left[\mathcal{L}^{\text{symmetric}}(\mathbf{R}, \psi; \mathcal{B}) + \lambda_{\text{repr}} \cdot \mathcal{L}^{\text{backward}}(\phi, \psi; \mathcal{B}) \right]$$

Policy Learning

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, g \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(g))} [\phi(s, a)^T \psi(g)]$$

Policy Learning

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, g \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(g))} \underbrace{[\phi(s, a)^T \psi(g)]}_{f_L(s, a, g)}$$

Policy Learning

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, g \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(g))} \underbrace{[\phi(s, a)^T \psi(g)]}_{f_L(s, a, g)}$$

Same as CRL [1], except for **this**

Key Assumptions on the Representations

[Assumption 1] The learned f_H under the symmetric InfoNCE objective has the following *Bayes-optimal* form:

$$f_H(u, v) = \log \frac{p_{\hat{\pi}}^{\gamma}(u^+ = v \mid u)}{p_{\hat{\pi}}^{\gamma}(v) \mathcal{C}} \quad [1]$$

[Assumption 2] The marginal distribution over the representations ψ is *uniform* on the unit hypersphere $\mathbb{S}^{d-1}$:

$$p_{\hat{\pi}}^{\gamma}(\psi) = \text{Unif}(\mathbb{S}^{d-1}) \quad [2]$$

[1] Eysenbach, Benjamin, et al. "Inference via interpolation: Contrastive representations provably enable planning and inference." 2024.

[2] Wang, Tongzhou, et al. "Understanding contrastive representation learning through alignment and uniformity on the hypersphere." 2020.

Tractable Reformulation

#1

$$\arg \min_q D_{\text{KL}} \left(q(\boldsymbol{w}_{1:K} \mid s, g) \parallel p_{\hat{\pi}}^{\gamma, K}(\boldsymbol{w}_{1:K} \mid s, g) \right)$$

Assumption 1 & Markov assumption

$$q(\boldsymbol{w}_{1:K} \mid s, g) = \prod_{i=1}^K q_i(\boldsymbol{w}_i \mid s, g)$$

#2

$$\arg \max_q \left\{ \mathbb{E}_{\boldsymbol{w}_{1:K} \sim q(\boldsymbol{w}_{1:K} \mid s, g)} \left[\sum_{i=0}^K f_H(\boldsymbol{w}_i, \boldsymbol{w}_{i+1}) \right] - \sum_{i=1}^K D_{\text{KL}} (q_i(\boldsymbol{w}_i \mid s, g) \parallel p_{\hat{\pi}}^{\gamma}(\boldsymbol{w}_i)) \right\}$$

Tractable Reformulation

#2

$$\arg \max_q \left\{ \mathbb{E}_{\mathbf{w}_{1:K} \sim q(\mathbf{w}_{1:K} | s, g)} \left[\sum_{i=0}^K f_H(w_i, w_{i+1}) \right] - \sum_{i=1}^K D_{\text{KL}} (q_i(\mathbf{w}_i | s, g) \parallel p_{\hat{\pi}}^{\gamma}(\mathbf{w}_i)) \right\}$$

Assumption 2 & $f_H(w_i, w_{i+1}) = \psi(w_i) R \psi(w_{i+1})$

$\psi(w_i) \rightarrow z_i$

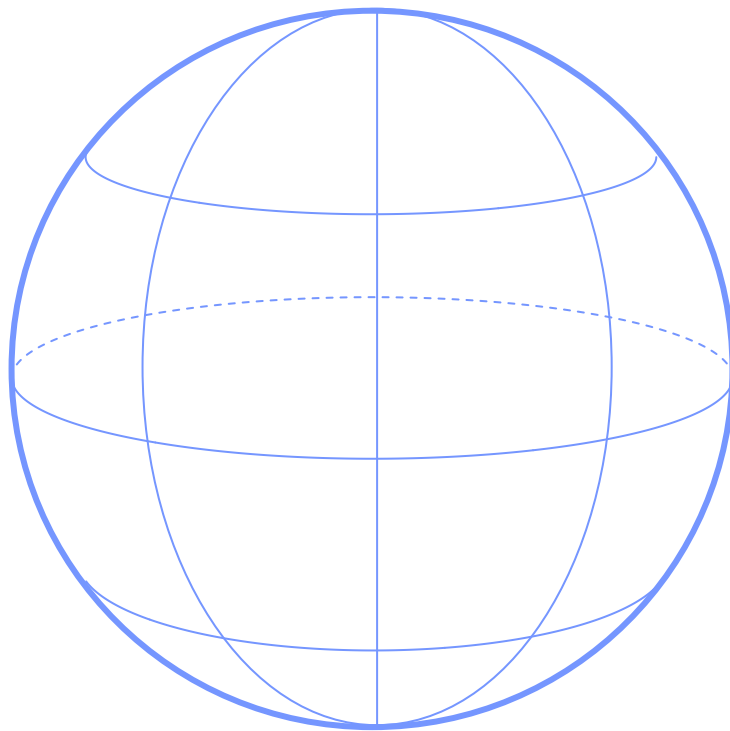
$q_i(w_i | s, g) \rightarrow q_i(z_i | s, g)$

#3

$$\arg \max_q \mathbb{E}_{\mathbf{z}_{1:K} \sim q(\mathbf{z}_{1:K} | s, g)} \left[\psi(s)^{\top} R z_1 + \sum_{i=2}^K z_{i-1}^{\top} R z_i + z_K^{\top} R \psi(g) \right] + \sum_{i=1}^K \mathcal{H} (q_i(\mathbf{z}_i | s, g))$$

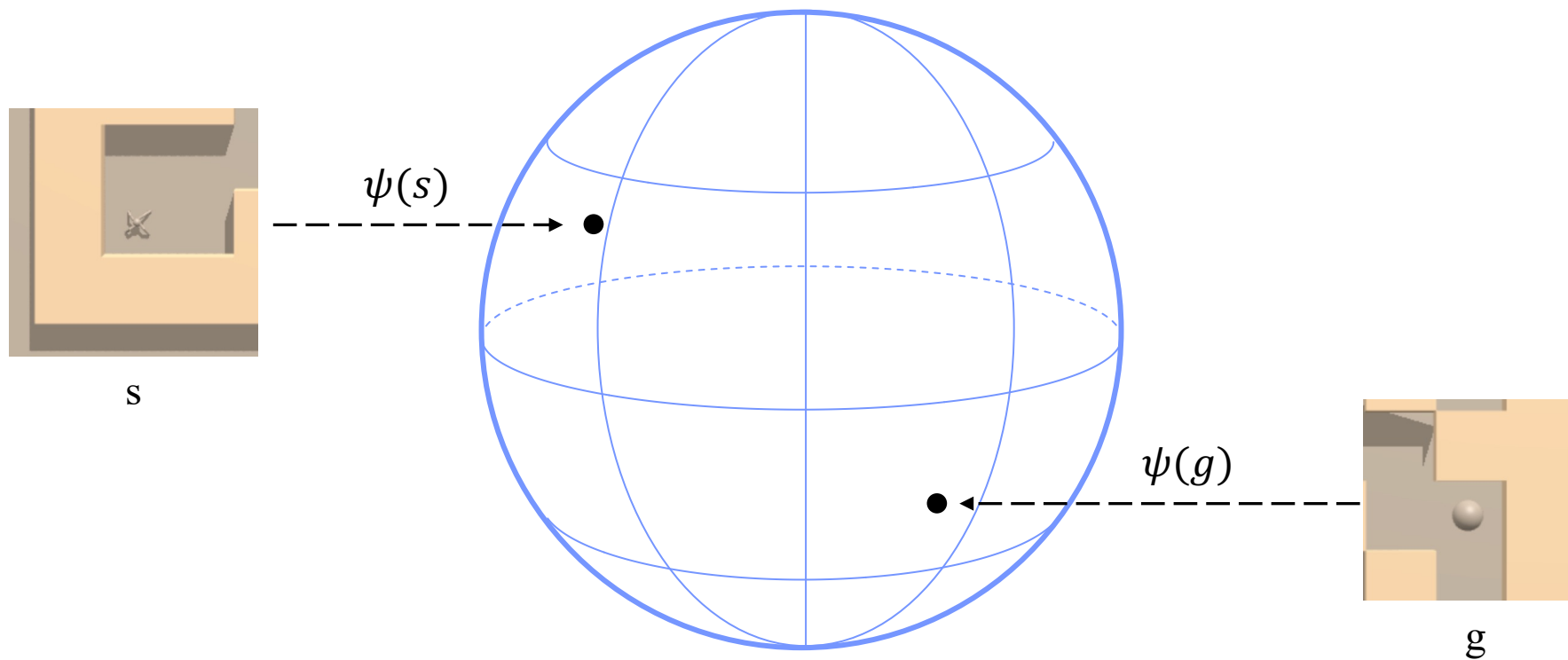
Geometric Interpretation

Geometric Interpretation



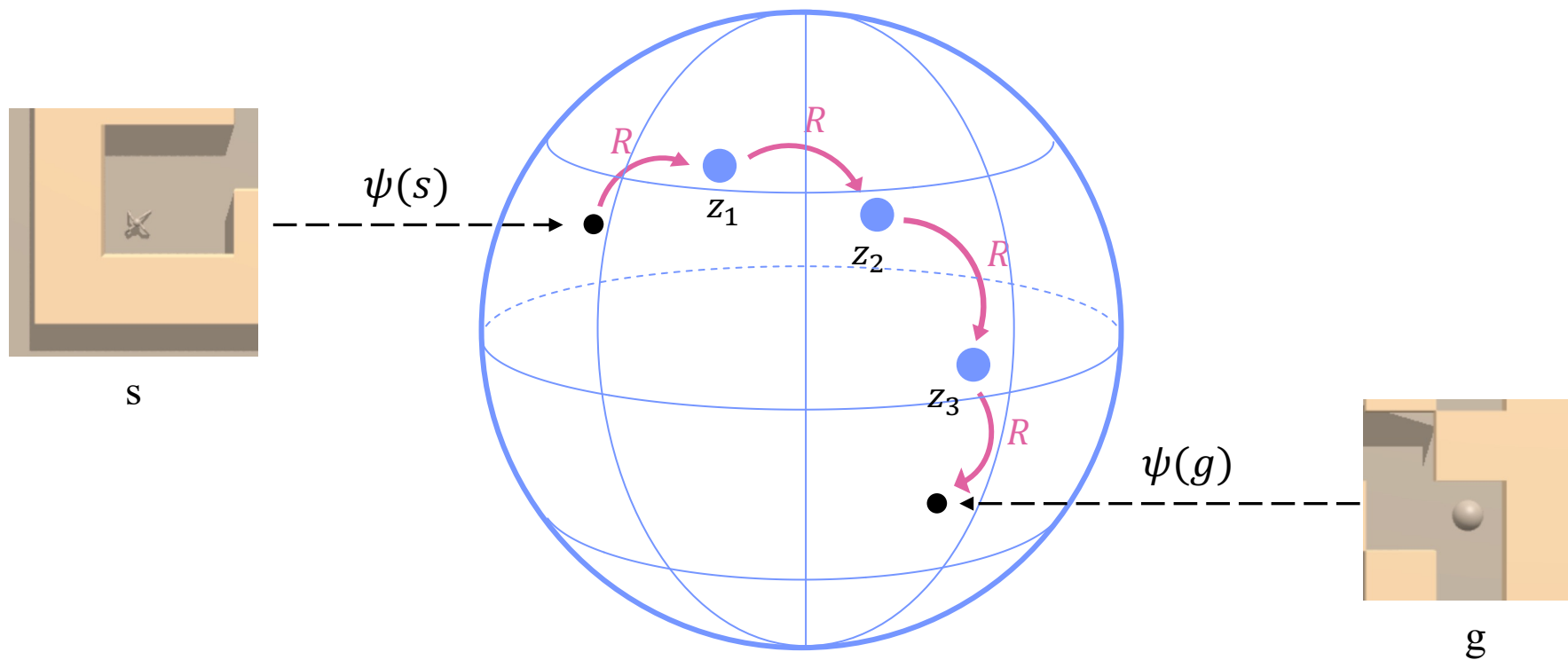
$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Geometric Interpretation



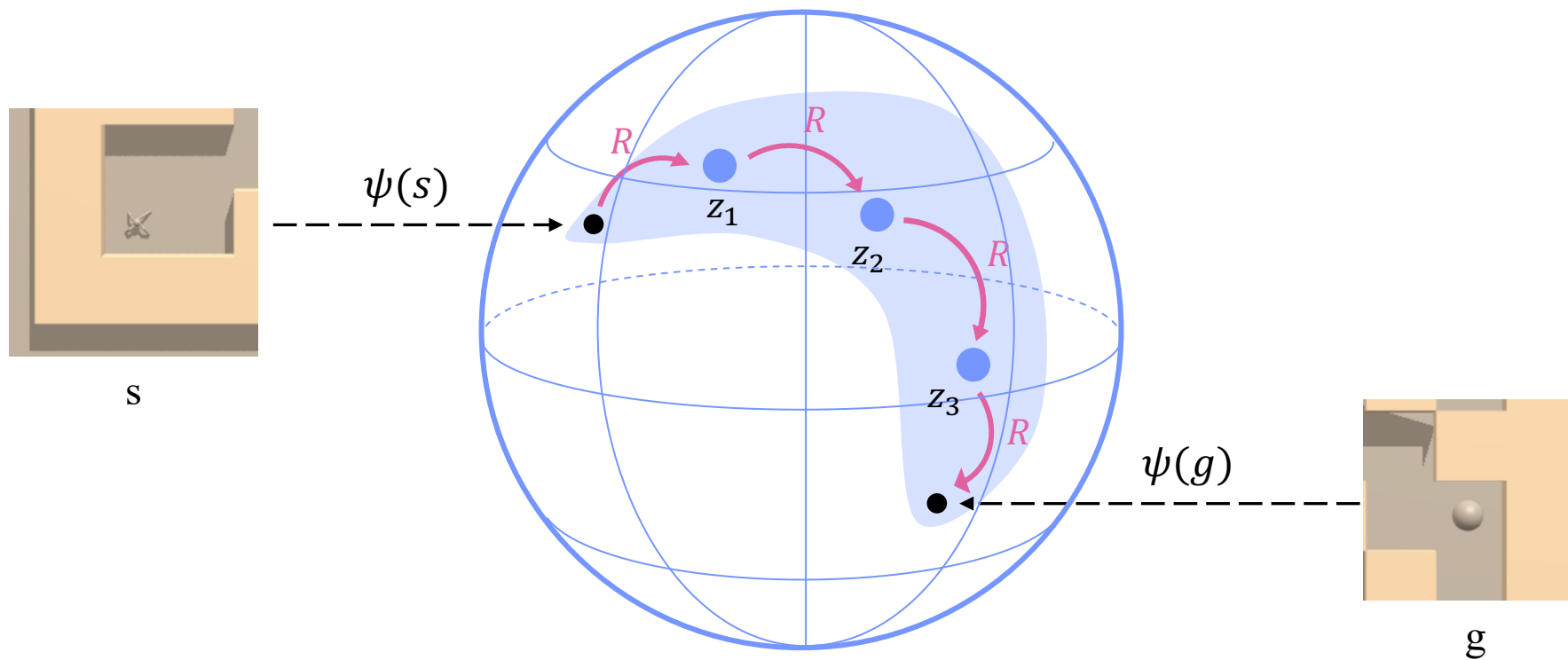
$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Geometric Interpretation



$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Geometric Interpretation



$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Closed form Solution

Closed form Solution

$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Closed form Solution

$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Parameterize q_i by normalizing **Gaussian samples** onto the unit **hypersphere** :

$$\tilde{z}_i \sim \mathcal{N}(\mu_i(s, g), \sigma_i^2(s, g)I), \quad z_i = \frac{\tilde{z}_i}{\|\tilde{z}_i\|_2} \sim q_i(\cdot | s, g)$$

Closed form Solution

$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

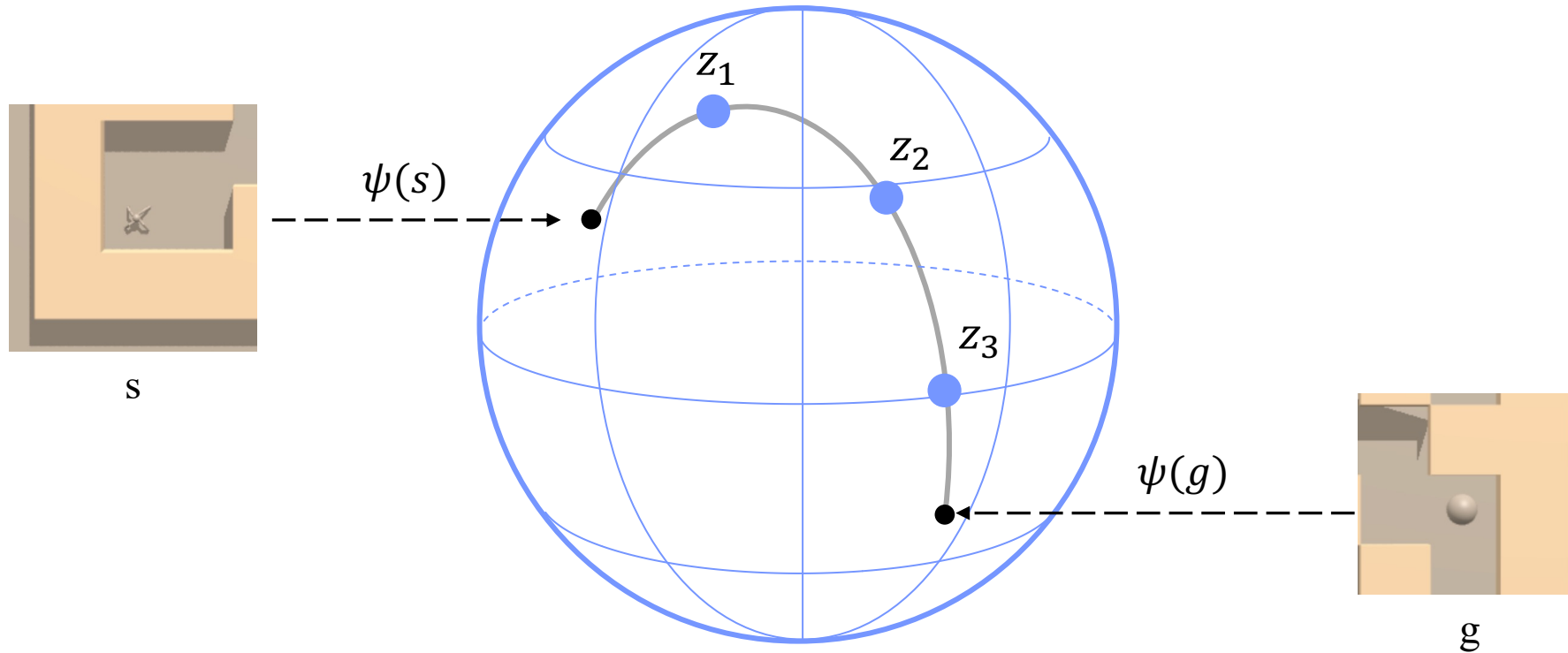
Under the small-variance assumption,

$$\begin{cases} \mu_i^*(s, g) = (R^T)^i \text{slerp}\left(\psi(s), R^{K+1}\psi(g), \frac{i}{K+1}\right), \\ \sigma_i^{*2}(s, g) = \frac{\alpha}{2 \cos(\frac{\theta}{K+1})} \end{cases} \quad \text{for } i = 1, \dots, K,$$

$$^*\theta = \arccos(\psi(s)^\top R^{K+1}\psi(g))$$

Closed form Solution

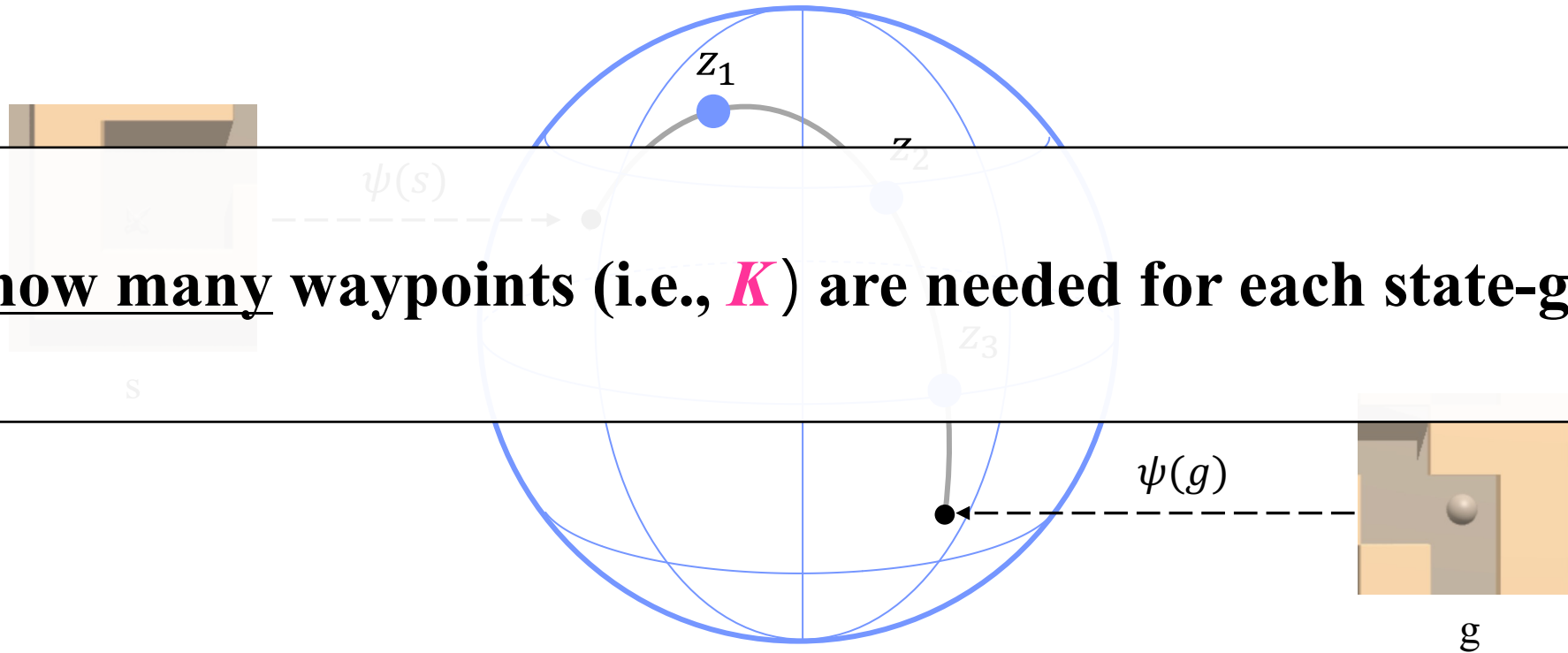
$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$



Closed form Solution

$$\arg \max_q \mathbb{E}_{z_{1:K} \sim q(z_{1:K} | s, g)} \left[\psi(s)^\top R z_1 + \sum_{i=2}^K z_{i-1}^\top R z_i + z_K^\top R \psi(g) \right] + \sum_{i=1}^K \mathcal{H}(q_i(z_i | s, g))$$

Then, how many waypoints (i.e., K) are needed for each state-goal pair?



K-selection strategy

K-selection strategy

$$f_H(s, s^+) = \psi(s)^\top R \psi(s^+) = (R^\top \psi(s))^\top \psi(s^+) \quad : \text{one geometric transition}$$

K-selection strategy

$$f_H(s, s^+) = \psi(s)^\top R \psi(s^+) = (R^\top \psi(s))^\top \psi(s^+) \quad : \text{one geometric transition}$$

$$\ell_K(s, g) = ((R^\top)^{K+1} \psi(s))^\top \psi(g) \quad : K+1 \text{ geometric transitions}$$

K-selection strategy (see Appendix A.4 for the detailed derivation)

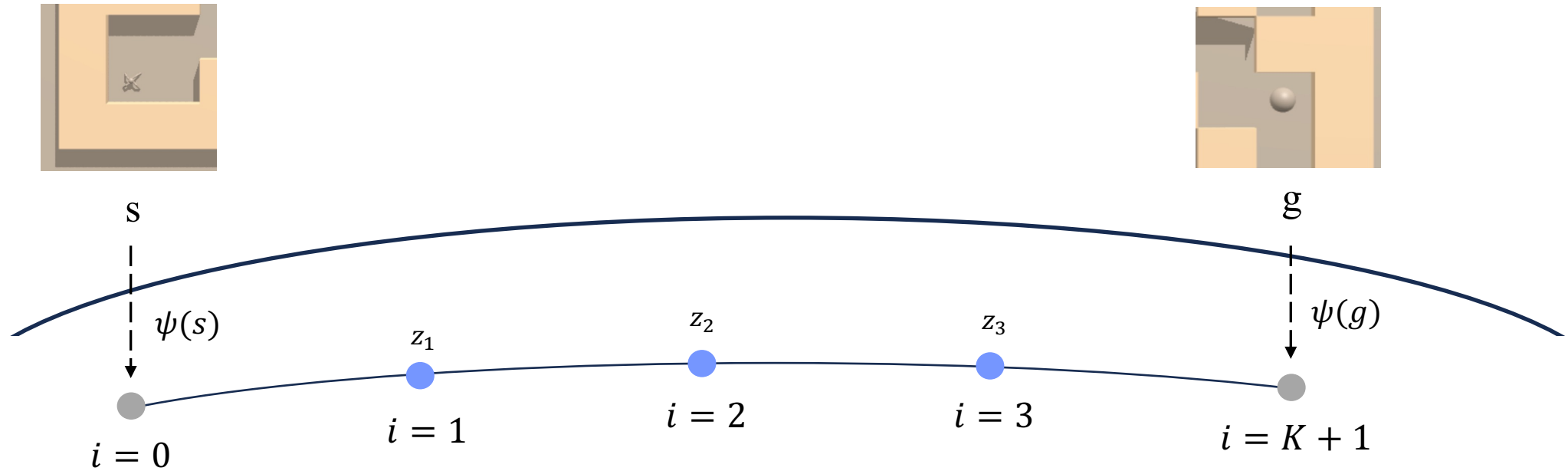
$$f_H(s, s^+) = \psi(s)^\top R \psi(s^+) = (R^\top \psi(s))^\top \psi(s^+) \quad : \text{one geometric transition}$$

$$\ell_K(s, g) = ((R^\top)^{K+1} \psi(s))^\top \psi(g) \quad : K+1 \text{ geometric transitions}$$

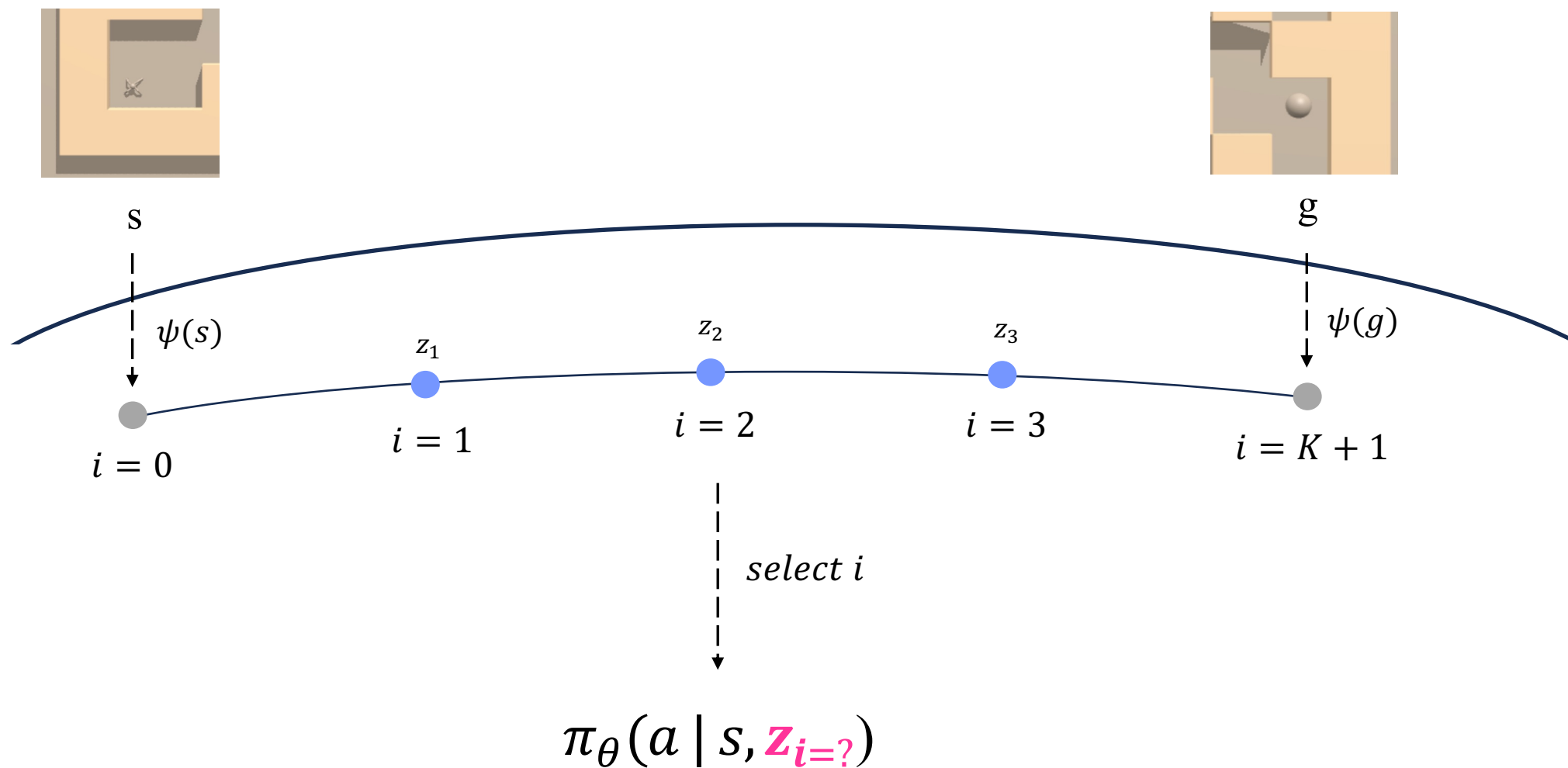
↓ vMF likelihood & Bayes' rule

$$\hat{p}(K \mid \psi(s), \psi(g)) = \text{Softmax}_{K \in \{0, \dots, K_{\max}\}} \left(\frac{\ell_K(s, g)}{\eta} \right)$$

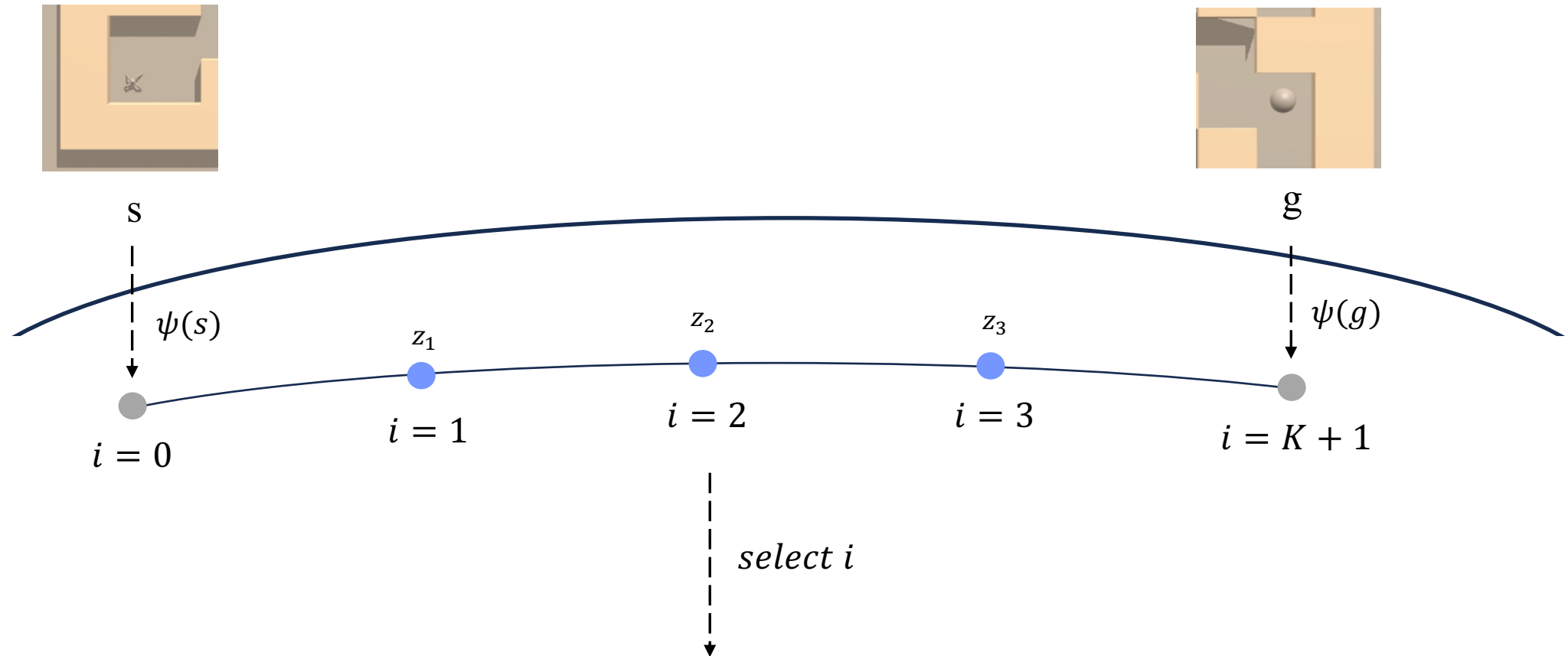
i – selection strategy



i – selection strategy

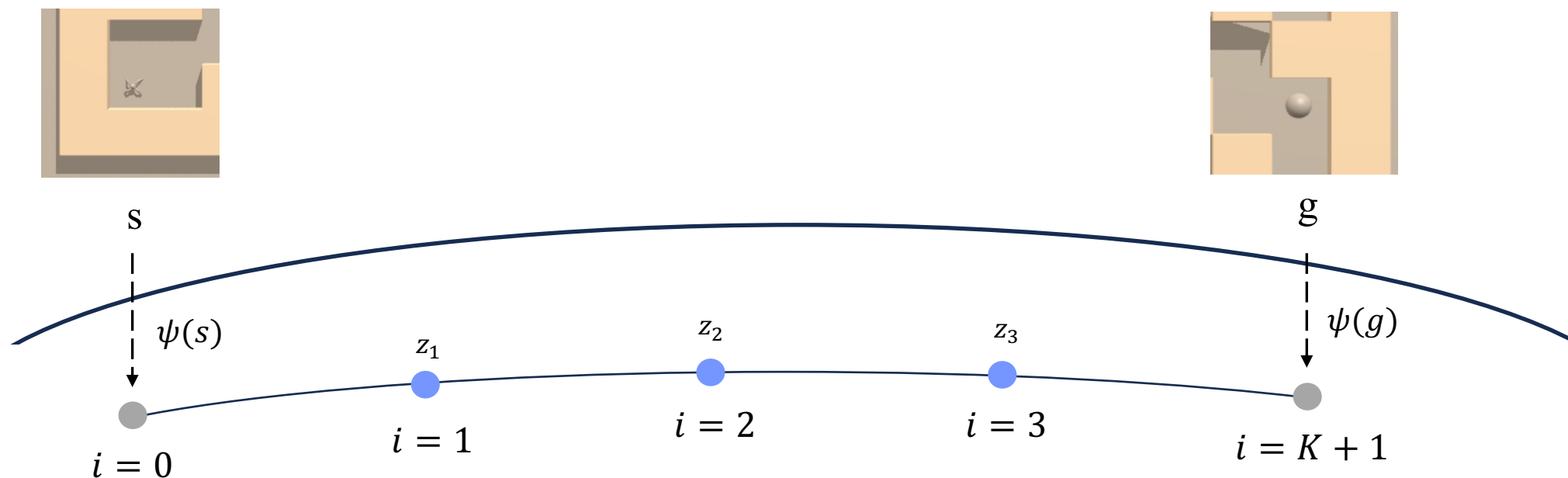


i – selection strategy



Are subgoals reachable by the low-level policy for all i ? 🤔

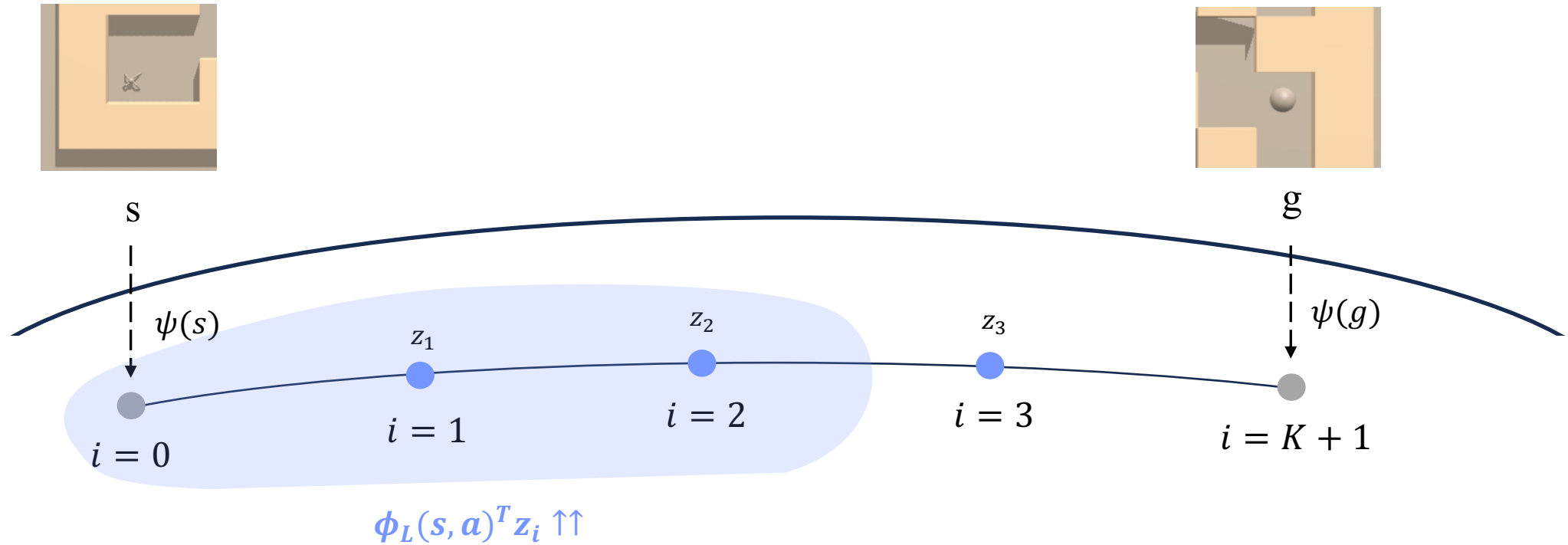
Proxy for reachability



- Recall the policy learning objective

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, w \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(w))} [\phi(s, a)^T \underbrace{\psi(w)}_{\mathbf{z}_i}]$$

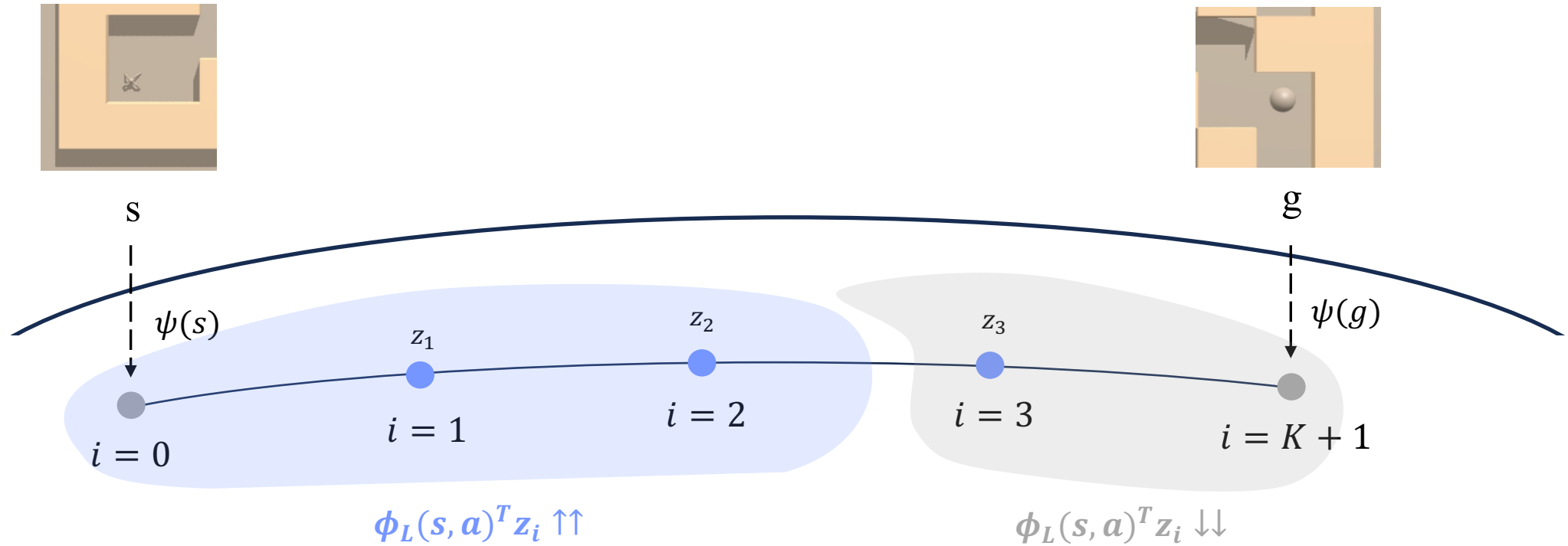
Proxy for reachability



- Recall the policy learning objective

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, w \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(w))} [\phi(s, a)^T \underbrace{\psi(w)}_{z_i}]$$

Proxy for reachability

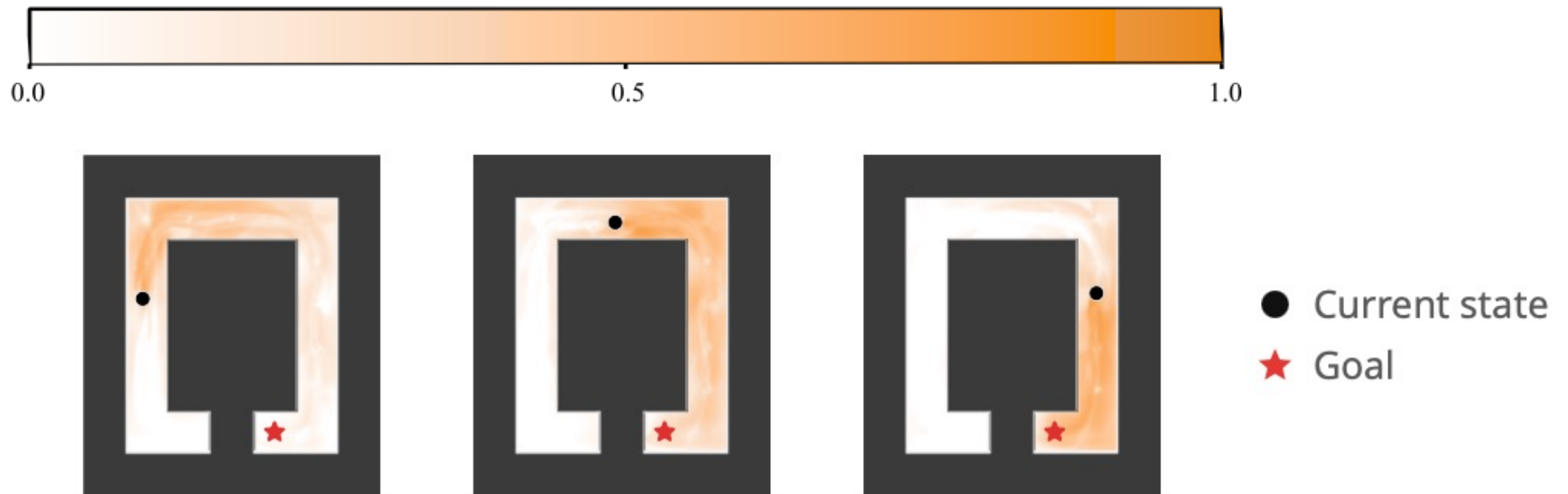


- Recall the policy learning objective

$$\max_{\theta} \mathbb{E}_{s \sim p_{\mathcal{D}}, w \sim p_{\mathcal{D}}^{\gamma}(\cdot | s), a \sim \pi_{\theta}(\cdot | s, \psi(w))} [\phi(s, a)^T \underbrace{\psi(w)}_{z_i}]$$

Low level reachability : visualization

$$\text{Low level reachability score : } f_L(s, a, w) = \phi(s, a)^T \underbrace{\psi(w)}_{\mathbf{z}_i}$$



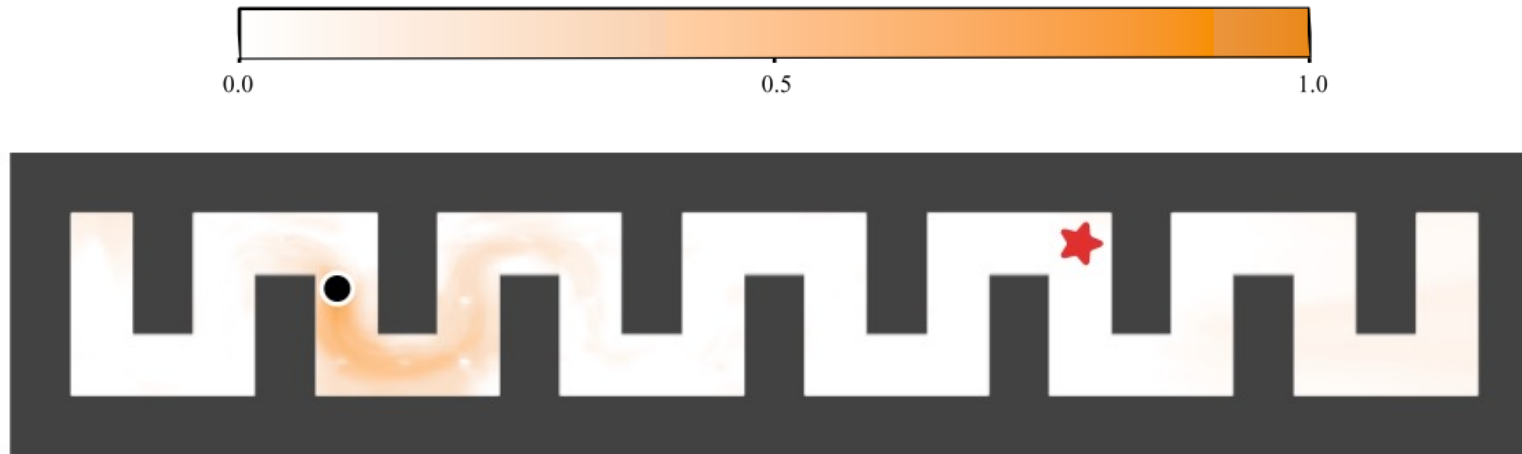
Higher score indicates greater estimated reachability of the current policy

i – selection strategy

i – selection strategy

1. Compute the f_L score for each candidate waypoint representation z_i

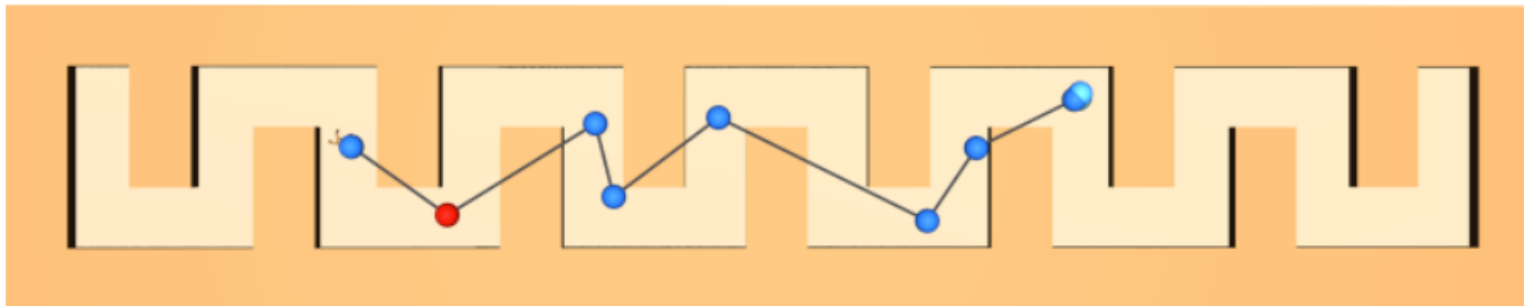
$$\rho_i = \phi(s, a_i)^\top z_i, \quad a_i \sim \pi_\theta(\cdot \mid s, z_i)$$



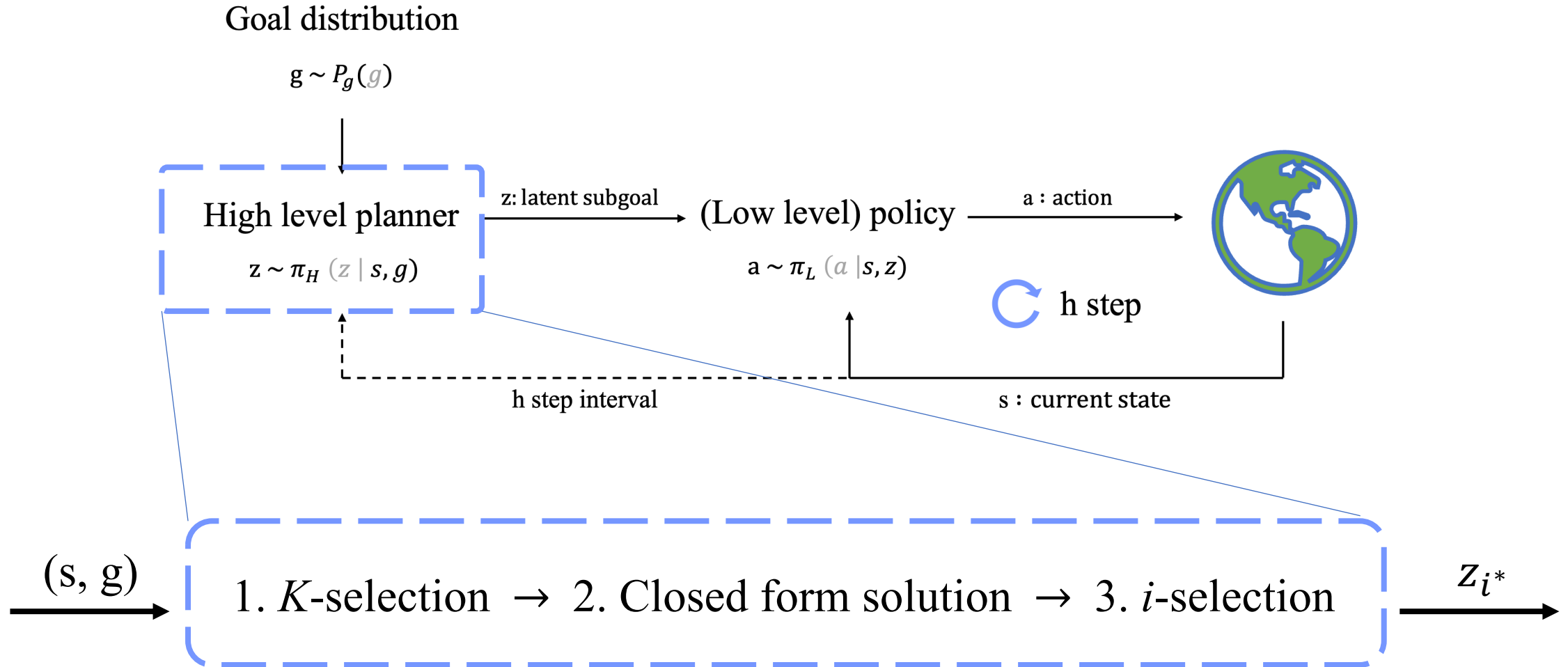
i – selection strategy

1. Compute the f_L score for each candidate waypoint representation z_i
2. Select the **furthest waypoint** whose score meets the threshold ϵ

$$i^* = \max \{i \in \{1, \dots, K + 1\} : \rho_i \geq \epsilon\}$$

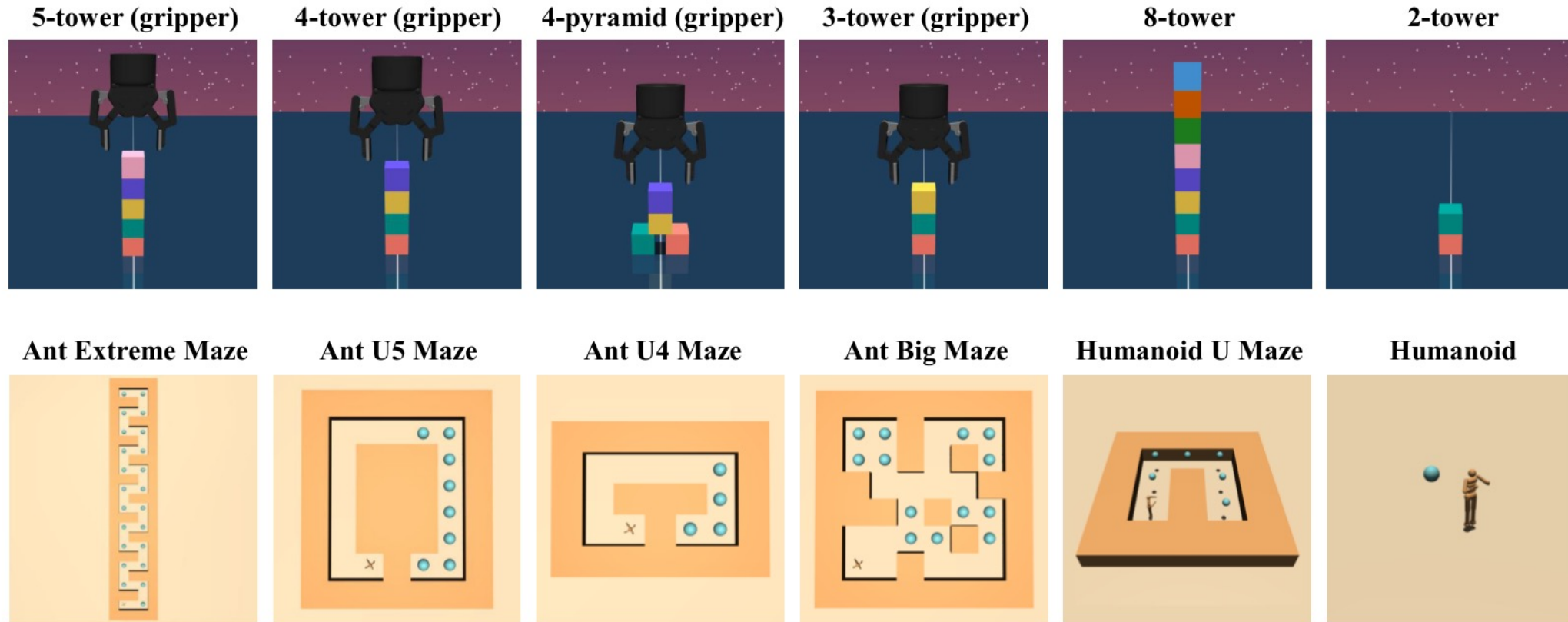


Overall planning algorithm



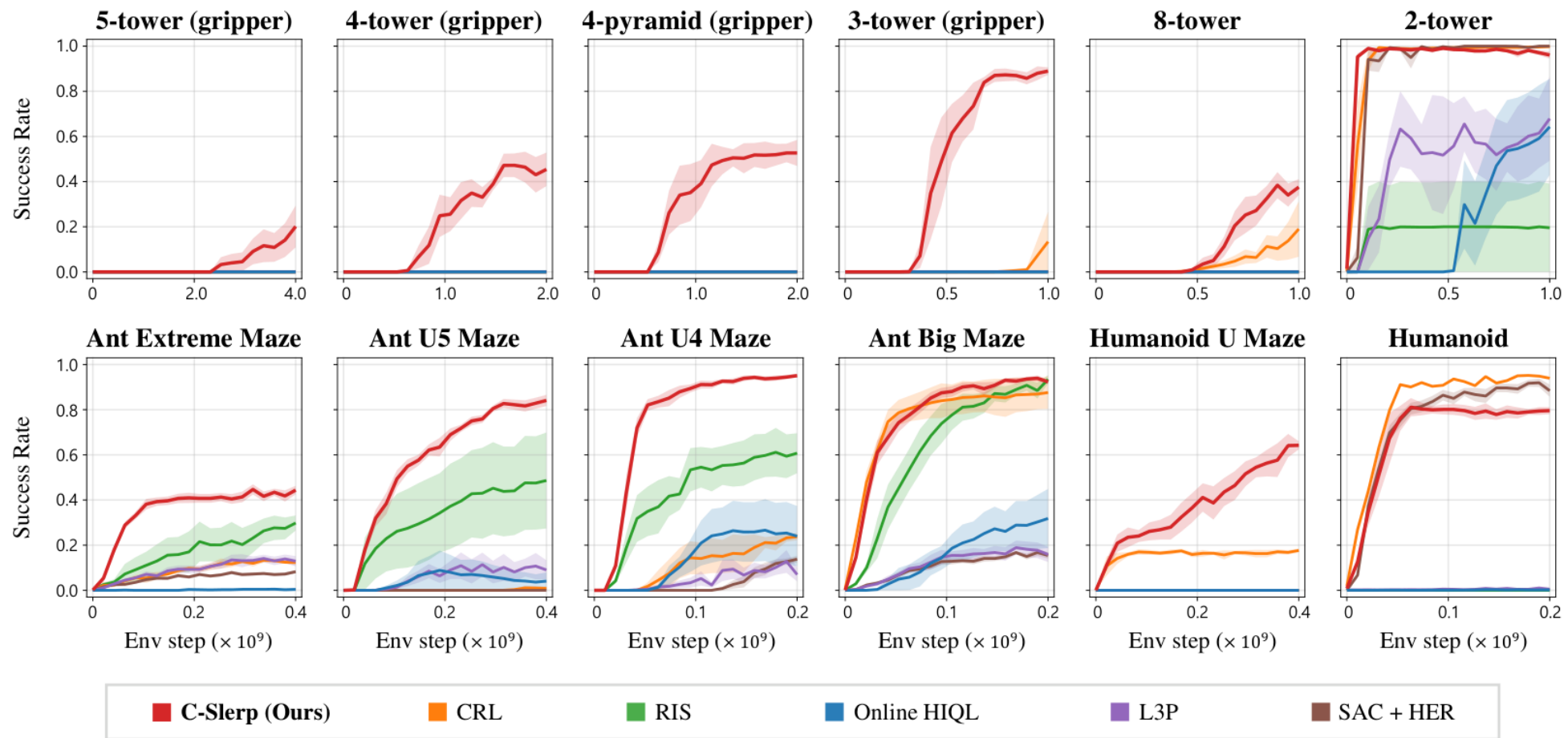
Experimental Results 🙄

Benchmark Environments



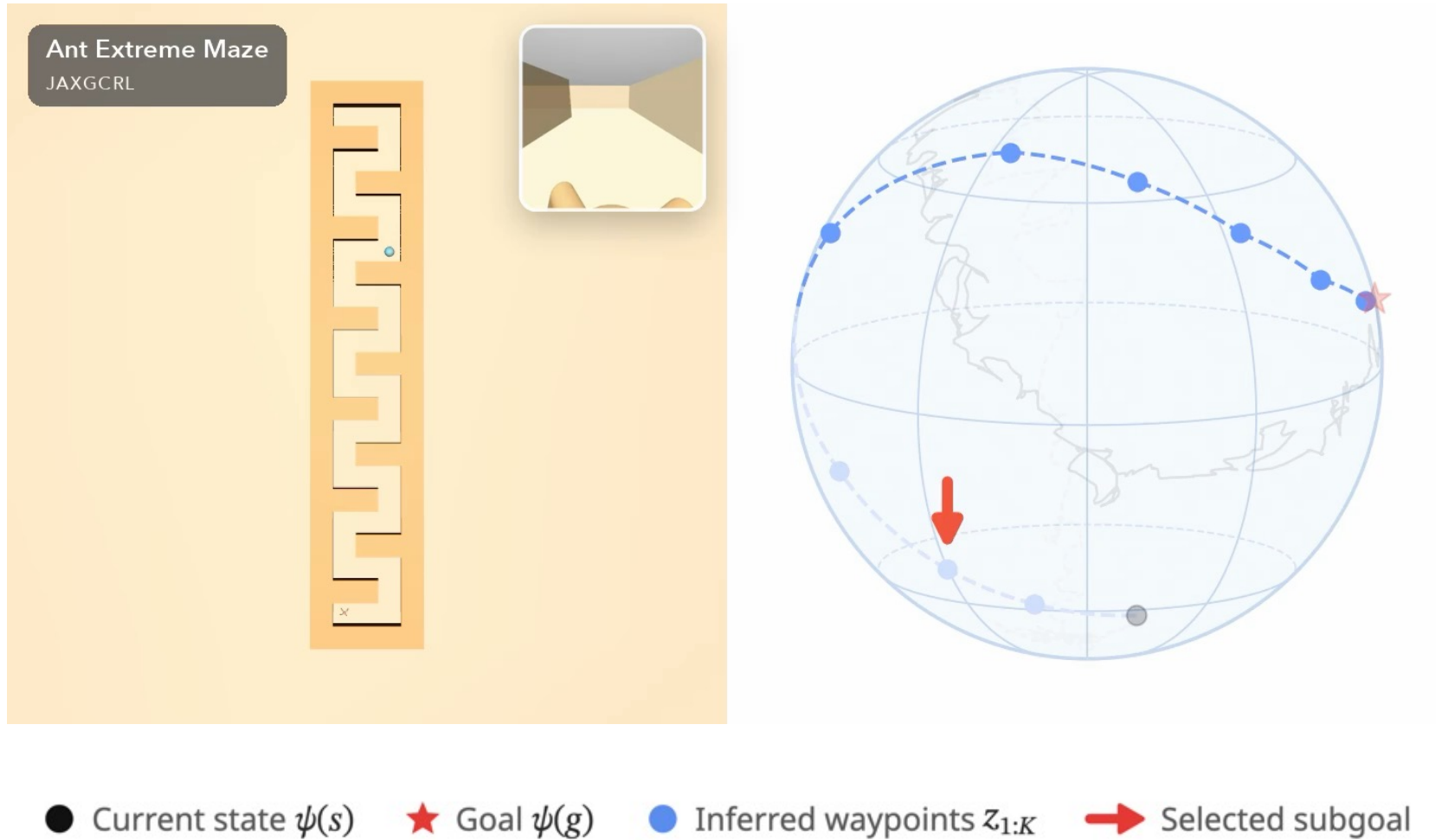
Six manipulation tasks (**BuilderBench**) and six locomotion tasks (**JaxGCRL**)

Main Results : BuilderBench (top) and JaxGCRL (bottom)

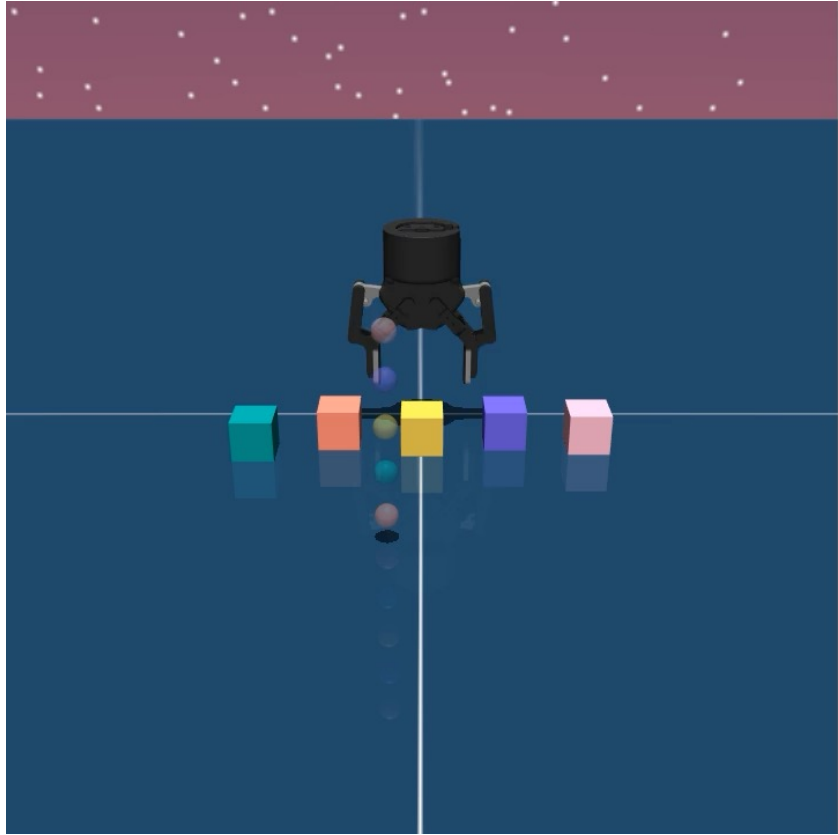
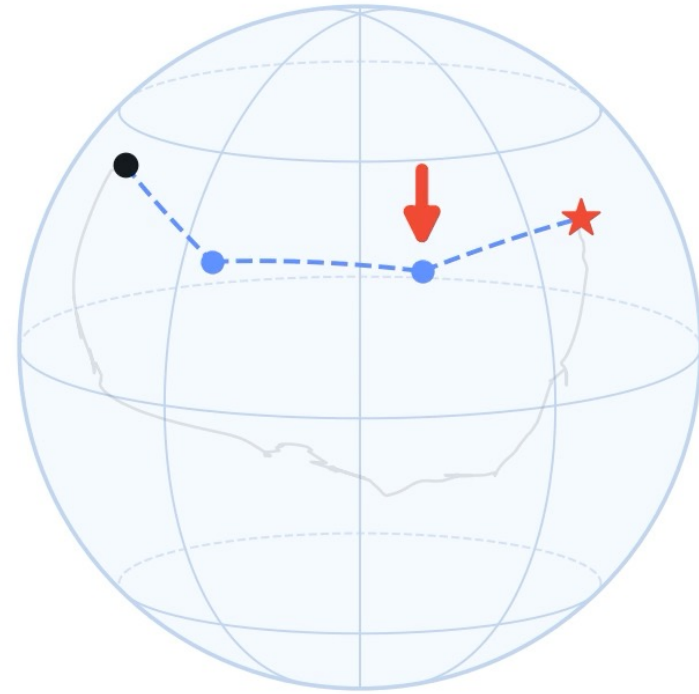


Our method solves long-horizon tasks where baselines **never achieve a single success**.

Latent Visualization : JAXGCRL

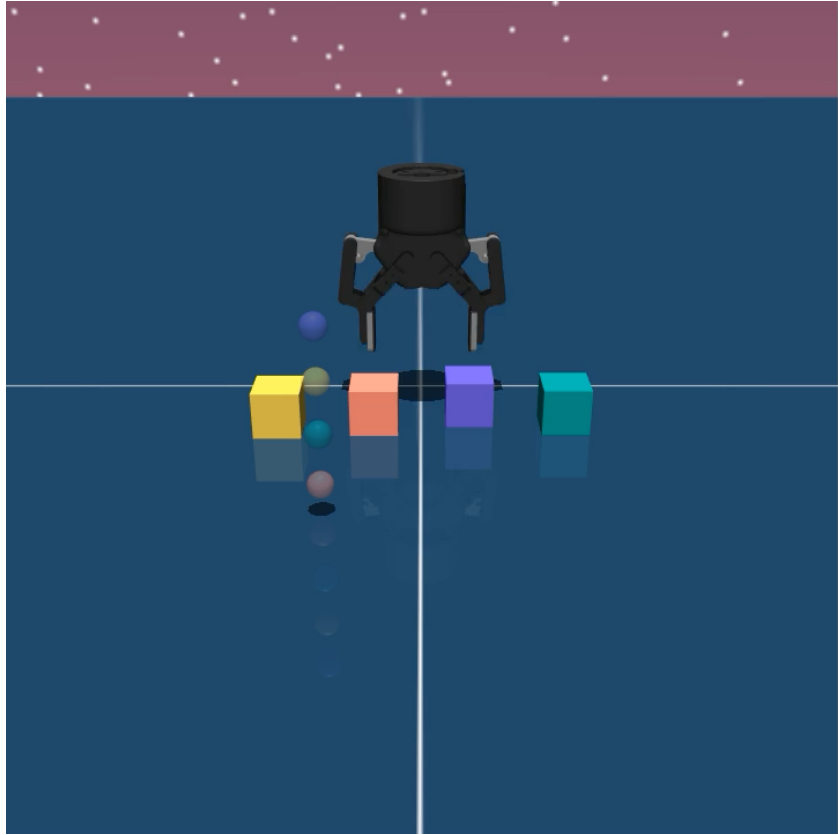
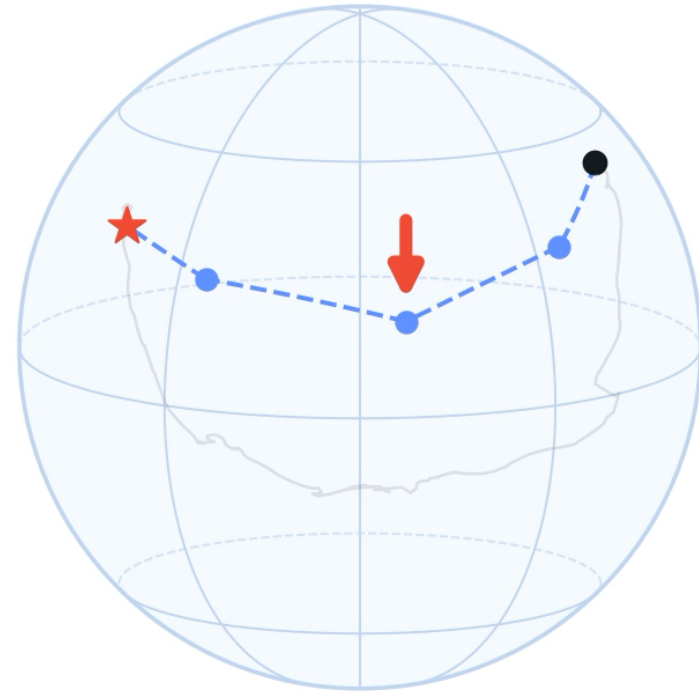


Latent Visualization : BuilderBench (5-tower)



● Current state $\psi(s)$ ★ Goal $\psi(g)$ ● Inferred waypoints $z_{1:K}$ → Selected subgoal

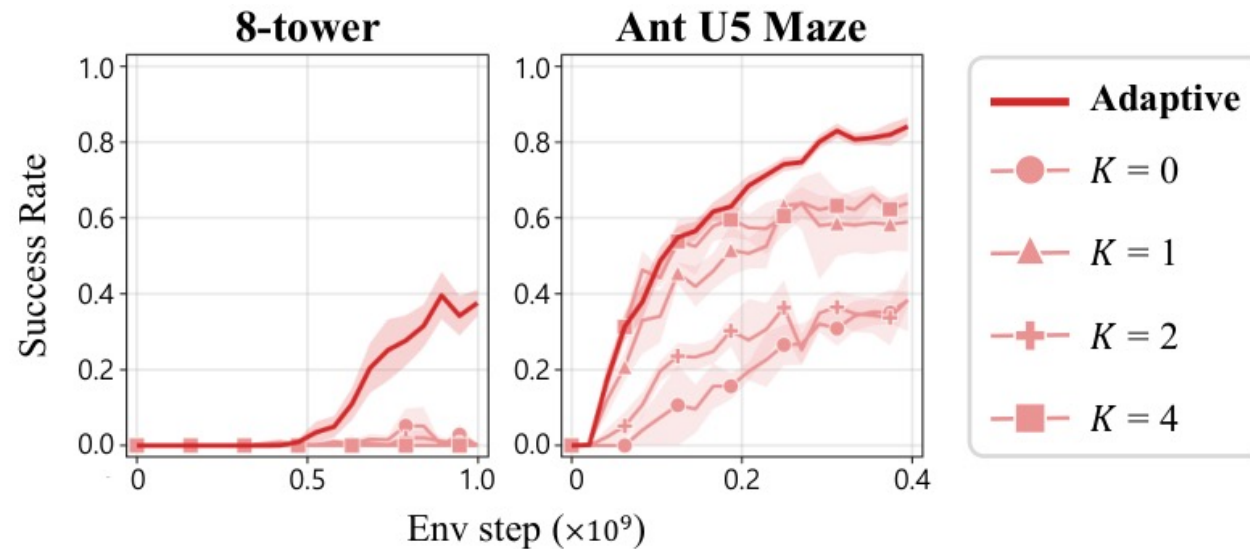
Latent Visualization : BuilderBench (4-tower)



● Current state $\psi(s)$ ★ Goal $\psi(g)$ ● Inferred waypoints $z_{1:K}$ → Selected subgoal

Does adaptivity matter? : ablation on K -selection

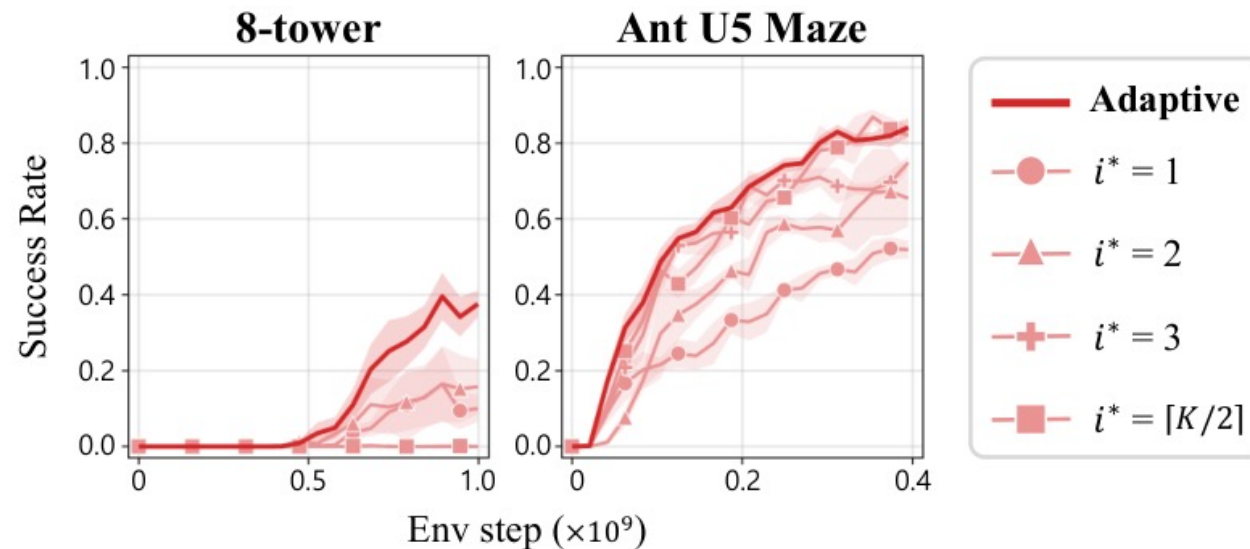
Fixing the number of waypoints — at *any* value — substantially **degrades** performance.



- More waypoints is *not* consistently better.
- The right K also changes *within* an episode as the agent nears the goal.

Does adaptivity matter? : ablation on i -selection

The best fixed choice of i **differs** across environments; **adaptive selection is best** in both.

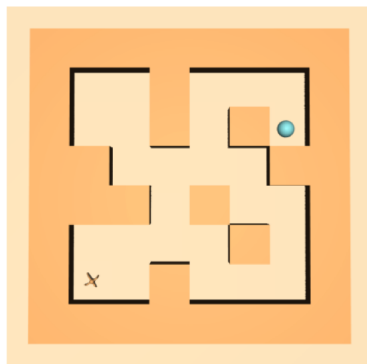


- Early waypoints are safe but slow; late ones make progress but may be unreachable.
- Adaptive i -selection method picks the furthest reachable one.

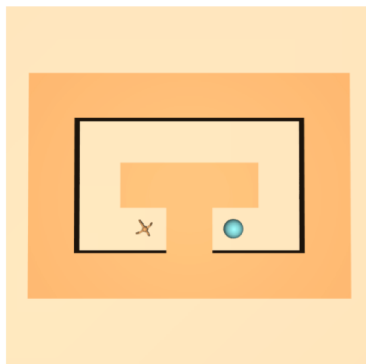
Exploration in single-goal settings

With only **one distant goal ever provided**, our method still achieves a high success rate — **directed exploration** without a goal curriculum.

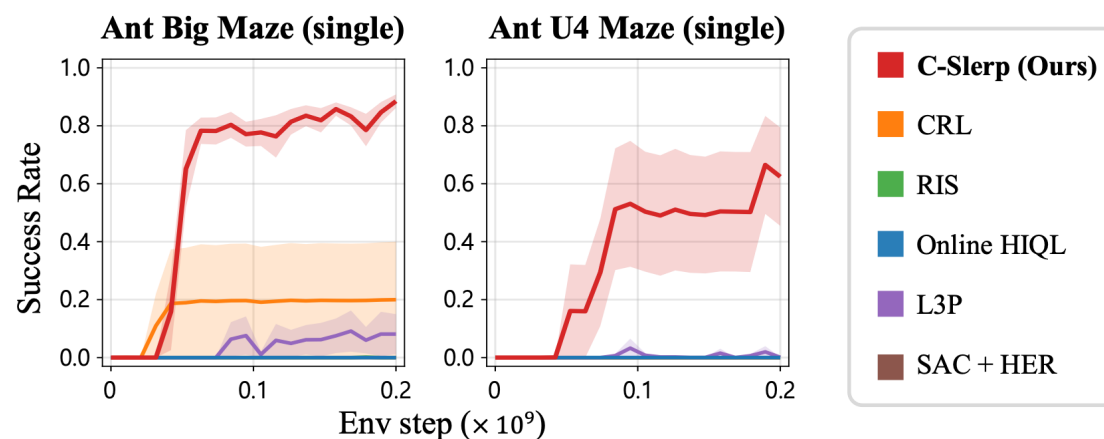
Ant Big Maze (single)



Ant U4 Maze (single)



(a) Single-goal environments



(b) Learning curves

Takeaways

Takeaways

1 Adaptive planning. Adapting task decomposition to the goal-reaching problem is key to the benefits of hierarchy — fixing the number of planning stages substantially degrades performance.

Takeaways

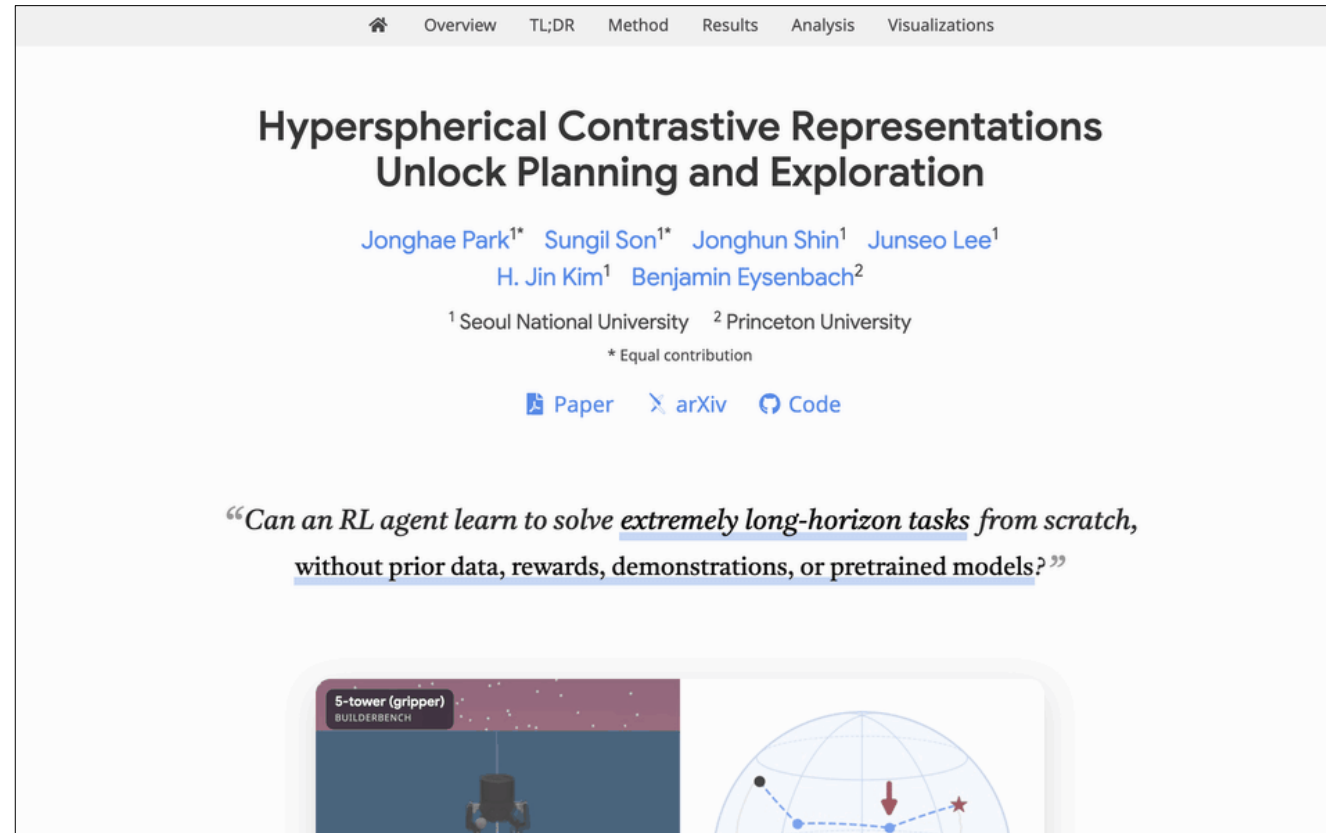
1 Adaptive planning. Adapting task decomposition to the goal-reaching problem is key to the benefits of hierarchy — fixing the number of planning stages substantially degrades performance.

2 Exploration. Our method solves long-horizon tasks where all evaluated baselines fail to achieve a single success, without a goal curriculum or dense rewards.

Takeaways

- 1 **Adaptive planning.** Adapting task decomposition to the goal-reaching problem is key to the benefits of hierarchy — fixing the number of planning stages substantially degrades performance.
- 2 **Exploration.** Our method solves long-horizon tasks where all evaluated baselines fail to achieve a single success, without a goal curriculum or dense rewards.
- 3 **No extra architecture.** Planning follows directly from spherical linear interpolation in the learned representation space — no separate high-level network, no planning graph.

For more results & derivations ..



<https://jonghaepark.github.io/c-slerp/>

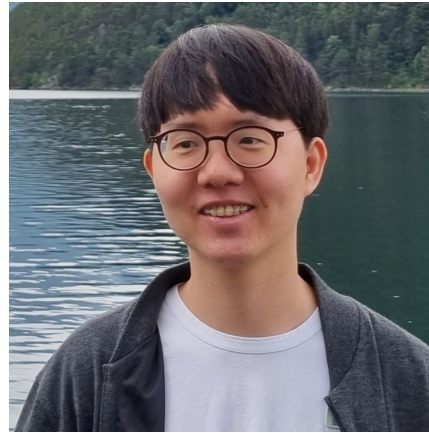
Thanks to my *collaborators*!



Sungil



Jonghun



Junseo



H. Jin Kim



Benjamin Eysenbach

Thank you

Any questions?