ID-WM · WM@BOOTH

Hierarchical World Models with Implicit Dynamics

Gaoyue Zhou[1,3] · Yvonne Wu[1] · Zichen Jeff Cui[1] · Nicolas Ballas[3] · Mido Assran[3] · Lerrel Pinto[1,2] · Yann LeCun[1,3]

[1]Courant Institute, NYU · [2]Meta-FAIR · [3]AMI Labs · gz2123@nyu.edu

id-wm.github.io

01 · THE PROBLEM

Long horizons break WM planning

Search complexity grows exponentially with the horizon and prediction errors accumulate over long rollouts, so most world model planners stay short horizon.

02 · WHY HIERARCHY FAILS

Skills and values are task-specific

Rigid macro actions, latent skills and value functions are brittle, yet without such anchors a powerful optimizer exploits the model and discovers "wormholes".

03 · OUR INSIGHT

Decouple reachability from control

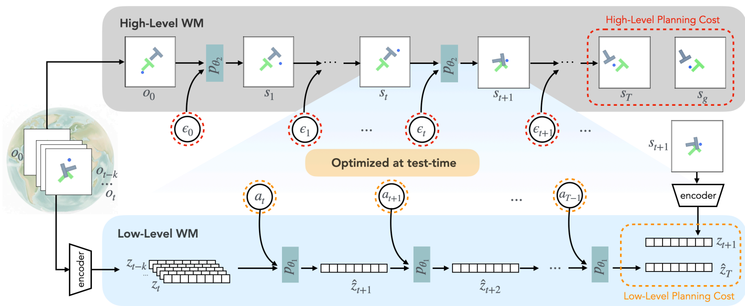
A high level diffusion world model samples feasible transitions across the state manifold, so the planner searches for *reachable subgoals* in a unit Gaussian; a low level world model grounds them into precise actions.

+38.5% over policy learning, offline RL and diffusion planners

+135% over single level world model with MPC

Method

subgoals and actions, optimized in interleaved fashion



f_{hi}: diffusion high level world model. f_{lo}: deterministic transformer low level world model. Subgoals $s_1, \dots, s_{T\Delta}$ and actions a_t are optimized in an interleaved fashion, both in latent space.

HIGH LEVEL: REACHABLE STATE GENERATOR

$$f_{hi}(\, s_{t+\Delta} \mid s_t \,, \Delta \,)$$

Conditional DDPM over states Δ steps away:

$$\mathcal{L}_{diff} = \mathbb{E}\, \| \, \varepsilon - \varepsilon_\theta(x_k, k, s_t) \, \|^2$$

At test time DDIM maps noise $z_t \sim \mathcal{N}(0, \mathbf{I})$ deterministically to a subgoal:

$$\hat{s}_{t+\Delta} = \mathcal{G}_\theta(s_t, z_t)$$

z_t is an **implicit high level action**. Unlike discrete skill indices or learned latent codes, which suffer mode collapse and topological discontinuities, the noise space is isotropic and smooth: every sample decodes to a physically valid reachable state.

LOW LEVEL: INVERSE DYNAMICS SOLVER

$$f_{lo}(\, s_{t+1} \mid s_{t-H:t} \,, a_{t-H:t} \,)$$

Deterministic transformer over H past latent states and actions, trained on $\|\hat{s}_{t+1} - s_{t+1}\|_2$. At planning time it acts as an inverse dynamics solver, recovering the actions that reach the next subgoal, never more than Δ steps away.

HIERARCHICAL MPC: CEM IN NOISE SPACE

$$\mathcal{L} = \| \, \hat{s}_H - s_g \, \|_2 + \alpha \, \mathcal{L}_{reg}(Z)$$

$$\mathcal{L}_{reg} = \mathbb{E}_i \, [\, (\, \max(0, |z_i| - \tau) \,)^p \,]$$

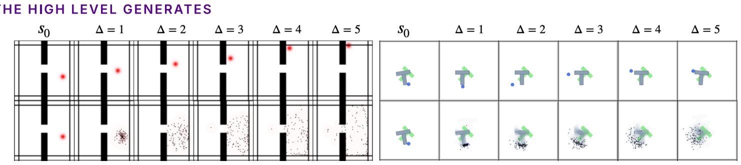
The barrier keeps latents in the prior's high density region, blocking the out of distribution transitions an unconstrained optimizer would otherwise exploit.

HIERARCHICAL PLANNING: PSEUDOCODE

```
in: obs o_0, goal o_g, models f_hi / f_lo, radius Δ,
horizon H
while o_g not reached:
  — high level: search noise
  Z ← CEM_Z ‖ ŝ_H - s_g ‖ + α L_reg(Z)
  s*_1..H ← DDIM(s_t, Z)   Δ steps apart
  for each subgoal s*_i:
    — low level: search actions
    a_1..Δ ← CEM_a ‖ f_lo(s_t, a) - s*_i ‖
      execute a_1..Δ, observe s_t
  replan from s_t
```

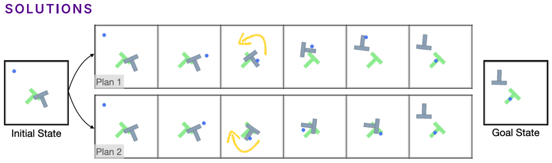
Only the two world models and CEM. The high level plan is refreshed at every MPC iteration.

WHAT THE HIGH LEVEL GENERATES



Top: ground truth Δ steps ahead. Bottom: 100 sampled latents overlaid.

ONE GOAL, SEVERAL SOLUTIONS



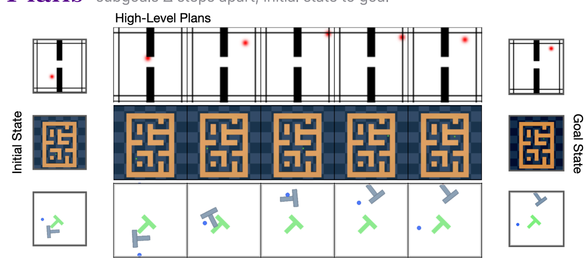
Unconstrained by the behaviour modes in the data: Plan 1 rotates the T block counterclockwise, Plan 2 clockwise.

TAKEAWAYS

- **Diffusion noise as a latent action space**, constraining search to a unit Gaussian. No reward, value function or pretrained skills.
- **Substantially outperforms state of the art baselines**, and is the only method that solves PushT with random goals.
- **Grounding hierarchy in local dynamics scales**, but latent search stays slower than a policy and grows with state dimension.

Plans

subgoals Δ steps apart, initial state to goal



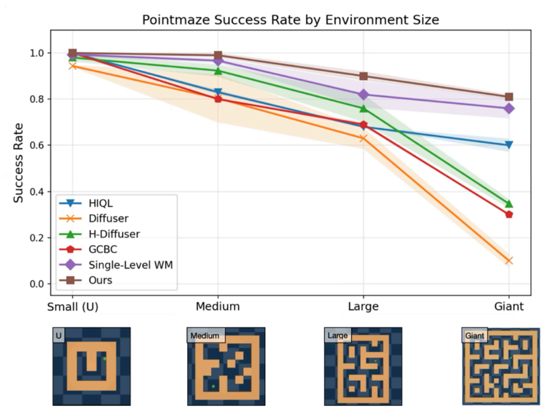
Results

success rate · 50 pairs · 3 seeds

Method	Med. Maze	Wall	PushT fix	PushT rand.	Aug fix	Aug rand.
HIQL	0.83	0.62	0.45	0.02	0.87	0.32
GCBC	0.80	1.00	1.00	0.21	1.00	0.15
Diffuser	0.80	0.07	0.00	0.00	0.01	0.01
H-Diffuser	0.92	0.10	0.01	0.01	0.55	0.27
Single level WM + MPC	0.97	0.88	0.19	0.21	0.10	0.10
ID-WM (ours)	**0.99**	**1.00**	**1.00**	**0.77**	**1.00**	**1.00**

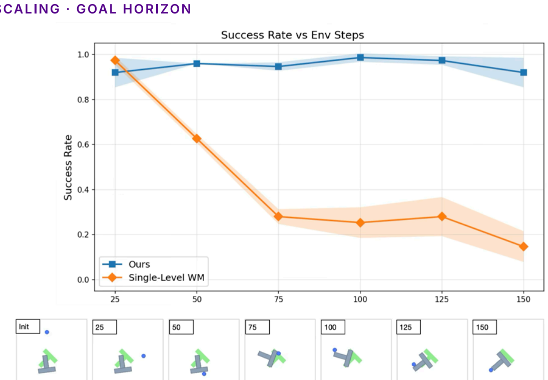
ID-WM is the only method above 30% on PushT with random goals; single level MPC fails once horizons exceed 100 steps.

SCALING · MAZE SIZE



Fixed data budget, so coverage thins as the maze grows. ID-WM stays strongest on the Giant maze.

SCALING · GOAL HORIZON



Goals 25 to 150 steps away. The single level WM decays past 50 steps.

Ground hierarchy in *local* dynamics — long horizon plans assemble themselves from short, valid segments.

NYU · Meta · Advanced Machine Intelligence