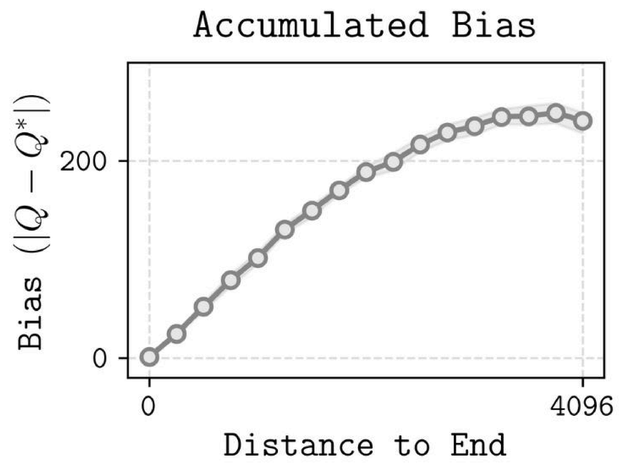
Offline RL with Hierarchical Action Chunking

Ahad Jawaid
The University of Texas at Dallas

ICLR 2026

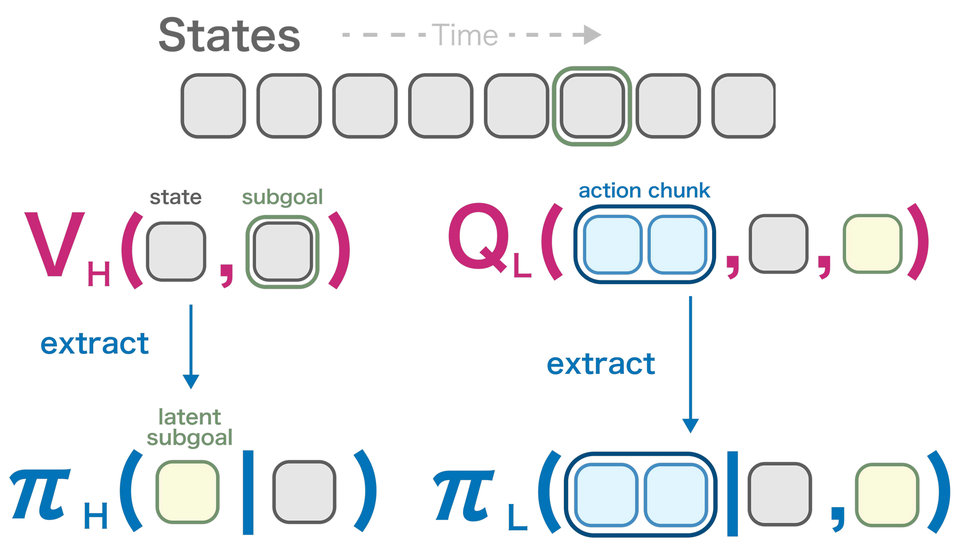
Motivation

- Offline RL suffers from compounding value errors over long chains of bootstrapped backups called the *"curse of horizon"*.



Overview

- **Hierarchical Implicit Q-Chunking (HiQC):** A hierarchical offline RL algorithm merging high-level latent planning with low-level action chunking.
- **Dual Horizon Reduction:** Compresses both planning (via subgoals) and execution (via action chunks).
- **Unbiased Value Propagation:** Enables stable, off-policy-free multi-step value backups.

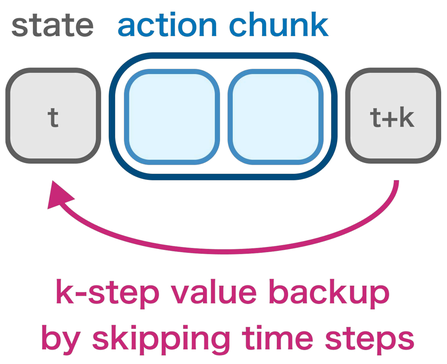


Methodology

- **High-Level (Horizon c):** Predicts latent subgoals z via Implicit V-Learning.
- **Low-Level (Horizon k):** Executes k-step action chunks a to reach z.
- **Chunked Critic:** Conditions Q-function on full chunk a to remove bias:

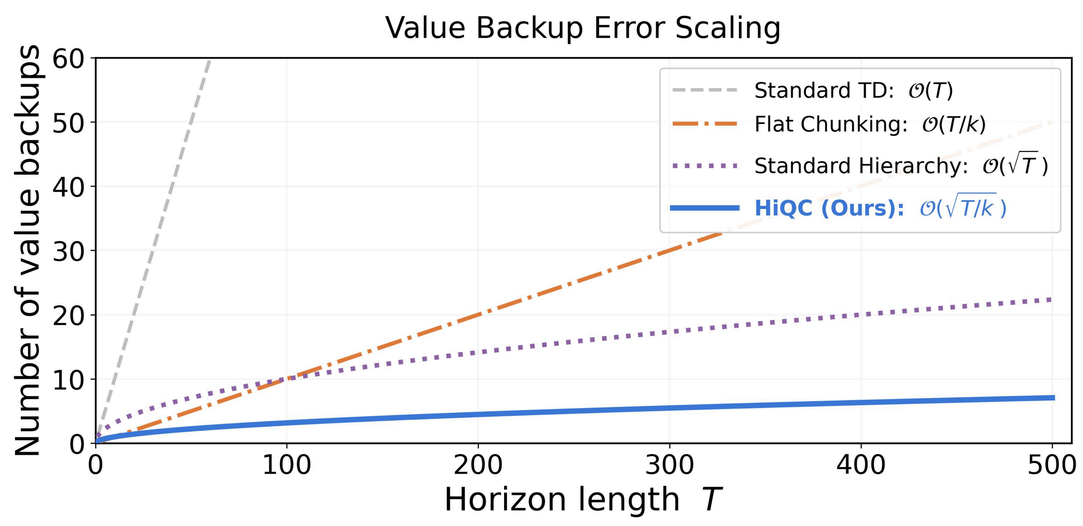
$$\mathcal{L}_{Q_L}(\theta) = \mathbb{E}\Big[\big(r(s,z) + \gamma^k V_L(s_{t+k}, z) - Q_L(s, \mathbf{a}, z)\big)^2\Big]$$

Scan Me!



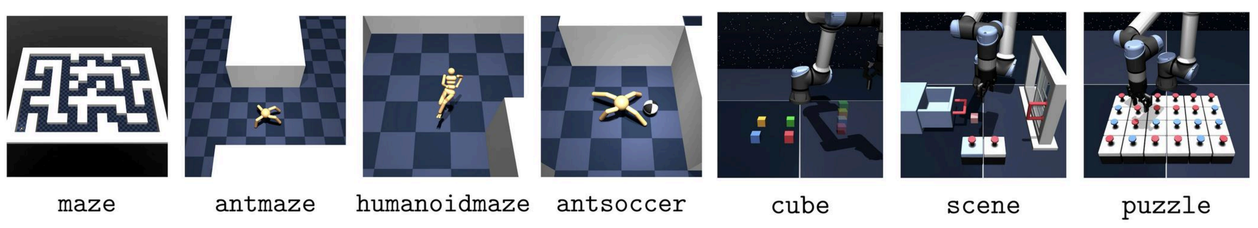
Theoretical Analysis

- Assuming bounded per-backup error, HiQC improves error scaling over task horizon T.

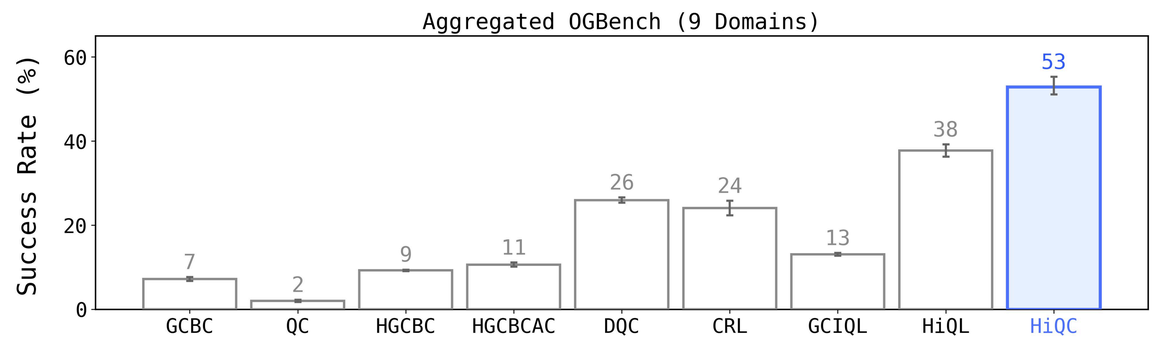


Results

- Evaluated on OGBench simulation environments.

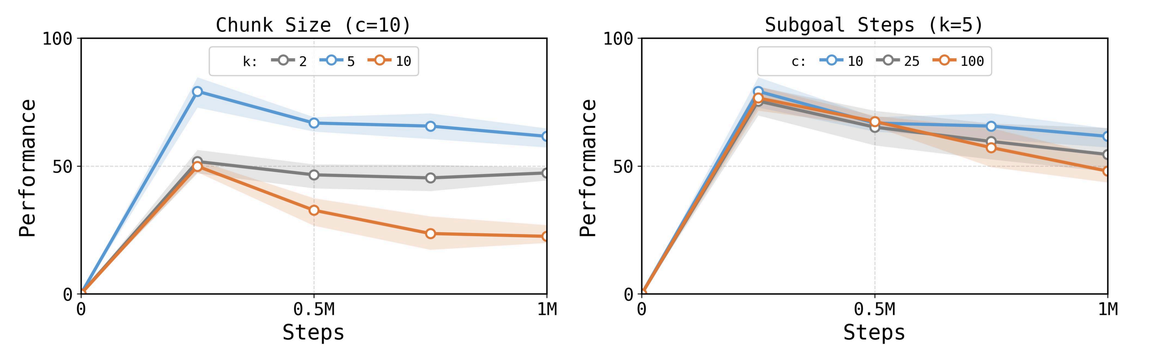


- Reported mean success rate at a 95% confidence interval over 4 seeds.



Task	BC	QC	HBC	HBCAC	DQC	CRL	GCIQL	HiQL	HiQC
antmaze-large	27 [24,30]	0 [0,0]	49 [48,51]	54 [52,56]	35 [33,38]	81 [78,84]	32 [25,37]	88 [86,89]	**93 [92,94]**
antmaze-giant	0 [0,0]	0 [0,0]	25 [22,27]	30 [27,33]	1 [0,1]	14 [11,18]	0 [0,1]	**68 [67,69]**	**68 [67,69]**
humanoid-large	1 [1,2]	0 [0,0]	4 [3,5]	6 [5,7]	1 [1,1]	17 [12,23]	1 [1,2]	34 [23,44]	**42 [39,44]**
humanoid-giant	0 [0,0]	0 [0,0]	1 [0,1]	0 [0,0]	0 [0,0]	4 [2,5]	0 [0,0]	10 [7,15]	**33 [29,39]**
pointmaze-large	29 [25,34]	15 [13,17]	0 [0,0]	0 [0,1]	62 [60,65]	36 [32,41]	29 [24,33]	47 [39,55]	**87 [81,92]**
pointmaze-giant	1 [0,1]	0 [0,0]	0 [0,0]	0 [0,0]	35 [29,40]	33 [27,39]	1 [0,3]	44 [35,53]	**80 [76,85]**
cube-triple	1 [1,2]	0 [0,0]	0 [0,0]	0 [0,0]	**8 [7,10]**	4 [3,4]	1 [1,2]	3 [2,4]	**9 [4,15]**
puzzle-4x6	0 [0,0]	0 [0,0]	0 [0,0]	0 [0,0]	**16 [14,17]**	5 [4,6]	5 [4,6]	4 [4,5]	1 [1,2]
scene-play	5 [4,5]	3 [2,3]	4 [4,5]	5 [5,6]	**76 [75,77]**	22 [21,23]	47 [44,49]	41 [38,44]	65 [61,68]

- Hyper parameter sensitivity analysis on scene-play:



Future Works & Limitations

- Extending this algorithm with a multimodal or natural language based state space.
- Using deeper levels of hierarchy.
- Using search based algorithm on the high level planner.
- Theory Limitations: Bounded error assumption may not strictly hold with deep neural networks.