LLM-Explorer: A Plug-in Reinforcement Learning Policy Exploration Enhancement Driven by Large Language Models

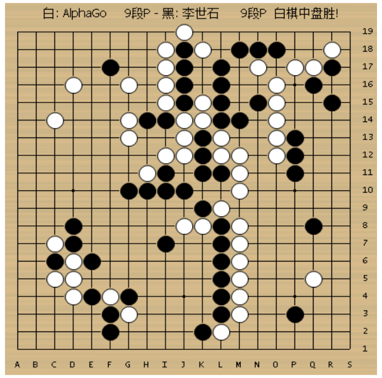
Qianyue Hao, Yiwen Song, Qingmin Liao, Jian Yuan, Yong Li
Department of Electronic Engineering, Tsinghua University

Speaker: Qianyue Hao

Backgrounds: RL before LLMs

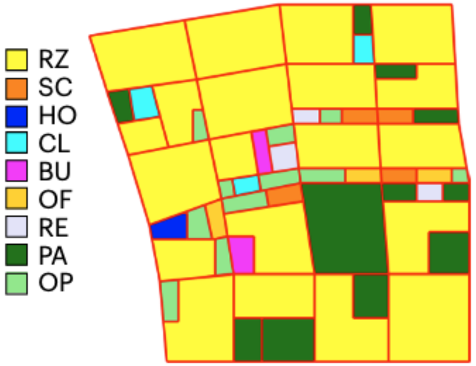
Wide applications across various tasks

Transportation



Go Game

Chip Design



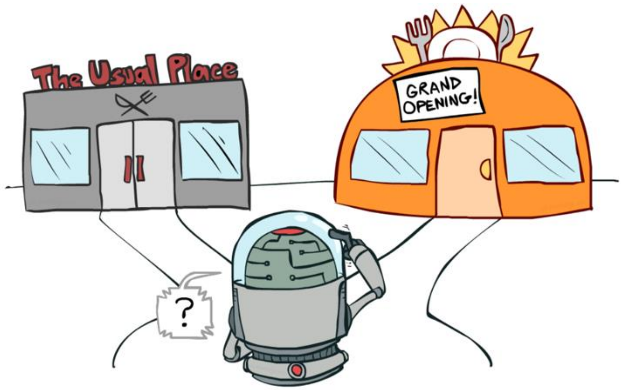
Urban Planning

(Super-) human level performance

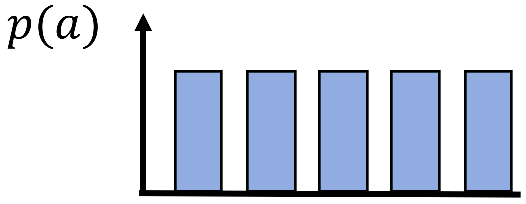
[1] Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of go without human knowledge. *Nature*, 2017.
[2] Mirhoseini A, Goldie A, et al. A graph placement methodology for fast chip design. *Nature*, 2021.
[3] Zheng Y, Lin Y, Zhao L, et al. Spatial planning of urban communities via deep reinforcement learning. *Nature Computational Science*, 2023.
[4] Zheng Y, Hao Q, et al. A Survey of Machine Learning for Urban Decision Making: Applications in Planning, Transportation, and Healthcare. *ACM Computing Surveys*, 2024.

Backgrounds: policy exploration in RL

Fixed stochastic processes

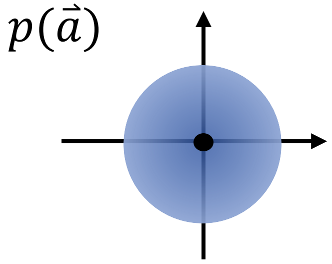


Uniform distribution



DQN and variants

Gaussian distribution



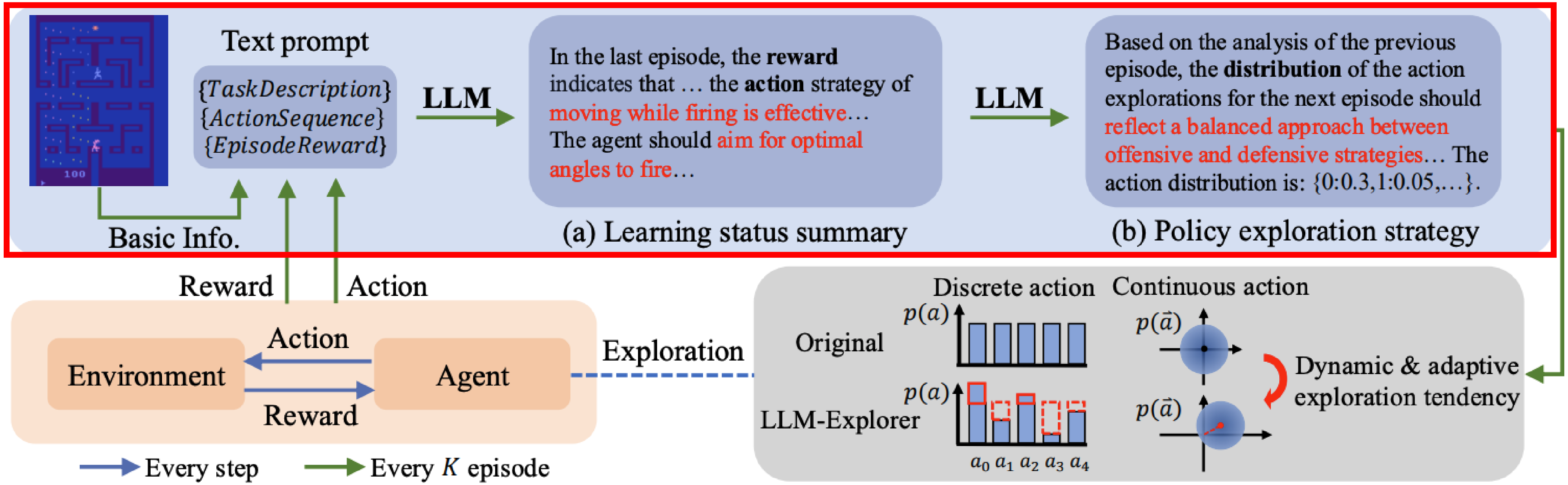
DDPG and variants

- **Lack flexibility**: preset stochastic processes applied uniformly across all kinds of tasks without any environment-specific design, neglecting the unique characteristics of different tasks.
- **Lack adaptability**: fail to flexibly adjust the policy exploration strategy based on the agent’s real-time learning status, potentially reducing the effectiveness of policy exploration.

[1] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature*, 2015.
[2] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning. *ICLR*, 2016.
[3] Fortunato M, Azar M G, Piot B, et al. Noisy networks for exploration. *ICLR*, 2018.

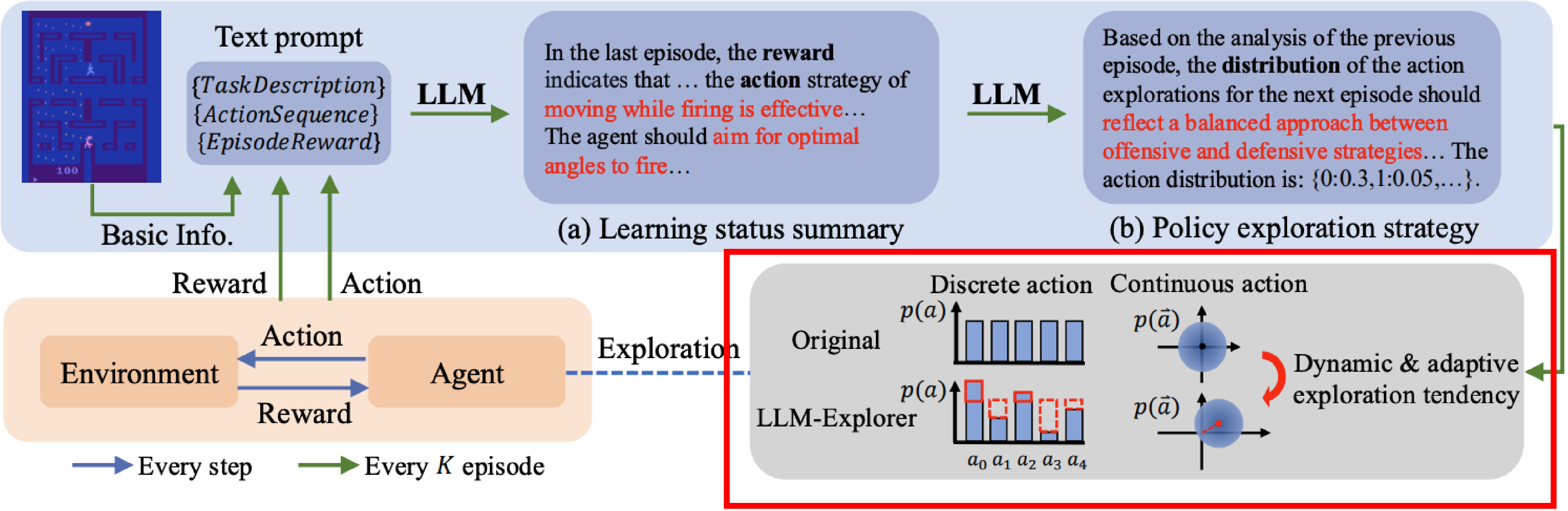
Method: LLM guides policy exploration

LLMs workflow to analyze the task feature and learning status

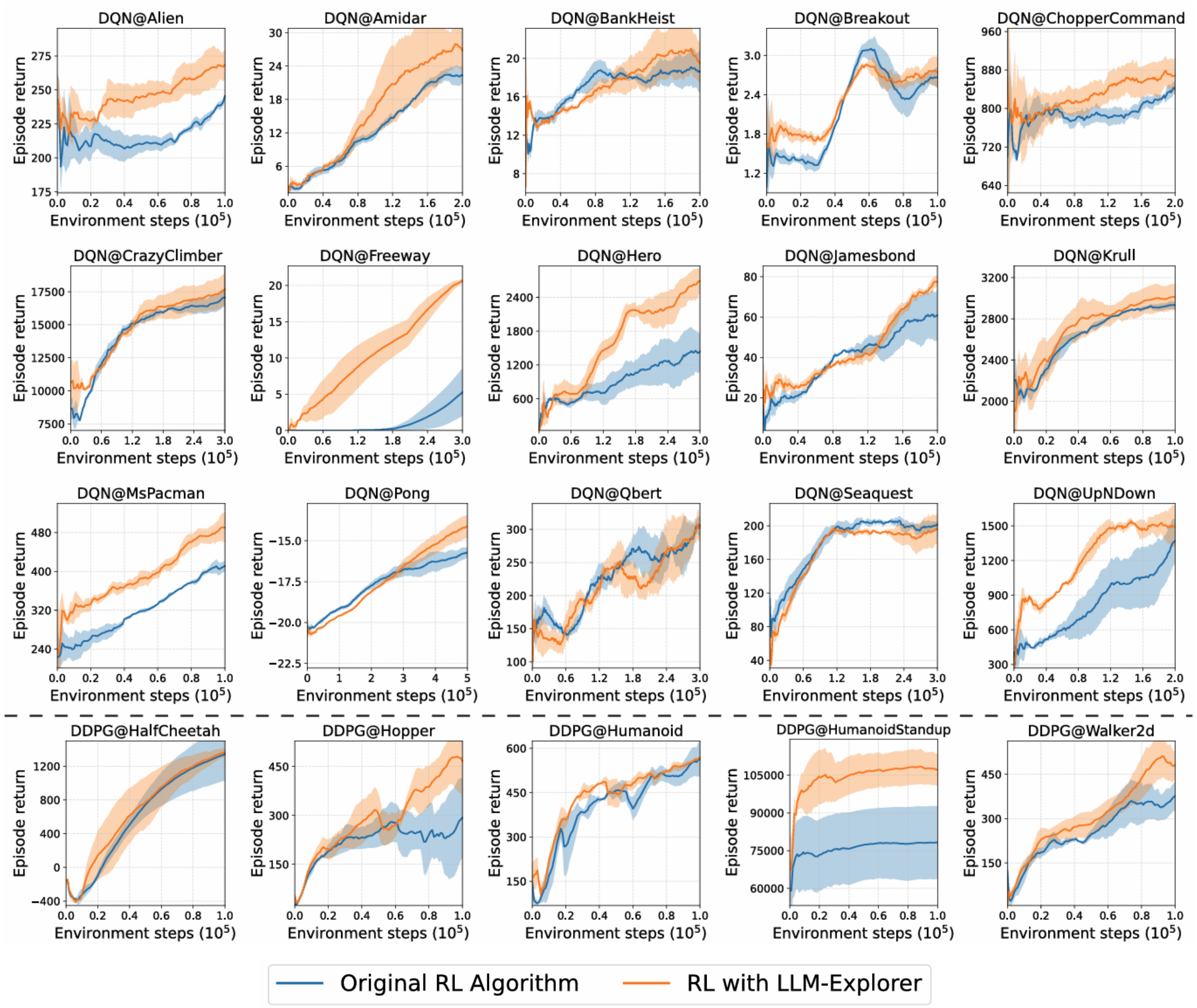


Method: LLM guides policy exploration

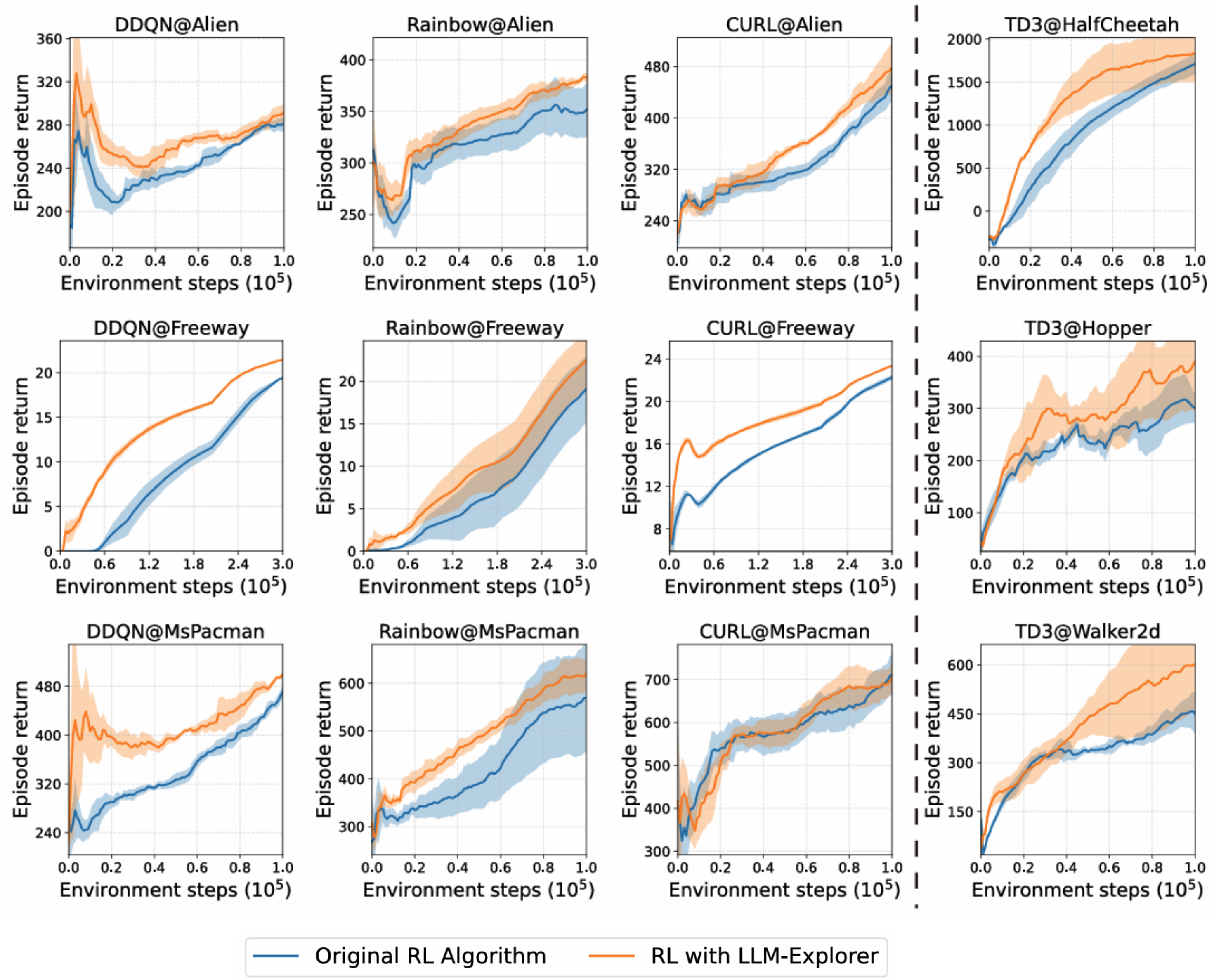
Plug-in design that is compatible with various existing RL algorithms



Performance: effective on various tasks



Performance: compatible with RL algorithms



Analyses: ablations on workflow design

Task	DQN	DQN+LLM-Explorer								
		Full design			w/o summarize & suggestion			w/o environment information		
		Score	Token in (k)	Token out (k)	Score	Token in (k)	Token out (k)	Score	Token in (k)	Token out (k)
Alien	0.26	**0.59**	248.73	179.59	0.51	111.07	112.54	0.38	186.41	165.90
Freeway	17.75	**69.71**	220.12	138.75	68.97	88.91	69.94	61.26	164.38	134.93
MsPacman	1.56	**2.75**	291.30	201.22	2.32	129.18	125.22	1.89	222.05	208.31

- The full design workflow is critical for achieving the best performance.
- Simplify the workflow can reduce computational consumption while still maintaining certain performance.