N18　脳情報科学

Hardware-Accelerated Reinforcement Learning

概要

We propose Pgx, a suite of reinforcement learning (RL) environments written in JAX and optimized for GPU/TPU accelerators. By leveraging JAX's auto-vectorization and parallelization over accelerators, Pgx can efficiently scale to thousands of simultaneous simulations over accelerators.

特徴

- Fast simulation: Pgx can efficiently scale to thousands of simultaneous simulations over accelerators and run fast on GPU/TPUs, similar to Brax [Freeman+21].

- Diverse set of games: Pgx offers over 20 environments. Pgx also offers miniature versions of game environments (e.g., miniature chess) to facilitate research cycles

- Baseline models: Evaluating agents in multi-agent games is relative, requiring baseline opponents for evaluation. Since it is not always easy to have appropriate baselines available, Pgx provides its own baseline models
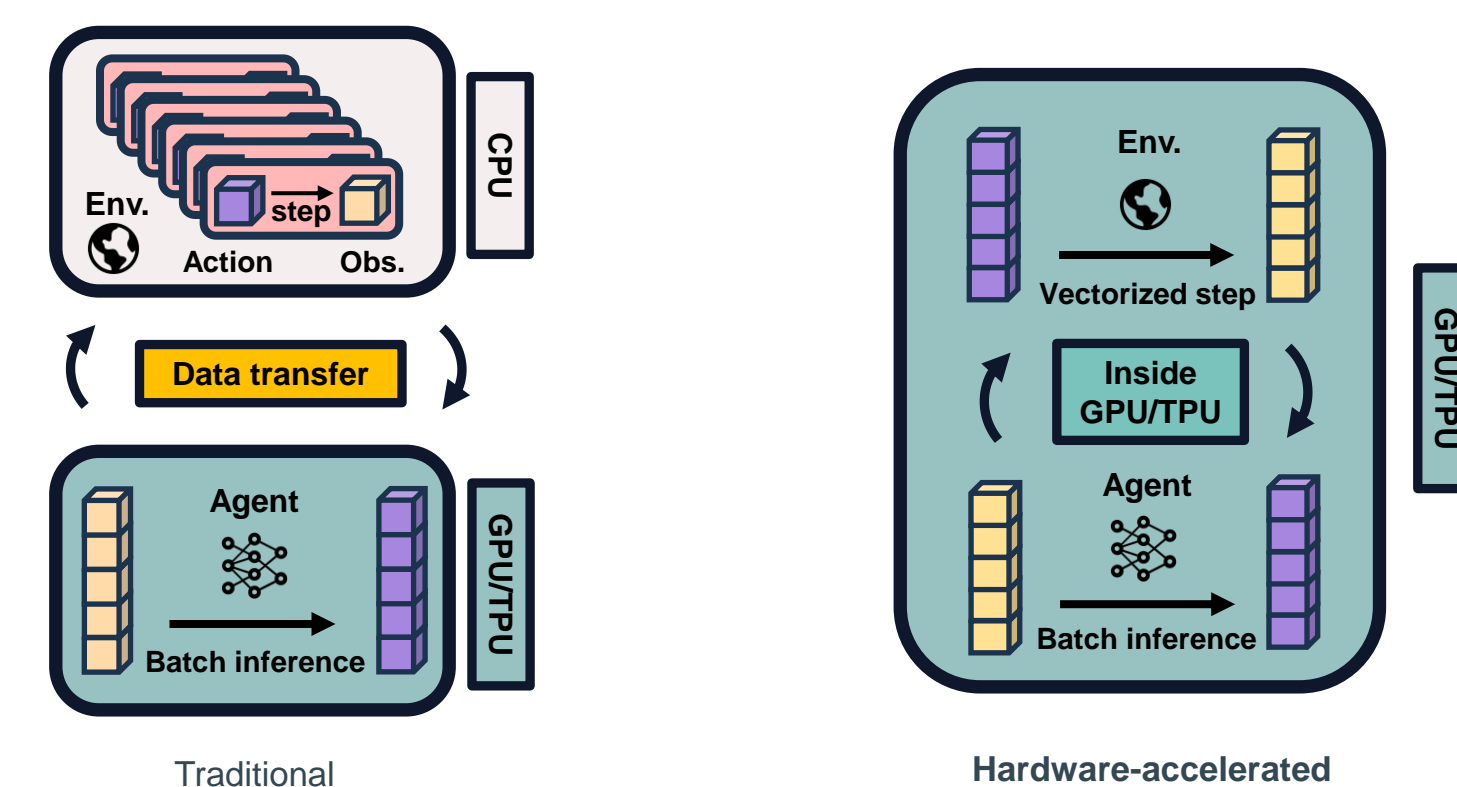
今後の展開

- Future work includes the development of RL and planning algorithms suitable for vectorized environments operating on GPU/TPUs.
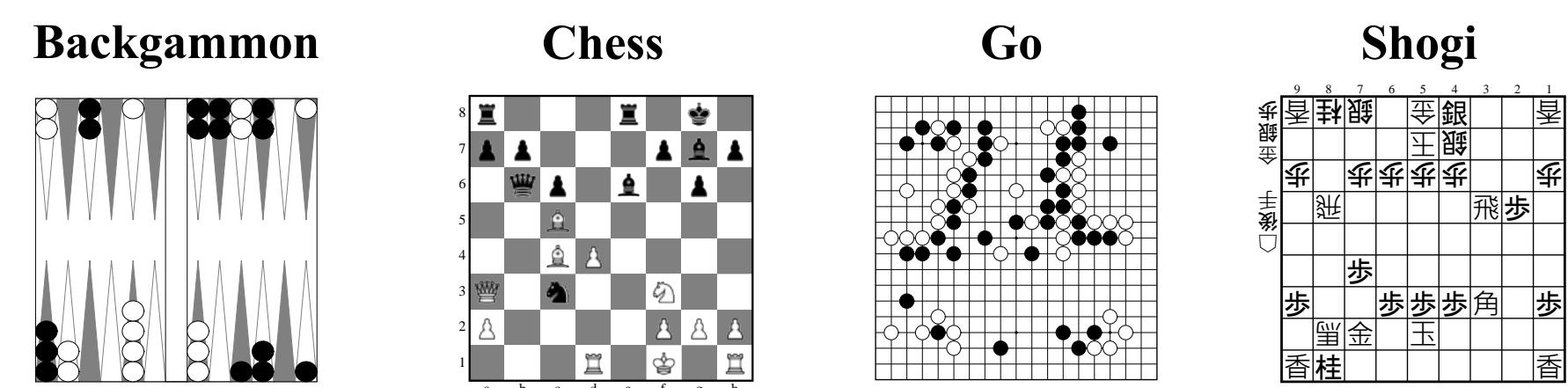
テーマ「万博、そしてその先へ 〜科学技術が描く未来〜」との関連

- A fast RL algorithm has the potential to explore and discover insights that humans cannot find.
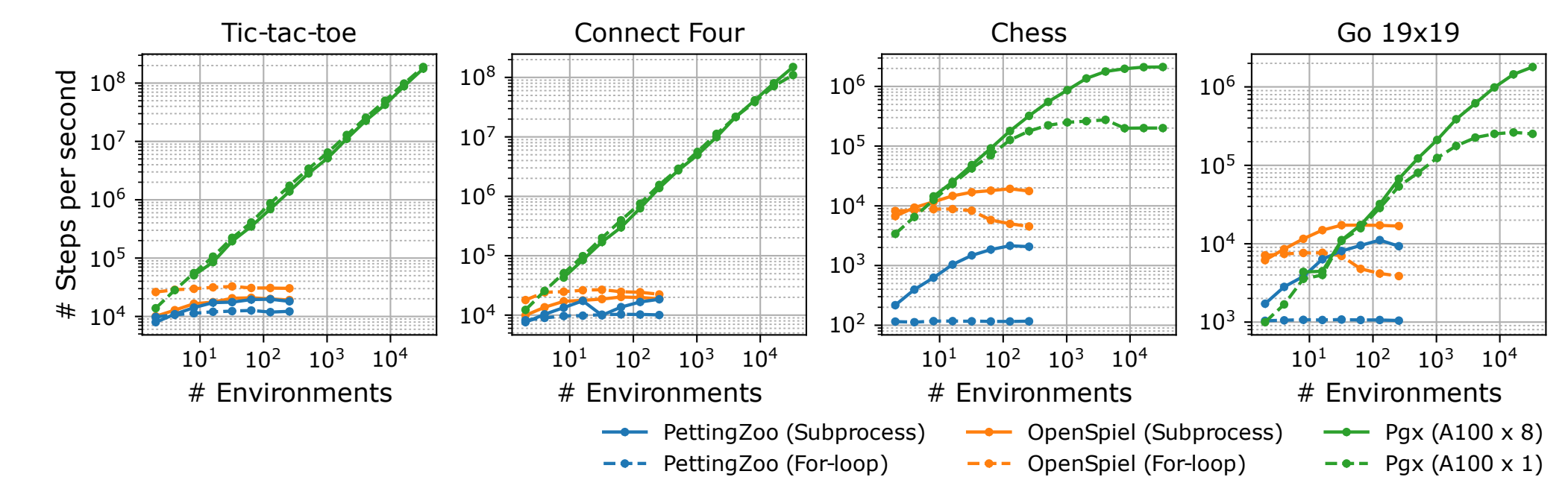
Hardware-accelerated RL environments

Available games in Pgx

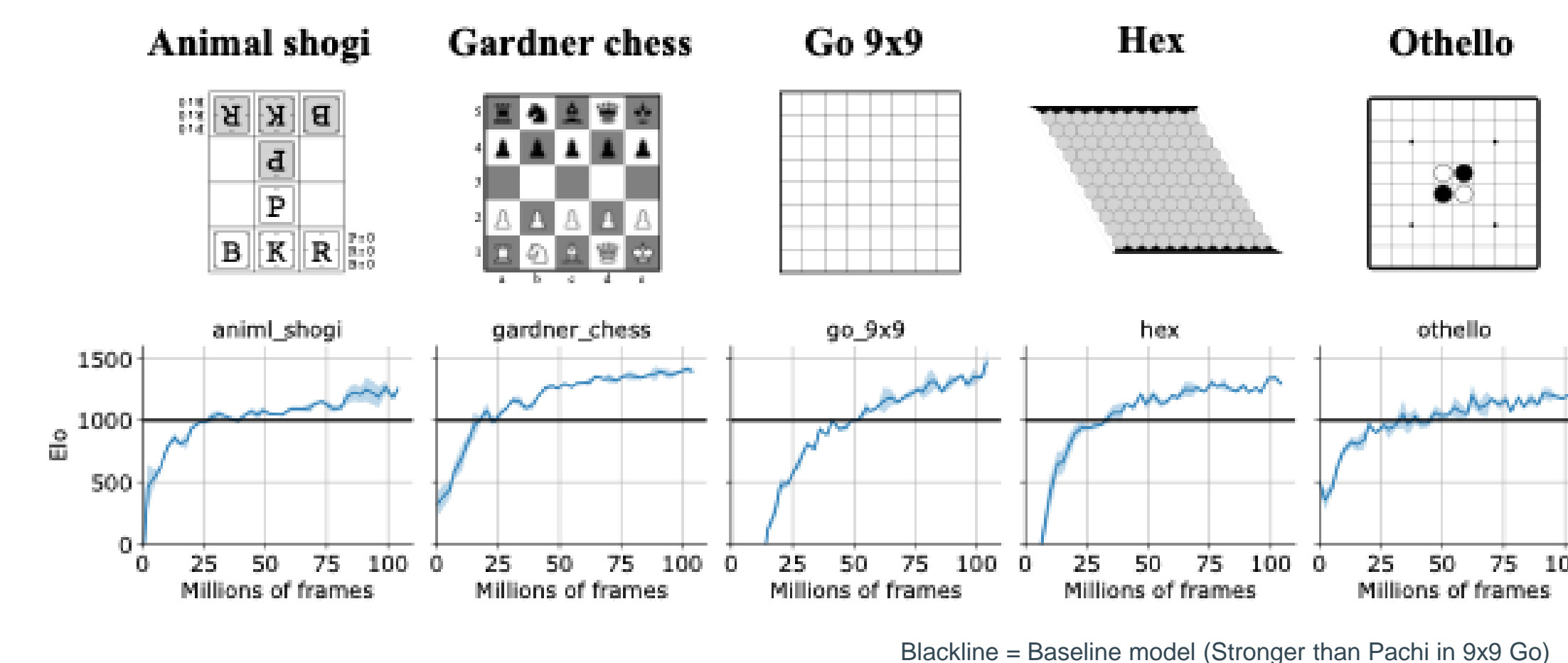+20 game environments including backgammon, chess, shogi and Go

Backgammon　Chess　Go　Shogi

Env Name	# Players	Obs. shape	# Actions	Tag
2048	1	$4 \times 4 \times 31$	4	perfect info. (w/ chance)
Animal shogi	2	$4 \times 3 \times 194$	132	perfect info.
Backgammon	2	34	156	perfect info. (w/ chance)
Bridge bidding	4	480	38	imperfect info.
Chess	2	$8 \times 8 \times 119$	4672	perfect info.
Connect Four	2	$6 \times 7 \times 2$	7	perfect info.
Gardner chess	2	$5 \times 5 \times 115$	1225	perfect info.
Go 9x9	2	$9 \times 9 \times 17$	82	perfect info.
Go 19x19	2	$19 \times 19 \times 17$	362	perfect info.
Hex	2	$11 \times 11 \times 4$	122	perfect info.
Kuhn poker	2	7	4	imperfect info.
Leduc hold'em	2	34	3	imperfect info.
MinAtar Asterix	1	$10 \times 10 \times 4$	5	Atari-like
MinAtar Breakout	1	$10 \times 10 \times 4$	3	Atari-like
MinAtar Freeway	1	$10 \times 10 \times 7$	3	Atari-like
MinAtar Seaquest	1	$10 \times 10 \times 10$	6	Atari-like
MinAtar Space Invaders	1	$10 \times 10 \times 6$	4	Atari-like
Othello	2	$8 \times 8 \times 2$	65	perfect info.
Shogi	2	$9 \times 9 \times 119$	2187	perfect info.
Sparrow mahjong	3	11×15	11	imperfect info.
Tic-tac-toe	2	$3 \times 3 \times 2$	9	perfect info.

Simulation throughput

End-to-end Gumbel AlphaZero [Danihelka+22] training

Animal shogi　Gardner chess　Go 9x9　Hex　Othello

Blackline = Baseline model (Stronger than Pachi in 9x9 Go)

Multi-GPU training
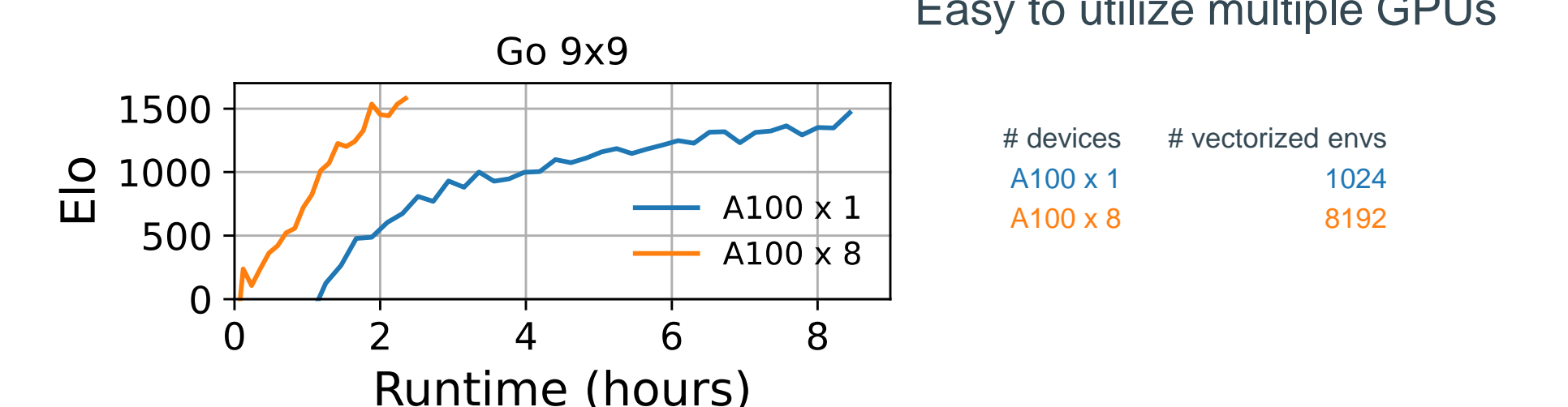
Easy to utilize multiple GPUs

# devices	# vectorized envs
A100 x 1	1024
A100 x 8	8192

連絡先: 脳情報解析研究所　担当　小山田創哲　E-Mail:koyamada@atr.jp