



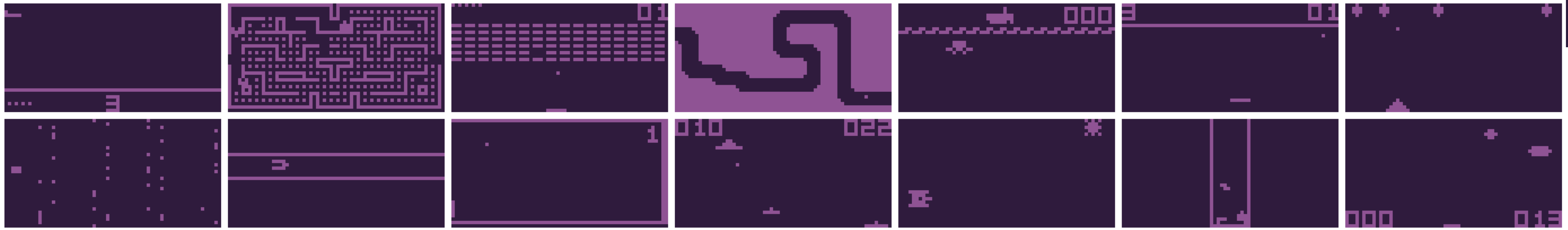
OCTAX: Accelerated CHIP-8 Arcade Environments for Reinforcement Learning in JAX

Waris Radji¹ Thomas Michel¹ Hector Piteau²

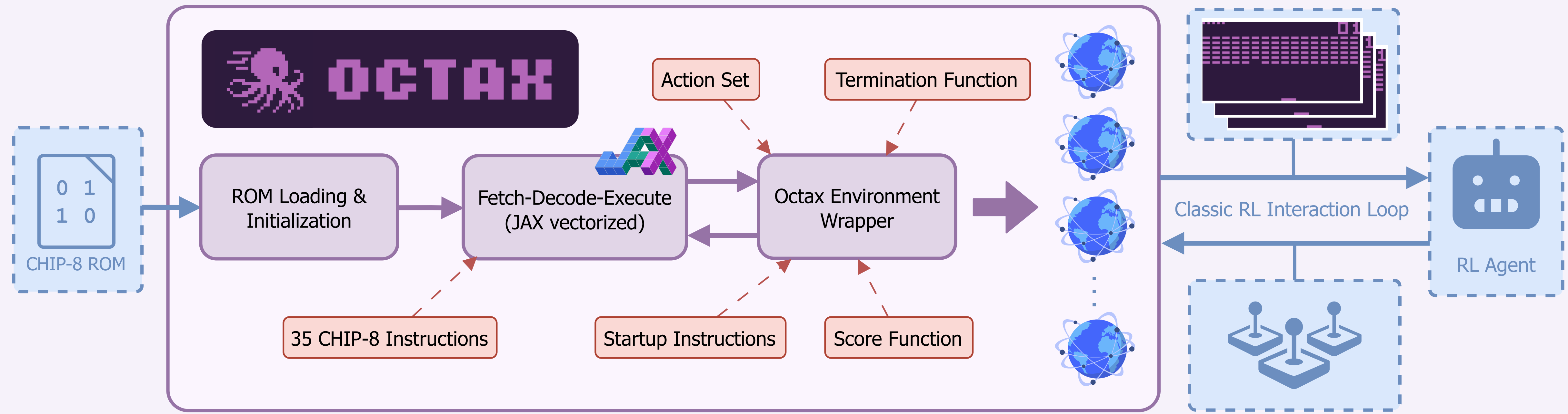
¹Univ. Lille, Inria, CNRS, Centrale Lille, UMR 9189-CRISTAL, France ²Independent Researcher



Inria



How Does OCTAX Work?



Game ROMs load into 4 KB memory and execute via a fetch–decode–execute loop using JAX’s `lax.switch` for GPU-compatible dispatch. The vectorised step runs thousands of instances in parallel, exposing a Gymnax-compatible interface. Observations are $(4 \times 64 \times 32)$ stacked boolean frames; actions map to game-specific CHIP-8 key subsets.

Abstract & Contributions

Deep Reinforcement Learning research demands extensive experimentation, yet computational constraints limit scale. **OCTAX** delivers a **JAX-native, end-to-end GPU** suite of classic arcade environments via CHIP-8 emulation, achieving **350K steps/s** (1.4M frames/s on RTX 3090), orders of magnitude faster than CPU emulators like EnvPool, and making large-scale RL experimentation accessible on a single consumer-grade GPU.

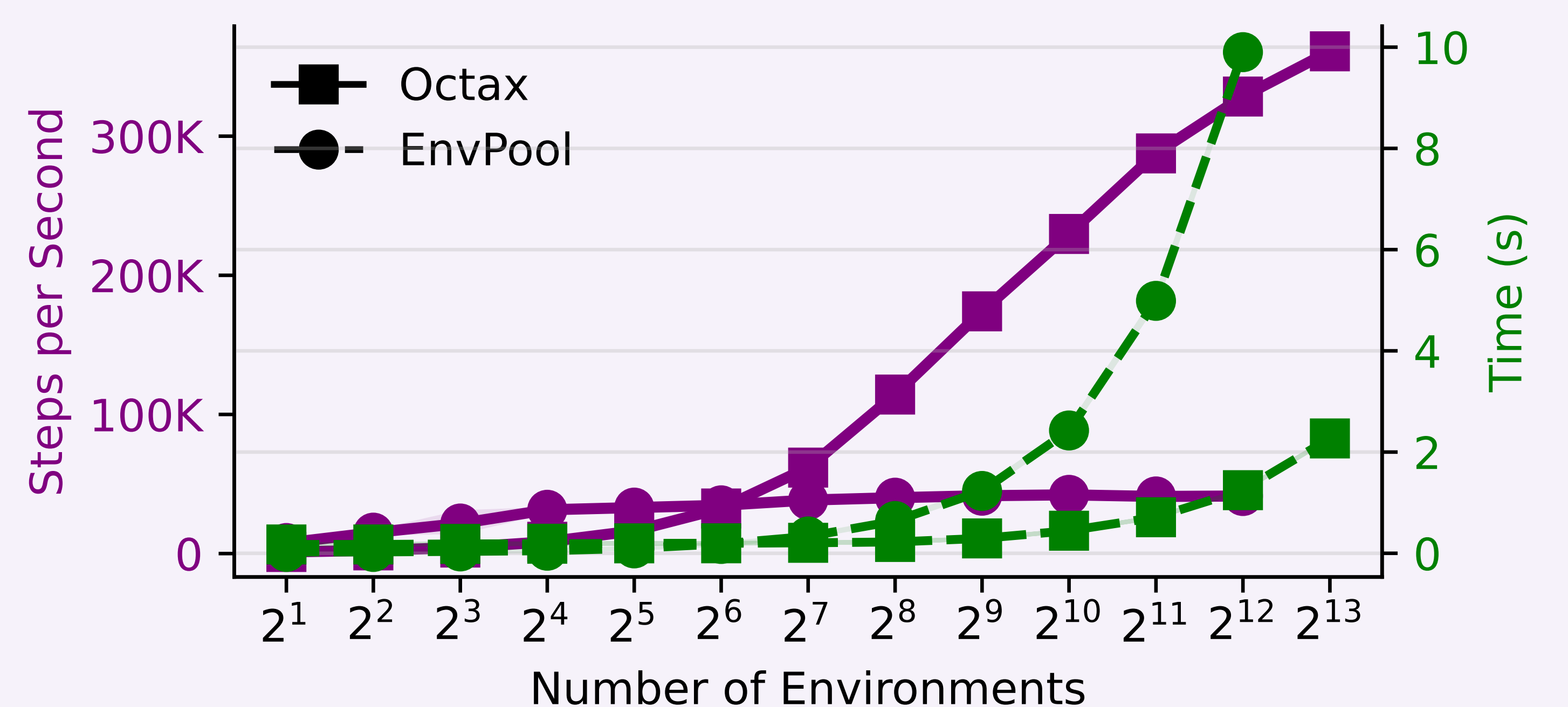
GPU-Accelerated Emulator. Vectorised JAX CHIP-8 emulator running thousands of parallel environments with perfect fidelity to original game mechanics.

21 Game Environments. Puzzle, action, strategy, exploration and shooter genres with Gymnax-compatible (and Gymnasium, through wrappers) APIs.

RL Benchmarks. We evaluate PPO & PQN across 16 games with 12 seeds in a few hours, revealing that **OCTAX** induces diverse and interesting training dynamics.

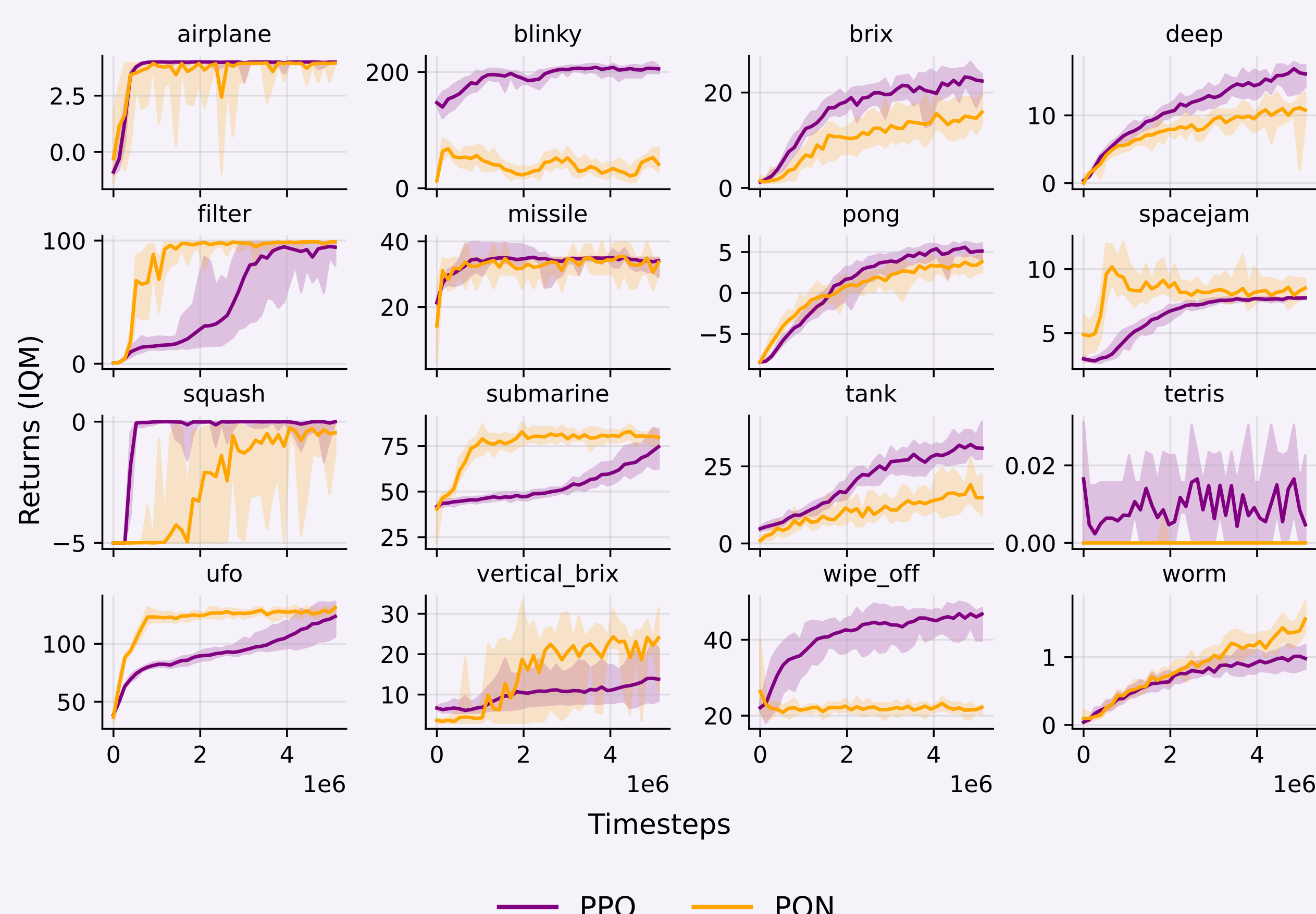
LLM-Assisted Generation. LLMs can write CHIP-8 assembly from natural language, auto-validated through compilation and RL training.

How Fast is OCTAX?



Near-linear scaling to **350 000 steps/s** (1.4 M frames/s) at 8 192 parallel environments on a consumer RTX 3090. EnvPool (ALE, all CPU cores) plateaus at $\approx 25\,000$ steps/s, a **14 $\times$ improvement** at high parallelisation. GPU memory usage: ≈ 2 MB per environment.

How Do RL Agents Learn?



Interquartile mean returns (10th–90th percentile) over 5 M timesteps, 12 seeds per game.

Can LLMs Design New Environments?

CHIP-8 assembly is compact and simple enough for LLMs to write directly.



Target Shooter, generated by Claude Opus 4.1 from a single natural-language description

Describe game → LLM writes ROM → Compile & validate → RL training