國立清華大學
NATIONAL TSING HUA UNIVERSITY

Elsa Lab
Since 2015

# Toward Synergism in Macro Action Ensembles

Yu-Ming Chen, Kuan-Yu Chang, Chien Liu, Tsu-Ching Hsiao, Zhang-Wei Hong and Chun-Yi Lee

*Elsa Lab, Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan*

NVIDIA
MEDIATEK
MOST 科技部
Ministry of Science and Technology

## Abstract

Macro actions have been demonstrated to be beneficial for the learning processes of an agent. A variety of techniques have been developed to construct more effective macro actions. However, they usually fail to provide an approach for combining macro actions to form a synergistic macro action ensemble. A synergistic macro action ensemble performs better than individual macro actions within it. Motivated by the recent advances of neural architecture search, we formulate the construction of a synergistic macro action ensemble as a sequential decision problem and evaluate the ensemble in a task. The formulation of sequential decision problem enables coherency in the macro actions to be considered. Also, our evaluation procedure takes synergism into account since the synergism among the macro action ensemble exhibits when jointly used by an agent. The experimental results show that our framework is able to discover synergistic macro action ensembles. We further perform experiments to validate the synergism property among the macro action ensemble.

## Proposed Methodology

Our proposed parallel asynchronous framework is composed of two phases: a construction phase and an evaluation phase.

### Construction phase

- The controller is an RL agent deciding a primitive action at each construction step based on the previously decided action sequence.
- The partial transition record is buffered in an evaluation queue waiting for the workers to evaluate the reward.
- The controller updates its policy in an off-policy manner using the data sampled from the replay memory.
- All primitive actions selected in an episode could be segmented into multiple macros that together form a macro ensemble.

### Evaluation phase

- A worker pool is a set of asynchronous parallel workers implemented as RL agents.
- Each worker in worker pool pop the partial transition record in the evaluation queue and evaluates the corresponding reward in its own copy of the target environment.
- The evaluated results are stored in a replay memory.



## Experimental Results
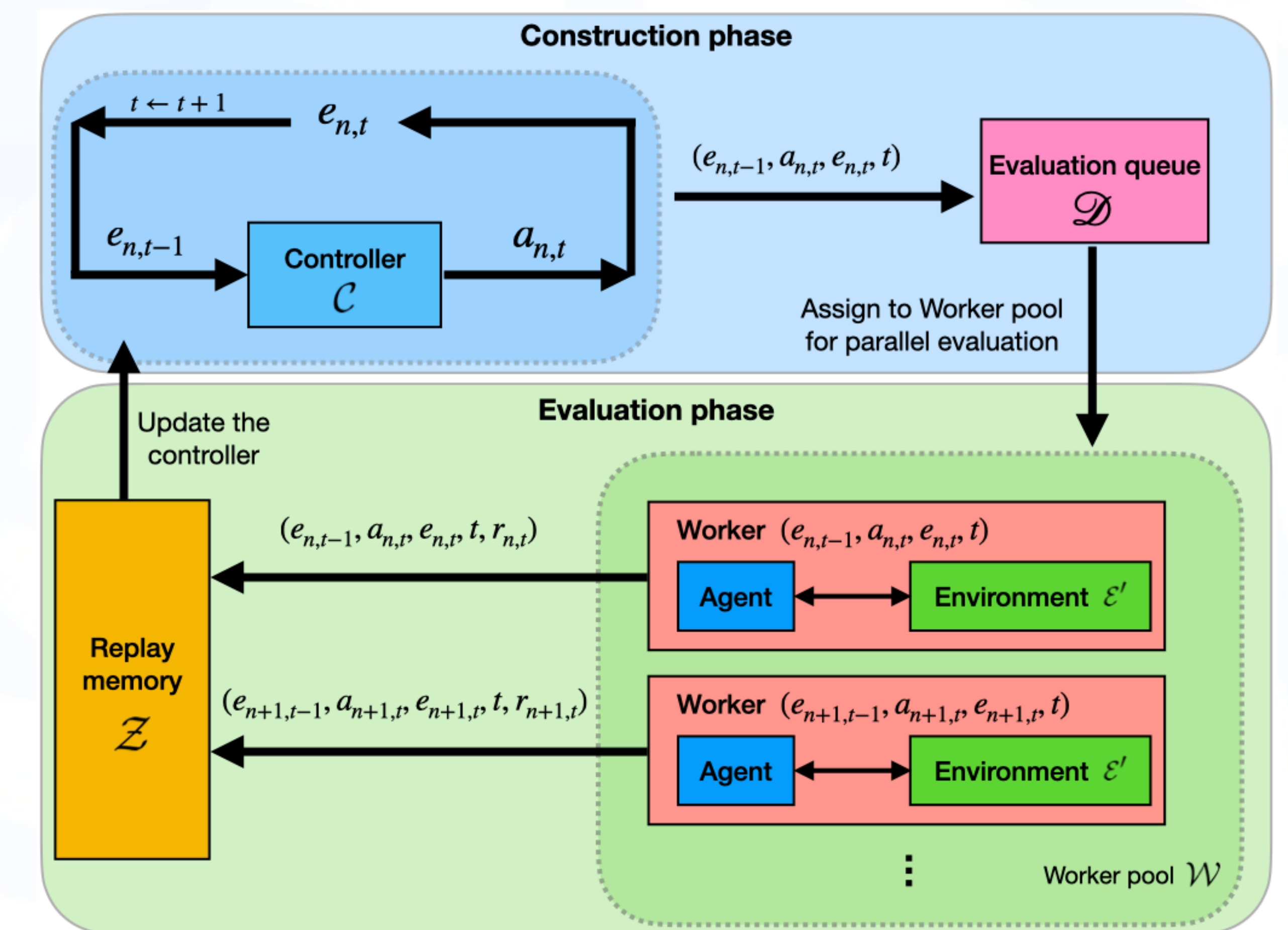


Seaquest: (UP, UP, UP) (UP, LEFT, RIGHT) (FIRE, FIRE, FIRE) (primitive actions)
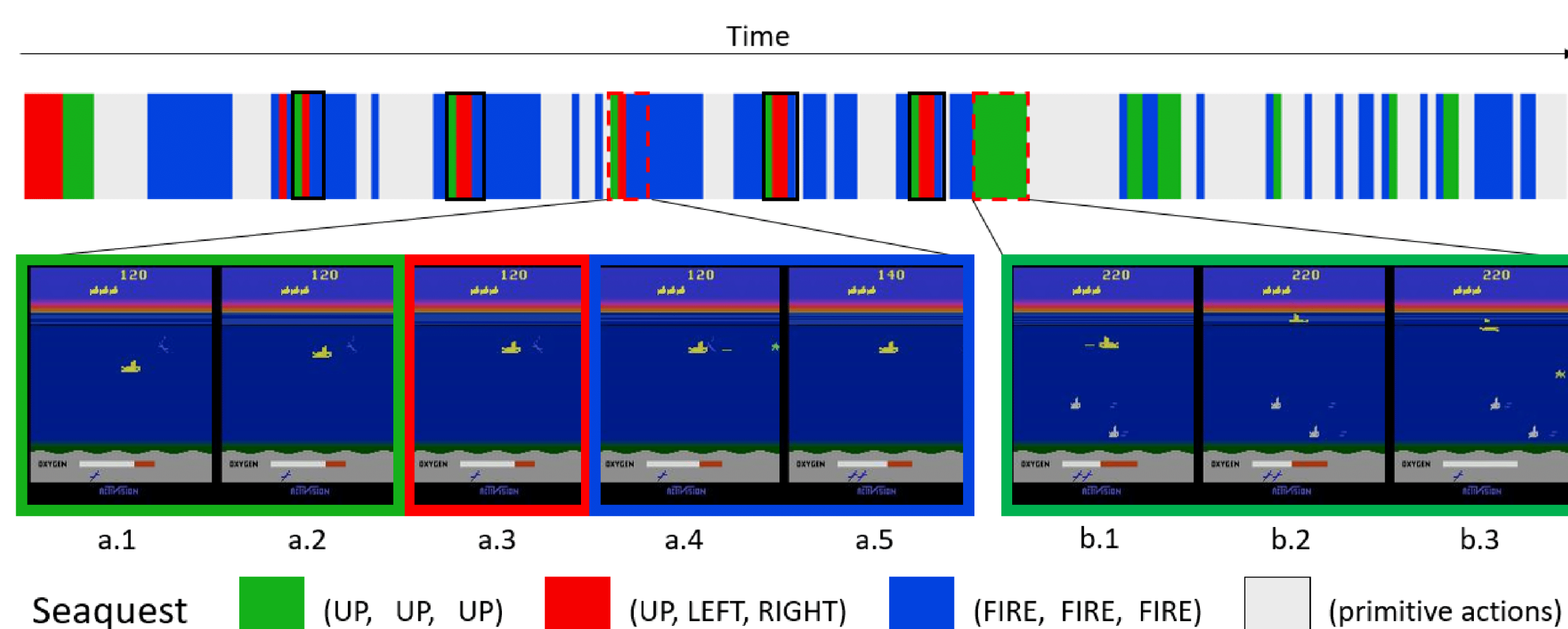
- The white segments correspond to the uses of primitive actions and the other colors (i.e., red, green and blue) represents three macros in the constructed macro action ensemble.
- The macro actions and primitive actions are able to be jointly and interleavedly used by a RL agent, which demonstrates the synergism property empirically.
- The primitive actions (depicted as white segments) and the macro actions are interleavedly used in the trajectory which further justifies that the macro actions are compatible to the primitive action space.
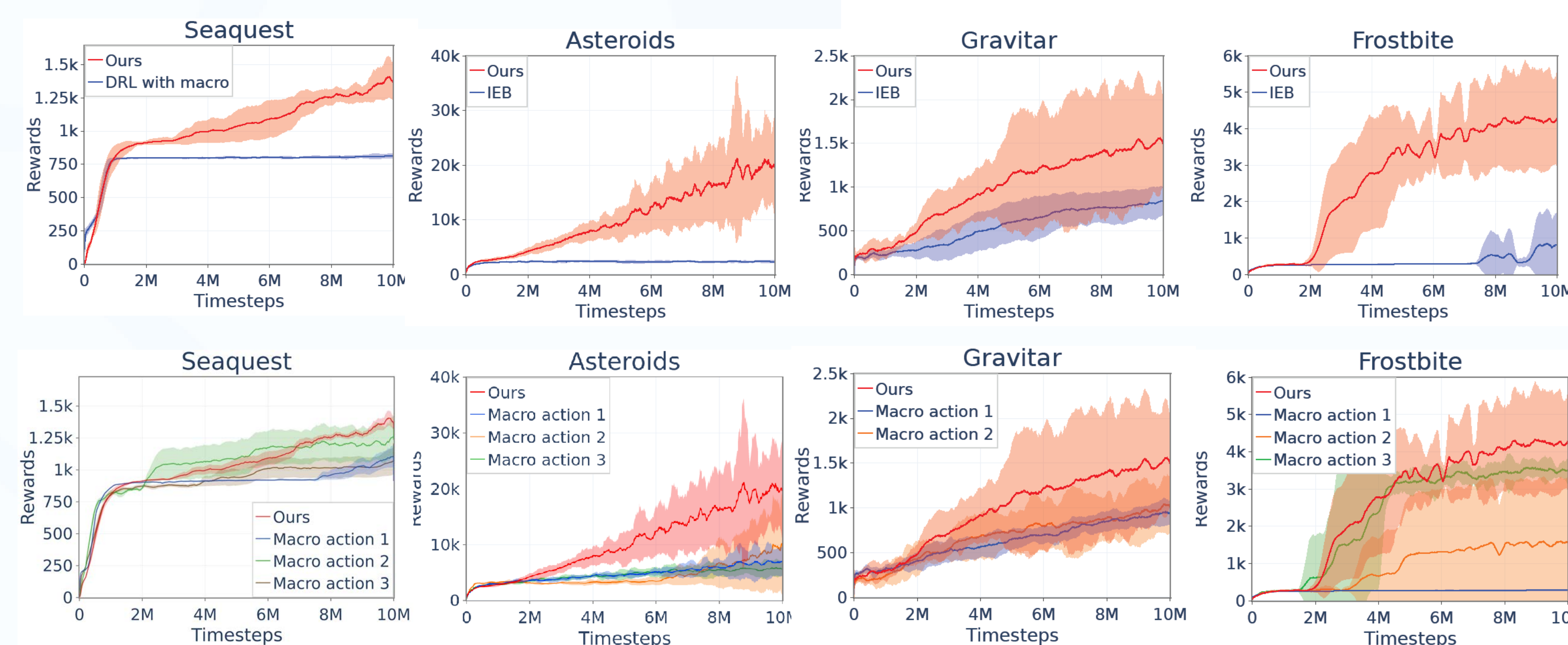


- The RL agents are trained from scratch for 10M timesteps based on the constructed macro ensemble.
- The performances of the macro ensembles constructed our methodology outperform those constructed by IEB.
- Our method can lead to higher performance than that corresponding to individual macros in the ensemble, showing the existence of the synergism property of the macro ensembles built by our method.

## Conclusion

In this paper, we presented a methodology for constructing macro action ensembles based on the MDP formulation. We proposed a parallel framework with an RL-based controller to generate candidate macro ensembles and evaluate them asynchronously. We evaluated the proposed methodology in a number of Atari 2600 environments against IEB. We demonstrated that our method is superior to IEB. We further provided analysis and verified the existence of the synergism property among the macro actions contained in the constructed macro ensemble. Moreover, our experimental results validated that the macro ensembles discovered by our method are complementary to the primitive action space, and outperformed this baseline in terms of episode rewards presented in appendix. The results show that the synergism property is consequential to the research field of macro action ensemble.

Questions?
cylee@cs.nthu.edu.tw