

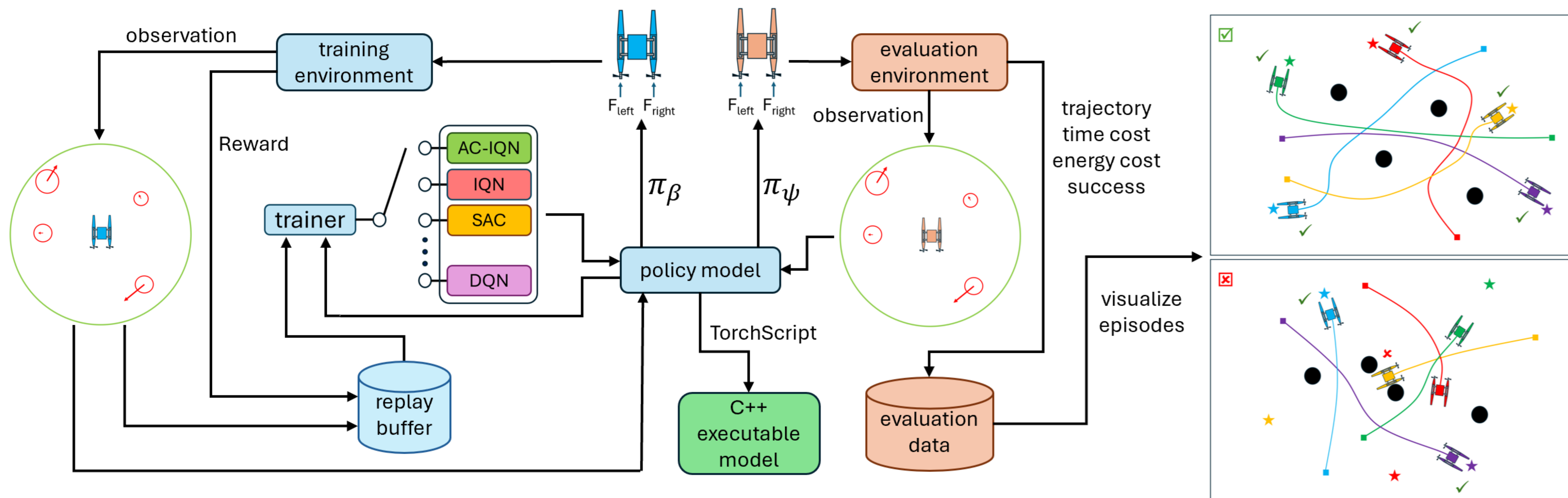


BENCHMARKING AND TRAINING SIMULATION FRAMEWORKS FOR USV NAVIGATION IN CONGESTED AND DISTURBANCE-FILLED ENVIRONMENTS

Xi Lin Brendan Englot
Stevens Institute of Technology



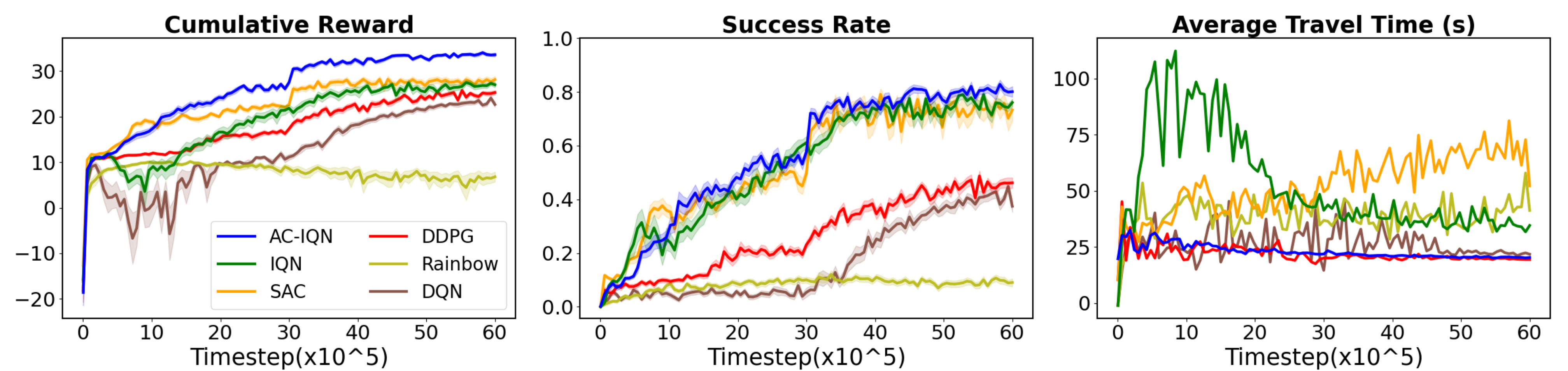
TRAINING FRAMEWORK



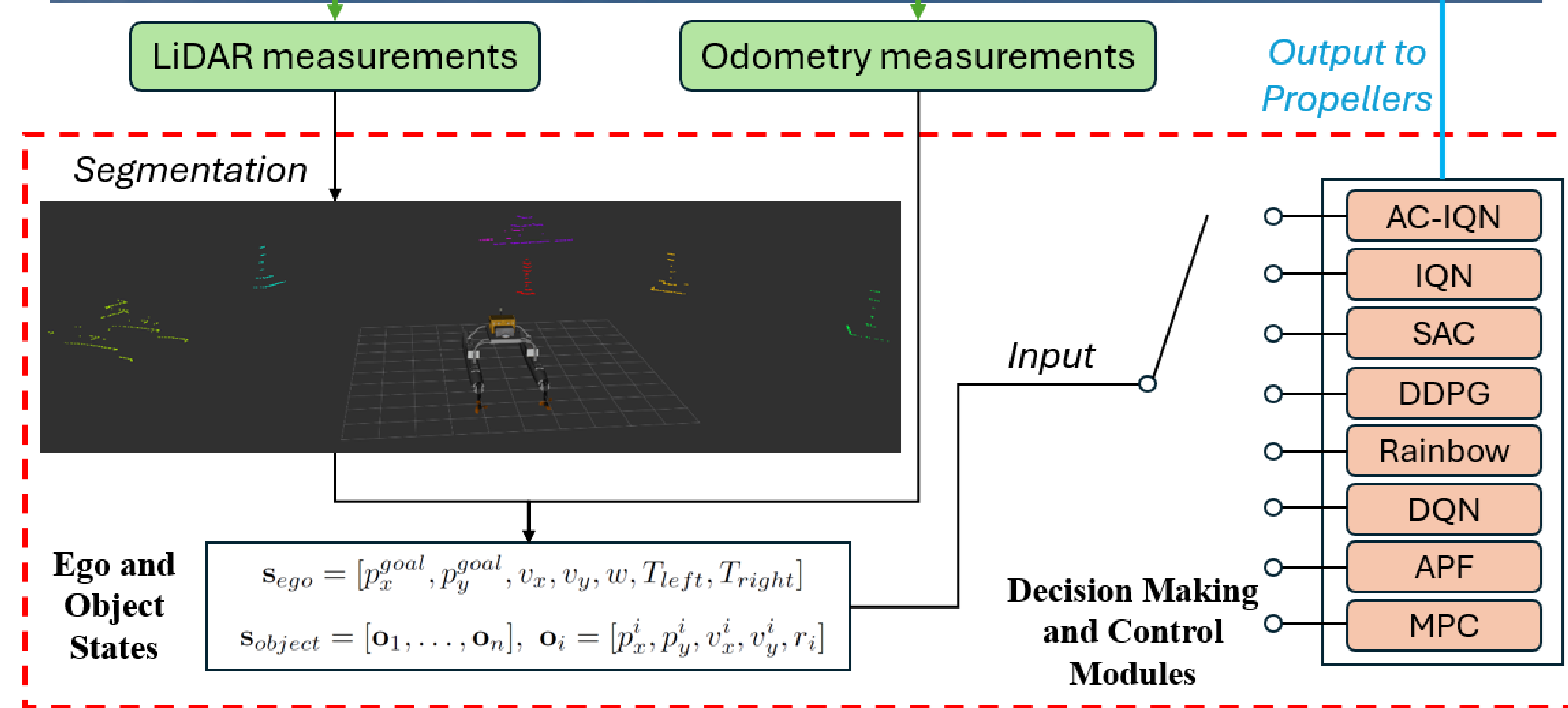
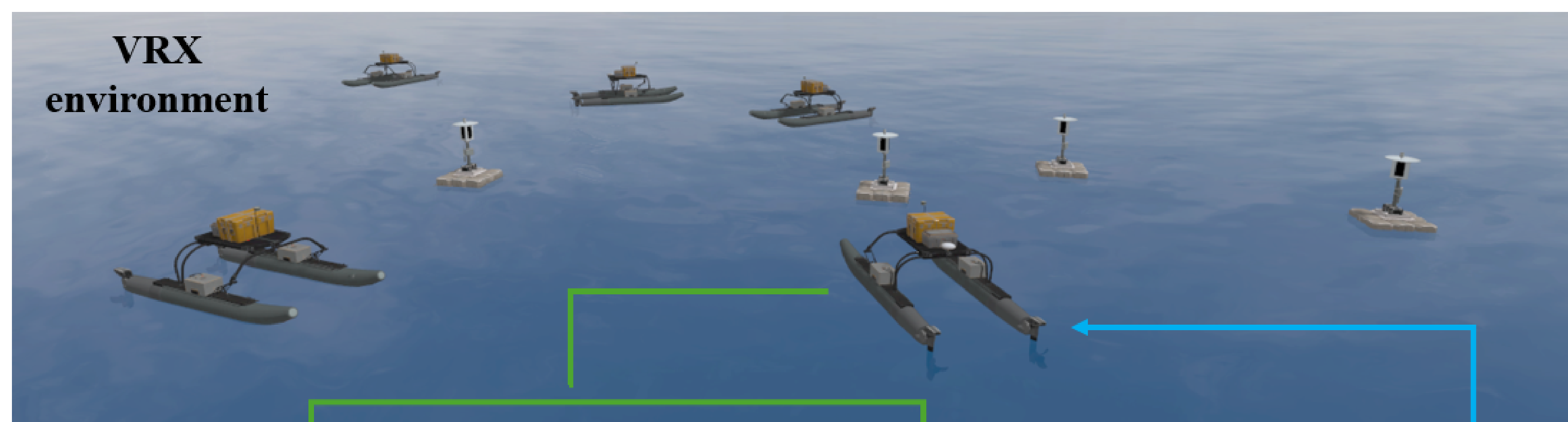
π_β is the behavior policy, for which we use ϵ -greedy to balance exploration and exploitation, and π_ψ is the greedy execution policy.

Table 1: Curriculum training process.

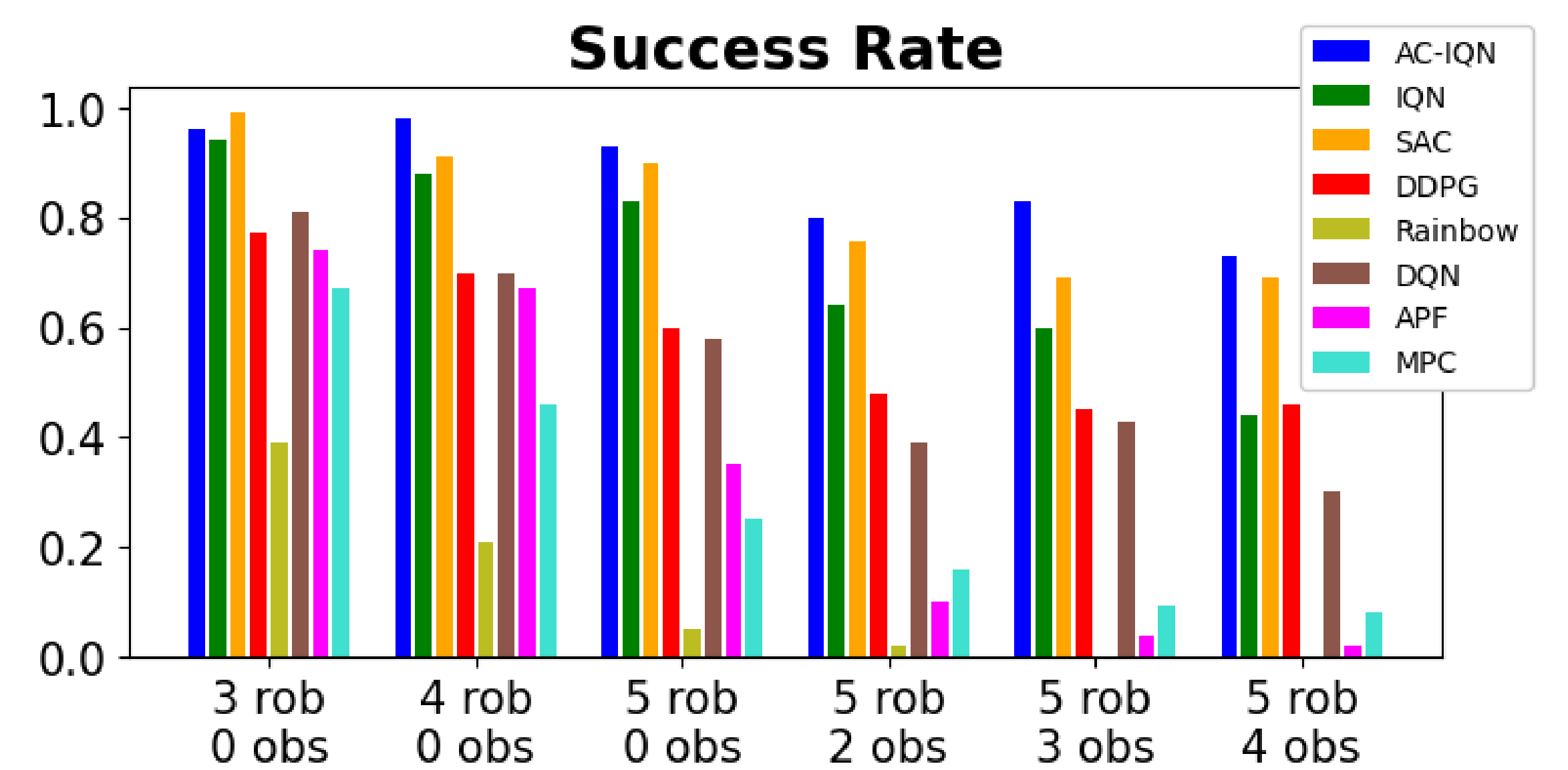
Timesteps (million)	1st	2nd	3rd	4th	5th	6th
Number of robots	3	4	5	5	5	5
Number of buoys	0	0	0	2	3	4
Min distance between start and goal (m)	30	35	40	40	40	40



BENCHMARKING FRAMEWORK



The proposed navigation system



Avg. travel time (s)	3 rob 0 obs	4 rob 0 obs	5 rob 0 obs	5 rob 2 obs	5 rob 3 obs	5 rob 4 obs
AC-IQN	17.22	17.85	20.22	21.06	21.97	22.81
IQN	29.19	30.06	36.77	36.65	41.21	39.64
SAC	24.09	24.75	29.96	29.50	31.54	32.35
DDPG	18.22	18.24	20.51	20.89	21.34	21.58
Rainbow	36.46	36.56	43.33	47.32	—	—
DQN	19.39	23.38	21.91	22.24	23.40	23.25
APF	22.37	23.07	25.76	23.82	24.62	24.53
MPC	17.96	19.10	20.71	28.80	33.16	33.61

TABLE II: Travel time data from VRX experiments. Rainbow based system fails completely in sets 5 rob 3 obs and 5 rob 4 obs.

PERFORMANCE OF ACTOR-CRITIC IMPLICIT QUANTILE NETWORKS (AC-IQN) BASED SYSTEM



This work was supported by the Office of Naval Research, grants N00014-20-1-2570 and N00014-24-1-2522.

