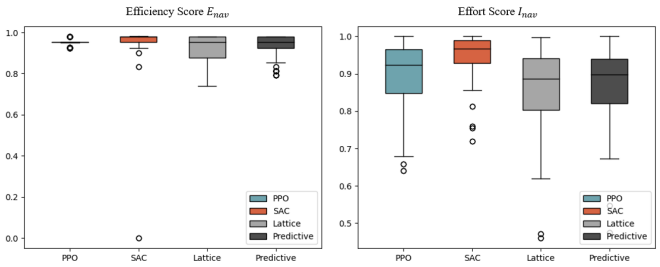

Fig. 1: Evaluation results for PPO, SAC, lattice-planning policy (Lattice), and predictive-planning policy (Predictive) in *Ship-Ice*.

APPENDIX I
GTSP BASELINE FOR *Area-Clearing*

We provide details about the *Area-Clearing* baseline policy that computes a robot path by solving a generalized traveling salesman problem (GTSP). This GTSP policy is inspired by an approach for robot coverage path planning [1]. For each box in the environment, we consider the path to clear the box across each edge of the clearance area boundary. For example, in the case of a rectangular clearance boundary, we consider 4 paths to clear each box across each of the 4 edges. We will refer to these as the box's *clearance paths*. We construct an auxiliary graph $G = (V, E)$ where each box constitutes a set S and contains vertices $v \in V$ that each correspond to a clearance path considered for the specific box. The robot's start position is encoded as a separate vertex. The edge cost between two vertices is the shortest path length to reach the start of one clearance path from the end of another (or the robot's start position). Solving the GTSP on this graph gives us a robot path that aims to clear each box individually along a series of clearance paths. The path is computed once and executed by the robot until it either (i) clears all boxes in the environment or (ii) finishes the full path. Note that this method does not necessarily clear all boxes and does not minimize the robot's effort as it ignores object-to-object interactions.

APPENDIX II
EVALUATIONS RESULTS

Here we present the evaluation results in simulations for *Ship-Ice* and *Area-Clearing*. In each environment, we evaluated all baseline policies for 200 episodes in 2D simulations. We fixed a constant random seed to evaluate each environment to ensure that all policies were tested against identical configurations.

A. *Ship-Ice Evaluations*

In *Ship-Ice*, the performance of SAC, PPO, lattice-planning policy, and predictive-planning policy are evaluated in 10% ice concentration with a 10-meter goal horizon. In each episode, the ice floes are randomly generated up to 10% concentration, and the ship is randomly initialized along a start line 10 meters away from the goal line.

Fig. 1 shows the evaluation results for PPO, SAC, lattice-planning policy, and predictive-planning policy. All baselines perform comparably in terms of efficiency with scores close

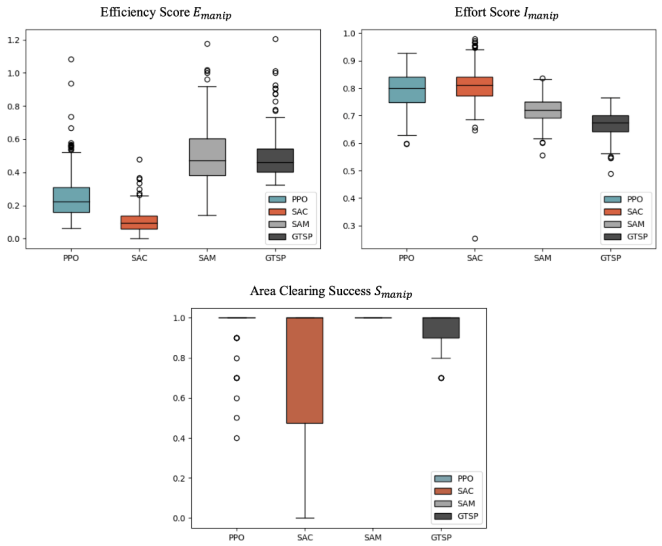

Fig. 2: Evaluation results for PPO, SAC, SAM, and GTSP in *Area-Clearing*.

to $E_{\text{nav}} = 1.0$. This is potentially due to the fact that evaluations are performed under a low ice concentration, so scenarios where long detours are required for avoiding large clusters of ice take place less frequently. In terms of collision avoidance, SAC gives the highest interaction effort scores, making it overall the best-performing model under 10% concentration setup.

APPENDIX III
Area-Clearing EVALUATIONS

We now evaluate the performance of PPO, SAC, SAM, and GTSP (Appendix I) policies for the *Area-Clearing* environment. For this evaluation, we used the environment configuration where the robot must clear 10 boxes from within a rectangular clearance boundary without any static walls or obstacles blocking its path. Each policy was run for 200 episodes, where the initial positions of the boxes are randomized after every episode. Also, the robot is initialized facing upwards at a random point close to the bottom of the environment. The episode terminates when all boxes are cleared out, or if the robot does not clear a box for 200 steps.

Figs. 2 show the evaluation results. All baselines achieve high task success scores, with SAM outperforming all baselines as it successfully clears all boxes in every episode. We also observe that SAM outperforms all baselines in efficiency, while SAC achieves the best effort score. This is because the SAM policy prefers stacking multiple boxes and clearing them simultaneously, while the SAC policy results in long paths with relatively minimal interaction with the boxes. However, in comparison with the GTSP planner, SAM achieves better effort scores as it considers object-to-object interactions while stacking boxes and minimizes the distance traveled by the boxes (lower effort). In comparison, the GTSP policy does not consider this and results in accidental stacking of boxes, which increases the robot's effort (see Appendix I).

REFERENCES

[1] M. Ramesh, F. Imeson, B. Fidan, and S. L. Smith, "Anytime Replanning of Robot Coverage Paths for Partially Unknown Environments," *IEEE Transactions on Robotics*, vol. 40, pp. 4190–4206, 2024.