Learning for TAMP using RAMP

Jack Collins*[1], Mark Robson*[2,3], Jun Yamada*[1], Mohan Sridharan[3], Karol Janik[2], Ingmar Posner[1]
*Equal contribution.
[1]Applied AI Lab, Oxford Robotics Institute, University of Oxford
[2]The Manufacturing Technology Centre, Coventry, United Kingdom
[3]University of Birmingham, Birmingham, United Kingdom
Correspondence to: jcollins@robots.ox.ac.uk

***Abstract*—RAMP, a benchmark for evaluating autonomous assembly and planning, provides an ideal environment for comparing approaches and assessing progress in Task and Motion Planning. RAMP is a recently released benchmark inspired by real-world industrial assembly tasks with the aim being to assemble a set of beams into predefined goal configurations. The benchmark is designed to be accessible with the physical benchmark created from 3D printed and other easily obtainable parts. The 3D printed parts can be extended upon and reconfigured into numerous designs, thereby making the benchmark open-ended. RAMP features a simulation environment to allow for virtual progress and therefore readily lends itself to learning-based approaches. To reduce the barrier to entry the benchmark includes a baseline and other conveniences, such as fiducial markers and layout templates, so users can focus on individual sub-tasks of the overall assembly challenge. RAMP has been designed intentionally to assess long-horizon planning and features many components that make it an ideal environment for learning and assessing task and motion planning.**

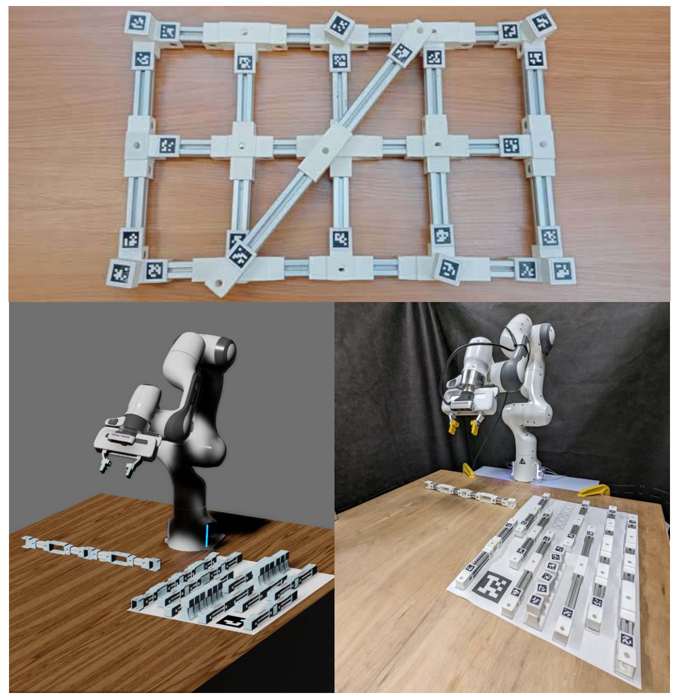


Fig. 1: *Top*: The benchmark beams and pegs with fiducial markers attached. *Bottom-Left*: Simulation environment with fixed beam on the left, additional beams and pegs laying on a layout template, Panda robot with optimised fingertips and wrist camera, and birds-eye camera looking down from above. *Bottom-Right*: Physical setup of the benchmark with the same setup as the simulation environment.

I. Introduction

The automation of offsite construction features many challenges, one of which is long-horizon planning of assembly sequences. Beams for the internal structure of a prefabricated building can be autonomously cut and punched to size using light-steel-roll-forming [21]. However, the assembly of these beams into a goal configuration (see Fig. 2) remains an entirely manual process as variation in the assembly can not be accounted for using traditional manufacturing techniques. Applying Task and Motion Planning (TAMP) is a compelling approach towards solving this problem. To this end, we have taken this industry-inspired problem and distilled it into a form available to researchers. The Robotic Assembly Manipulation and Planning benchmark or RAMP [4] is intended to focus research effort and catalyse progress in long-horizon industrial assembly tasks.

Designing a robotic manipulation benchmark is notoriously challenging as relying upon shared hardware systems and components across research labs is impractical. However, we pose that the greatest limitations to long-horizon industrial assembly tasks is not hardware but open challenges in perception, reasoning, manipulation skill creation, diagnostics, fault recovery and goal parsing. RAMP embodies these challenges whilst also allowing researchers to engage only in particular sub-tasks aligned with their specific research interests. This is possible through the availability of a simulation environment; the inclusion of fiducial markers for part identification and state estimation; and through the inclusion of a baseline that users can choose to leverage and build upon.

RAMP has been designed to overcome the common pitfalls of benchmarks for robotic manipulation with the intent that the benchmark should remain challenge-driven, accessible and open-ended. **Challenge-Driven.** RAMP is inspired by open challenges experienced in offsite construction. Three categories of assemblies (*easy*, *medium*, and *hard*) have been proposed with an evaluation protocol that reflects the industry requirement for repeatable assembly of parts alongside evaluation metrics that promote speed and completeness.

Accessible. Beams for RAMP are constructed from 3D printed parts and extruded aluminium profiles (see Fig. 1, top),

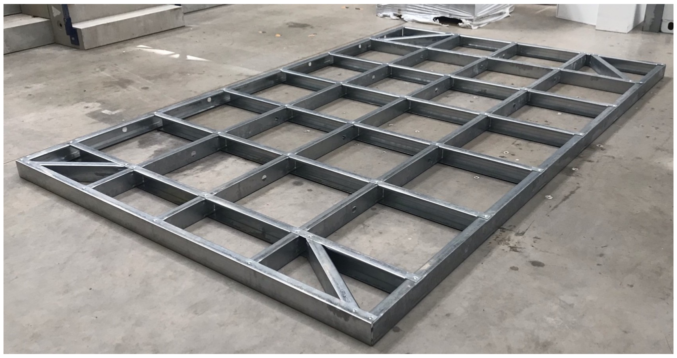


Fig. 2: Light steel-rolled beams automatically cut and punched to size, and manually assembled and fastened. Such assemblies take a lot of time and effort as each beam and assembly can be unique within a prefabricated building.

both widely accessible across research labs. The part files and detailed instructions on how to construct your own benchmarking set are publicly available[1]. Additionally, RAMP distributes with a high-fidelity simulation environment created in Nvidia Isaac [19] (see Fig. 1, bottom-left) mirroring a typical real-world setup. A baseline method has been implemented in both simulation and in a real-world setup to allow researchers to build upon or to use as substrate in conjunction with their own sub-task solutions.

Open-Ended. While RAMP has been released with a number of predefined goal configurations, the beams and 3D printed parts can be reconfigured in a variety of ways. New parts can also be readily created to broaden the spectrum of reachable configurations (e.g. parts for creating 3D structures). The benchmark protocol has been configured to ensure the goal configurations are achievable with a single-arm system, although, RAMP readily lends itself to grow with community capabilities to extend to more challenging setups (e.g. multi-agent system configurations, inclusion of deformable cabling, etc.). Our vision is to see RAMP grow into a community-driven effort which evolves to meet the needs of the community.

II. Literature Review

Benchmarks can have a significant impact in catalysing communities into progressing the state-of-the-art. Several domains that have seen great advancements through the use of benchmarks include object detection [15], visual odometry [12] and reinforcement learning [1]. A number of robot manipulation benchmarks have been proposed [18, 9, 2, 3, 23, 6] with a subset of environments covering TAMP problem domains [8, 24, 16, 10].

The most notable benchmark for TAMP proposed by Lagriffoul *et al.* [16] introduces five simulated environments that assess a range of properties that make TAMP difficult. Assessed properties include infeasible task actions, large task spaces, motion/task trade-offs, non-monotonicity and non-geometric actions [16]. Environments for their benchmark include Towers of Hanoi, block environments and a kitchen environment. RAMP assess four of the five properties introduced by Lagriffoul *et al.* but does not, in its current configuration, assess non-geometric actions where there is a non-geometric state change in objects.

There are several commonalities shared across most environments used to assess TAMP methods [8, 24, 16, 10]. One commonality is the extensive use of simulation with no corresponding physical benchmark. Another common assumption is that perfect state information is available, which is the case in simulation although not the case in real-world environments. The scenes used in the environments are often toy problems that have objects with only simple geometries, although, ThreeDWorld [8] provides a more realistic environment in this sense. Finally, the environments require only simple pick-and-place manipulations.

Our benchmark, RAMP, builds on components of these environments whilst also providing some unique contributions useful to the TAMP community. First, it is designed to be accessible, providing both a simulation and physical benchmark that can be assessed using the same metrics and using the same evaluation protocol. Second, it is challenge-driven including challenging objects and manipulations that are inspired by industry. Finally, it is open-ended with the base components capable of generating many unique beams and a great number of goal configurations, useful for learning and assessing generalisable policies. Overall, we make it easy for new users to get started with a simulation environment, baseline code and fiducial markers whilst also creating an environment that mimics industry assembly challenges.

III. Benchmark

RAMP is designed to assess long-horizon assembly tasks with the core principles of providing a challenge-driven, accessible and open-ended benchmark. The benchmark consists of an extensible set of base parts used to construct beams, a predefined set of goal configurations, metrics for assessing performance, an evaluation protocol to promote reproducibility and a high-fidelity simulation environment. We provide an overview of each of these components below. For specifics, we refer the reader to [4].

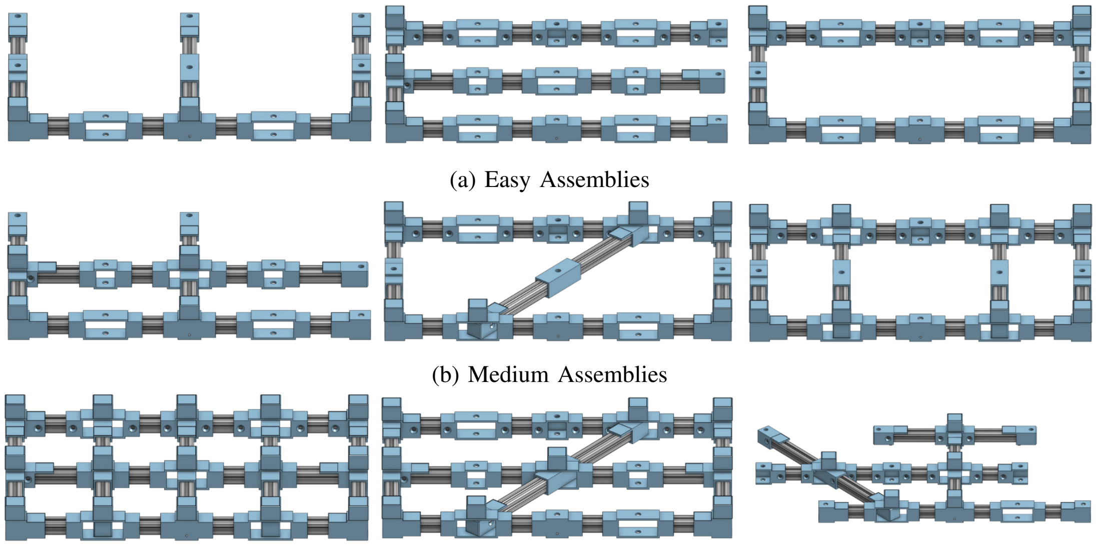
A. Base Parts

The base set of parts are released as STL files which are made publicly available. These parts can be composed into beams in any number of ways with 20×20 mm extruded-profile aluminium profile used to link parts together. We define nine beams along with 15 pegs—see Fig.1 (top)—for the benchmark. Space is made on select parts for a $25 \times 25mm$ April Tag [20], this is to assist with beam identification and localisation. For an exact bill of parts and an assembly guide please visit the RAMP benchmark website[2].

Other parts have also been developed as part of the benchmark and can be found at the project website. Parts include

[1] https://sites.google.com/oxfordrobotics.institute/ramp/create-your-own

[2] https://sites.google.com/oxfordrobotics.institute/ramp



(a) Easy Assemblies

(b) Medium Assemblies

(c) Hard Assemblies

Fig. 3: Three classes of assemblies for assessing the capabilities of any proposed solutions to the benchmark.

robot fingertips (useful for picking beams), clamps and peg holders that can be 3D printed as well as a A2 layout template to assist in laying out parts repeatably. We anticipate the number of beams, assemblies and parts to evolve as the benchmark grows to include more challenges over time.

B. Tasks

The goal configurations for RAMP can be seen in Fig. 3 with three classes of difficulty: easy, medium and hard. The increase in difficulty across the classes is evident with harder assemblies requiring more physical interactions to complete as well as additional skills not required for the easy assembly class. These goal configurations should act as a holdout set with similar designs also achievable using the same method but with the relevant goal conditioning.

To pass a goal for planning, any modality is supported, including images, natural language, etc. We include an Extensible Markup Language (XML) based parser alongside the benchmark that is used for describing both the beam configurations and goal configurations.

C. Performance Measures

To assess performance, metrics have been chosen to reflect the needs of industry, which are (i) a complete assembly, and (ii) an assembly completed in the shortest amount of time. Thus the primary metric is completion percentage with the secondary metric being time to completion. Time to completion must include the planning time and execution time. Results should be presented similar to that found in Fig.4.

D. Evaluation Protocol

For reproducibility, we suggest users of the benchmark follow several constraints, although, these are guidelines and we anticipate the benchmark evolving to best suit the needs of the community. To apply the benchmark an entire class of assemblies must be attempted with each assembly repeated five times consecutively without any changes to the method except for the conditioning goal. There are no constraints on systems, any robot may be used to attempt the benchmark. Similarly, any gripper and sensors are able to be used with exploration of new modalities encouraged.

The arrangement of the robot and beams should be similar to those seen in Fig. 1, bottom-right, with the starting beam fixed rigidly to the table and the layout template lying outside the assembly area. The rigid attachment of the first beam to the table top allows for the assembly to be attempted with single-arm systems, an easy constraint to remove in future to allow for multi-agent systems.

E. Simulation

RAMP includes a simulation environment configured to mirror a typical setup with a Franka Emika Panda, two real-sense cameras and beams laid out on the A2 layout template (see Fig. 1 bottom-left). The simulation environment is built using the Nvidia Isaac simulator [19] and in particular, leverages signed distance field collisions [17] to faithfully represent the physical setup. The simulation environment provides observations, including state-based, image-based and robot proprioceptive data, and also provides a high-level action interface with Cartesion position and impedance control.

The simulation environment facilitates virtual progress and testing but equally important, unlocks the ability to apply a broad range of learning-based approaches. To expand the current environment to further take advantage of learning-based approaches and the open-ended benchmark we look to create procedurally generated beams and assemblies whilst also leveraging the ability of Nvidia Issac to run many simulation environments simultaneously.

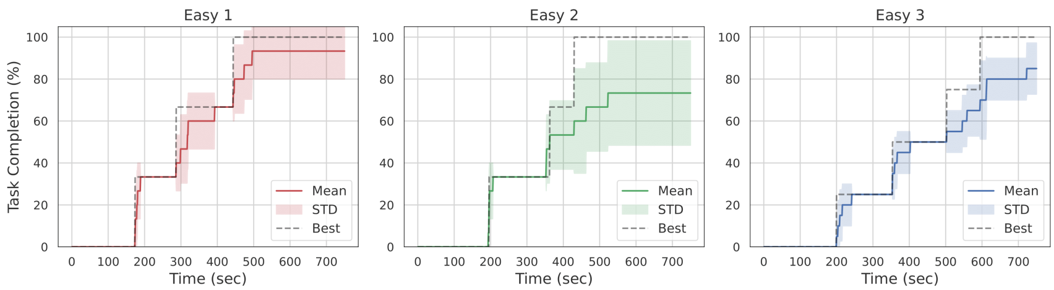


Fig. 4: Results of the baseline method across the three assemblies in the easy class of the benchmark, with the percentage of completion of the desired assembly expressed as a function of the time (planning and execution, in seconds). The best and average of the repeats are plotted along with the standard deviation.

IV. Baseline

RAMP is released with a baseline method that is made publicly available, therefore lowering the barrier to entry by allowing adopters to build upon and use it within their own systems. The baseline takes a traditional approach to solve the easy class of assemblies with the implementation and performance of the baseline discussed in the following sections.

A. Baseline Implementation

The baseline method can be separated into two components, the task planner and skill execution. The task planner is adapted from the REBA robot architecture[22] and reasons over the problem at two resolutions, coarse and fine. At the coarse resolution, the task planner computes a sequence of abstracted actions, with each action transition then expanded upon by the fine resolution planner. The fine resolution planner outputs an ordered list of actions which are then executed open-loop by the robot. The fine resolution actions are *move(robot, place)*, *pick_up(robot, part)*, *put_down(robot, part)*, *assemble_square(robot, joint)*, *assemble_cap(robot, joint)*, *fasten(robot, joint, joint, peg)* and *push(robot, beam)*.

Each action is executed on the robot using hand-designed skills created intentionally as the simplest approach for the baseline. The skills rely upon information available from XML files that describe the beam configurations; April Tag identification and state estimation [20]; and proprioceptive information available from the robot. For simple skills, such as *move*, *pick_up*, and *put_down*, the *MoveIt* [13] motion planner is used to find a valid path. For all other skills, all of which have contact-rich interactions, force-feedback and Cartesian impedance control is used to compose each skill.

B. Results

Performance of the baseline method evaluated using the evaluation protocol and metrics from Section III is presented in Fig. 4. To summarise quantitatively the baseline method performance across five repeats of each of the three easy assemblies, on average the baseline achieves 84% in an average time of $580sec$. A significant portion of failure cases are caused by an inability to fasten beams together with a peg. This is often caused by poor state estimation of the hole location or because of alignment issues between two beams.

V. Looking Ahead

There are many challenges and opportunities to leverage learning-based approaches to overcome limitations in current TAMP methods. Some current challenges and limitations as highlighted by Guo *et al.* [14] and others include TAMP for real-world applications with imperfect perception and execution [5], online closed-loop planning [11], multi-robot cooperation [7], and domain knowledge representations. RAMP provides an ideal environment for developing and testing solutions that overcome some of these challenges as the problem domain of RAMP already encompasses these challenges or can be readily extended to encompass these challenges. RAMP, being a benchmark, also allows for direct comparison between proposed approaches and as it is designed to be accessible, will allow for reproducible research and comparison of methods across institutions.

We see learning-based approaches for TAMP as an interesting prospect as applied to RAMP as the problem domain is non-trivial to construct, task planning in itself is time intensive and open-loop planning is a major limitation in achieving high success rates. Learning-based approaches offer an opportunity to make significant progress on some of these challenges and are facilitated within the benchmark framework with the availability of a high-fidelity simulation environment and an open-ended domain that offers the opportunity of many possible beam and goal configurations to assist with generalisation and testing.

VI. Conclusion

We introduce RAMP, a benchmark designed to evaluate progress on automated assembly and planning, as an ideal environment for designing, training and assessing learning-based TAMP methods. RAMP is inspired by industrial assembly tasks featured in the domain of offsite construction and distilled into a form that is easily accessible to researchers. As such, RAMP is designed to be challenge-driven, accessible and open-ended. RAMP is released with a high-fidelity simulation environment and a baseline approach that users can build upon or use as substrate in conjunction with their own sub-task solutions.

References

[1] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. Openai gym, 2016.

[2] B Calli, A Walsman, A Singh, S Srinivasa, P Abbeel, and A M Dollar. Benchmarking in Manipulation Research: Using the Yale-CMU-Berkeley Object and Model Set. *IEEE Robotics & Automation Magazine*, 22(3):36–52, 2015. ISSN 1558-223X.

[3] J Collins, J McVicar, D Wedlock, R Brown, D Howard, and J Leitner. Benchmarking Simulated Robotic Manipulation through a Real World Dataset. *IEEE Robotics and Automation Letters*, 2019. ISSN 2377-3774.

[4] Jack Collins, Mark Robson, Jun Yamada, Mohan Sridharan, Karol Janik, and Ingmar Posner. Ramp: A benchmark for evaluating robotic assembly manipulation and planning. 5 2023. URL https://arxiv.org/abs/2305.09644.

[5] Aidan Curtis, Xiaolin Fang, Leslie Pack Kaelbling, Tomás Lozano-Pérez, and Caelan Reed Garrett. Long-horizon manipulation of unknown objects via task and motion planning with estimated affordances. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 1940–1946. IEEE, 2022.

[6] Sudeep Dasari, Jianren Wang, Joyce Hong, Shikhar Bahl, Yixin Lin, Austin Wang, Abitha Thankaraj, Karanbir Chahal, Berk Calli, Saurabh Gupta, David Held, Lerrel Pinto, Deepak Pathak, Vikash Kumar, and Abhinav Gupta. Rb2: Robotic manipulation benchmarking with a twist, 2022.

[7] Danny Driess, Jung-Su Ha, and Marc Toussaint. Deep visual reasoning: Learning to predict action sequences for task and motion planning from an initial scene image. *arXiv preprint arXiv:2006.05398*, 2020.

[8] Chuang Gan, Siyuan Zhou, Jeremy Schwartz, Seth Alter, Abhishek Bhandwaldar, Dan Gutfreund, Daniel L K Yamins, James J DiCarlo, Josh McDermott, Antonio Torralba, and Joshua B Tenenbaum. The threedworld transport challenge: A visually guided task-and-motion planning benchmark towards physically realistic embodied ai. pages 8847–8854, 2022.

[9] I Garcia-Camacho, M Lippi, M C Welle, H Yin, R Antonova, A Varava, J Borras, C Torras, A Marino, G Alenyà, and D Kragic. Benchmarking Bimanual Cloth Manipulation. *IEEE Robotics and Automation Letters*, 5 (2):1111–1118, 2020. ISSN 2377-3766.

[10] Caelan Reed Garrett, Tomás Lozano-Pérez, and Leslie Pack Kaelbling. Pddlstream: Integrating symbolic planners and blackbox samplers via optimistic adaptive planning. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 30, pages 440–448, 2020.

[11] Caelan Reed Garrett, Chris Paxton, Tomás Lozano-Pérez, Leslie Pack Kaelbling, and Dieter Fox. Online replanning in belief space for partially observable task and motion problems. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5678–5684, 2020.

[12] A Geiger, P Lenz, and R Urtasun. Are we ready for autonomous driving? The KITTI vision benchmark suite. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3354–3361, 2012. ISBN 1063-6919.

[13] Michael Görner, Robert Haschke, Helge Ritter, and Jianwei Zhang. Moveit! task constructor for task-level motion planning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 190–196. IEEE, 2019.

[14] Huihui Guo, Fan Wu, Yunchuan Qin, Ruihui Li, Keqin Li, and Kenli Li. Recent trends in task and motion planning for robotics: A survey. *ACM Computing Surveys*, 2023.

[15] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25:1097–1105, 2012.

[16] Fabien Lagriffoul, Neil T Dantam, Caelan Garrett, Aliakbar Akbari, Siddharth Srivastava, and Lydia E Kavraki. Platform-independent benchmarks for task and motion planning. *IEEE Robotics and Automation Letters*, 3: 3765–3772, 2018.

[17] Miles Macklin, Kenny Erleben, Matthias Müller, Nuttapong Chentanez, Stefan Jeschke, and Zach Corse. Local Optimization for Robust Signed Distance Field Coll ision. *Proc. ACM Comput. Graph. Interact. Tech.*, 3(1), 5 2020.

[18] H Mnyusiwalla, P Triantafyllou, P Sotiropoulos, M A Roa, W Friedl, A M Sundaram, D Russell, and G Deacon. A Bin-Picking Benchmark for Systematic Evaluation of Robotic Pick-and-Place Systems. *IEEE Robotics and Automation Letters*, 5(2):1389–1396, 2020. ISSN 2377-3766.

[19] NVIDIA. Nvidia isaac sim. URL https://developer.nvidia.com/isaac-sim.

[20] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*, pages 3400–3407, 2011.

[21] Ryan E Smith. Off-site and modular construction explained. *National institute of building sciences*, 2016.

[22] Mohan Sridharan, Michael Gelfond, Shiqi Zhang, and Jeremy L. Wyatt. REBA: A refinement-based architecture for knowledge representation and reasoning in robotics. *J. Artif. Intell. Res.*, 65:87–180, 2019.

[23] A Suarez, V M Vega, M Fernandez, G Heredia, and A Ollero. Benchmarks for Aerial Manipulation. *IEEE Robotics and Automation Letters*, 5(2):2650–2657, 2020. ISSN 2377-3766.

[24] Ryan Yang, Tom Silver, Aidan Curtis, Tomas Lozano-Perez, and Leslie Pack Kaelbling. Pg3: Policy-guided planning for generalized policy generation. 2022.