

ALOHA 2: An Enhanced Low-Cost Hardware for Bimanual Teleoperation

ALOHA 2 Team, Jorge Aldaco¹, Travis Armstrong¹, Robert Baruch¹, Jeff Bingham¹, Sankh Chan¹, Kenneth Draper¹, Debidatta Dwibedi¹, Chelsea Finn^{1,2}, Pete Florence¹, Spencer Goodrich¹, Wayne Gramlich¹, Torr Hage¹, Alexander Herzog¹, Jonathan Hoech¹, Thinh Nguyen¹, Ian Storz¹, Baruch Tabanpour¹, Leila Takayama^{1,3}, Jonathan Tompson¹, Ayzaan Wahid¹, Ted Wahrburg¹, Sichun Xu¹, Sergey Yaroshenko¹, Kevin Zakka¹ and Tony Zhao^{1,2}

¹Google DeepMind, ²Stanford University, ³Hoku Labs, Authors listed in alphabetical order, with contributions listed in Appendix.

Diverse demonstration datasets have powered significant advances in robot learning, but the dexterity and scale of such data can be limited by the hardware cost, the hardware robustness, and the ease of teleoperation. We introduce *ALOHA 2*, an enhanced version of *ALOHA* that has greater performance, ergonomics, and robustness compared to the original design. To accelerate research in large-scale bimanual manipulation, we open source all hardware designs of *ALOHA 2* with a detailed tutorial, together with a MuJoCo model of *ALOHA 2* with system identification. See the project website at aloha-2.github.io.

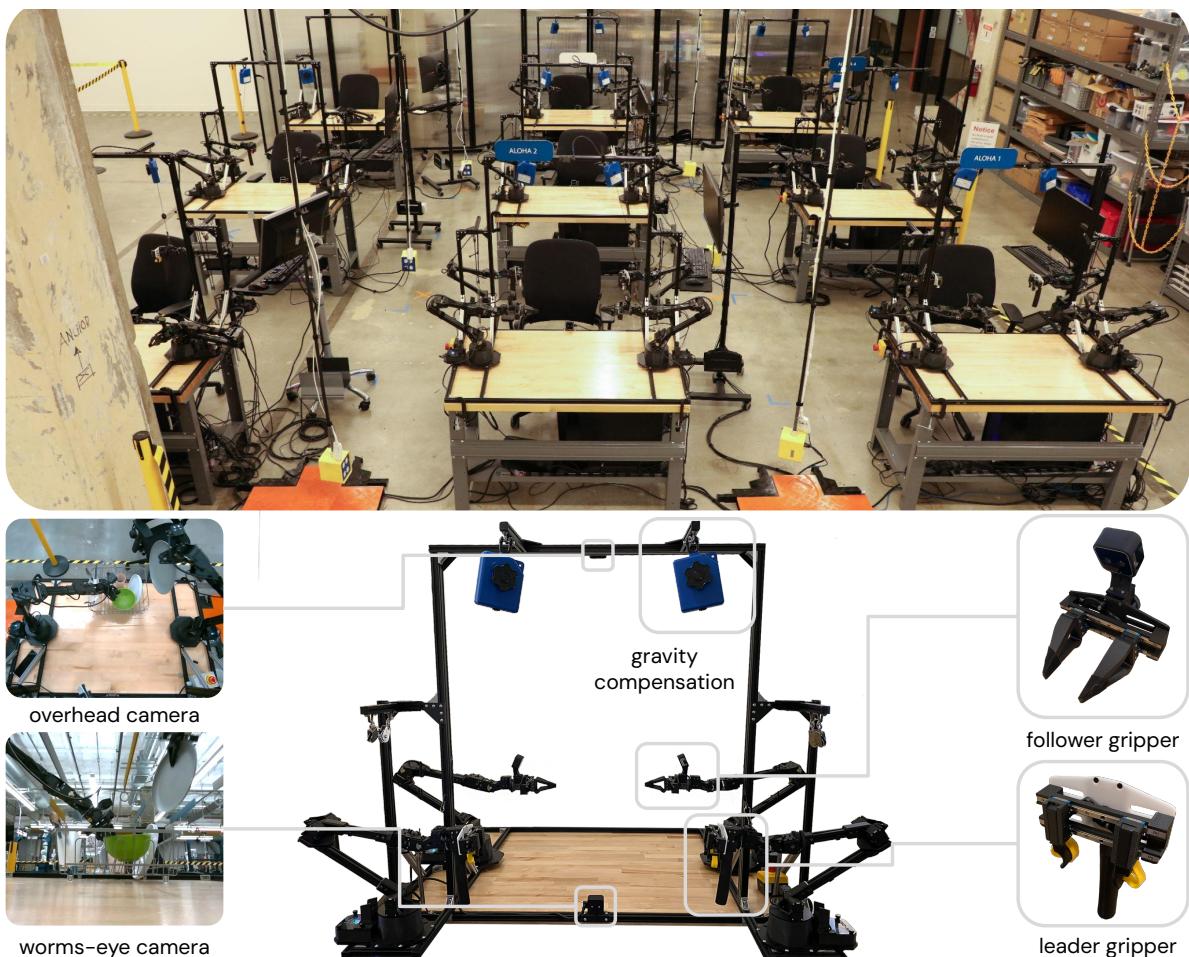


Figure 1 | *ALOHA 2: An Enhanced Low-Cost Hardware for Bimanual Teleoperation*. **Top:** The fleet of *ALOHA 2* robots capable of collecting 1000s of demonstrations per day. **Bottom:** A detailed image of an *ALOHA 2* workcell with gravity compensation, the redesigned leader and follower grippers, and images from the frame-mounted cameras.

ALOHA 2

ALOHA 2, like the original [Zhao et al. \(2023\)](#), consists of a bimanual parallel-jaw gripper workcell with two ViperX 6-DoF arms ([Trossen Robotics, a](#)) (the "follower"), along with 2 smaller WidowX arms ([Trossen Robotics, b](#)) (the "leader"). The WidowX contains the same kinematic structure as the ViperX, in a smaller form factor. The follower joints are synchronized with the leader arms, and users teleoperate the follower arms by backdriving, or "puppeteering", the leader arms. The setup also contains cameras that produce images from multiple viewpoints, allowing collection of RGB data during teleoperation. The robots are mounted on a 48" x 30" table, with an aluminum cage to provide additional mount points for cameras and a gravity compensation system, which are detailed below.

To support research on complex manipulation tasks, we aim to significantly scale up data collection on the ALOHA platform, including the number of robots in use, the amount of hours of data per robot, and the diversity of data collection. This scaling process shifts the requirements and scope relative to the first ALOHA platform. For ALOHA 2, we build on the strengths of ALOHA platform while also targeting the following areas for further improvement:

- **Performance and Task Range:** We seek to enhance key components that enable ALOHA's performance, including grippers and controllers, to enable a broader set of manipulation tasks.
- **User Friendliness and Ergonomics:** To optimize data collection at scale, we prioritize user experience and comfort. This includes improvements to the responsiveness and ergonomic design of user-facing systems.
- **Robustness:** We want to increase system robustness to minimize downtime caused by diagnosis and repairs. This involves simplifying mechanical designs and ensuring the overall ease of maintenance for a larger robot fleet.

To this end, we make the following concrete improvements:

- **Grippers:** We create a new low-friction rail design for both the leader and follower grippers. For the leader robots, this improves teleoperation ergonomics and responsiveness. For the followers, this improves latency and the force output of the grippers. In addition, we upgrade the grip tape material on the fingers to improve durability and grasping of small objects.
- **Gravity Compensation:** We create a passive gravity compensation mechanism using off-the-shelf components that improves the durability compared to ALOHA's original rubber band system.
- **Frame:** We simplify the frame surrounding the workcell, while maintaining the rigidity of the camera mounting points. These changes allow space for both human-robot collaborators and props for the robot to interact with.
- **Cameras:** We use smaller Intel RealSense ([Keselman et al., 2017](#)) D405 cameras and custom 3D-printed camera mounts to reduce the footprint of the follower arms, which less inhibits manipulation tasks. These cameras also have a larger field of view, provide depth, have global shutter, and allow for more customization compared to the original consumer-grade webcams.
- **Simulation:** We model the exact specifications of the ALOHA 2 robot in a MuJoCo model in MuJoCo Menagerie, which allows improved data collection, policy learning, and evaluation in simulation for challenging manipulation tasks.

We find that these improvements make it easier to teleoperate challenging tasks like folding a T-shirt, tying a knot, throwing objects, or industrial tasks with tight tolerances. These improvements make it easier to collect 100s of demonstrations on these tasks per robot per day.

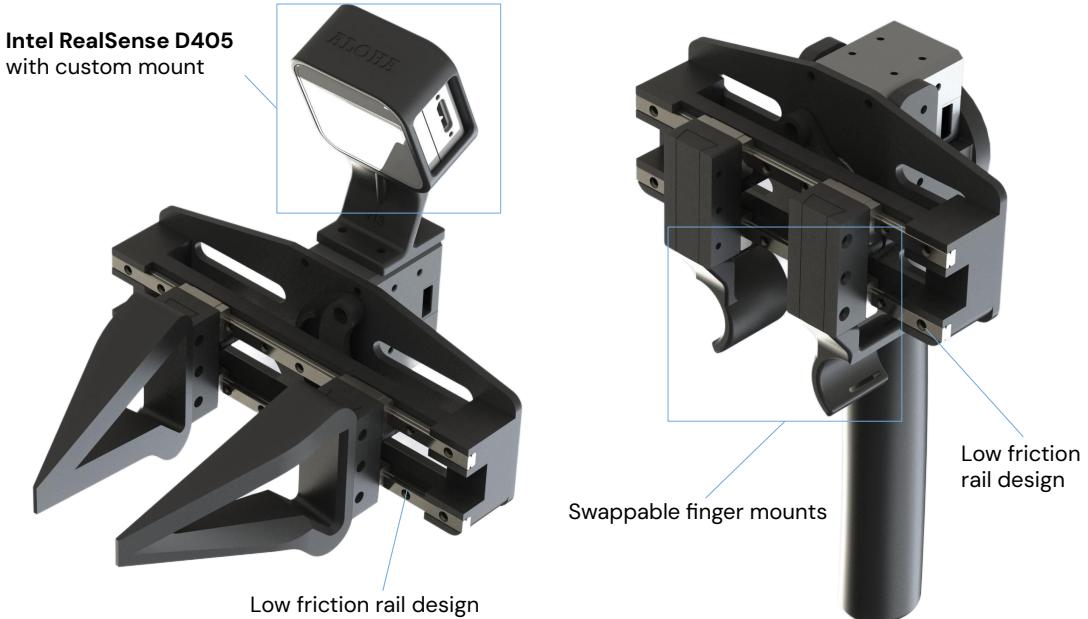


Figure 2 | **Renderings of leader and follower gripper designs.** **Left:** follower gripper with Intel RealSense D405, custom 3d-printed camera mount, and low friction follower rail design. **Right:** leader gripper with swappable finger mounts and low friction leader rail design.

Hardware

Leader Grippers (*The device the human holds in their hands*)

For smoother teleoperation and improved ergonomics, we replace the original scissor leader gripper design from *ALOHA* with a low friction rail design with reduced mechanical complexity. To further reduce strain, we also lower the backdriving friction by swapping the original leader's gripper motors (XL430-W250-T) with a lower-friction alternative (XC430-W150-T), which have a lower gear ratio and use lower-friction metal instead of plastic gears.¹ The new design requires approximately 10 times less force to open and close than the previous *ALOHA* scissor design (Fig. 4). The lower friction significantly reduces the operator's hand fatigue and strain during long data collection sessions, notably on the lumbrical muscles² responsible for opening the gripper.

When deciding on the linear rail design, we also considered two additional designs. First is the original *ALOHA* scissor design, which uses custom 3d printed rotors and rails to backdrive the leader gripper motor. In addition, we evaluate a spring loaded trigger design, where pulling the trigger closes the gripper and releasing it opens the gripper to its neutral open position (See Figure 3 for images and renderings of the evaluated grippers). We had 6 users teleoperate *ALOHA* 2 to unwrap candy using the original *ALOHA* scissor design, the linear rail design, and the trigger design. While users had varying preferences, the linear rail was rated well by nearly all operators.

Follower Grippers (*The end-effector of the robot, i.e. the robot fingers that manipulate objects*)

We design and manufacture low-friction follower grippers, replacing the original design from *ALOHA*. The lower friction reduces perceived latency between leader and follower grippers, significantly

¹<https://www.youtube.com/watch?v=JwAkSwwaOA4>

²<https://my.clevelandclinic.org/health/body/25060-anatomy-of-the-hand-and-wrist>



Figure 3 | *The leader grippers evaluated for ALOHA 2.* **Left:** the original ALOHA scissor design. **Middle:** low friction rail design which was ultimately chosen based on user studies. **Right:** rendering of the trigger design that was also considered.

improving user experience during teleoperation. We show the difference in latency between the old design and new design in the supplementary video. The new grippers are also capable of applying 2 times more force than the old design, allowing for stronger and more stable grasps of objects.

In addition, we improve the compliance of the gripper mechanism by removing the original PLA + acrylic structure, and replacing it with 3D printed carbon fiber nylon. Both the gripper fingers and the supporting structure can deform when loaded, improving the safety of the system.

We keep the "see-through" design of the finger links from the original ALOHA. In addition, we make improvements to the gripping tape on the fingers. We find that the original gripping tape on ALOHA wears out over time, and the roundness at the tips of the fingers makes picking up small objects difficult. We apply a polyurethane gripping tape to the inside of the gripper. We also apply strips of tape on the outside of the finger to increase traction for tasks that require manipulating objects with the outer side of the gripper.

Gravity Compensation

We design a more robust *passive* gravity compensation for the leader robots to ease operator wear during teleoperation. We construct this using off-the-shelf components, including adjustable hanging retractors that allows operators to tune the load balancing forces to their comfort level.

We run a study to evaluate the *passive* hardware gravity compensation system against an *active* software-based system. We develop the active system using inverse dynamics from the MuJoCo model to calculate equivalent torques for the gravity load and then command these torques to the joints of the leader robot. To conduct the study, 6 users teleoperate the robots for 10 minute sessions, and attempt to perform a precise task of inserting shapes into corresponding holes in a box. The operators attempt the task on both systems in a randomly assigned order. We compare performance on the task across the two systems, and find that on average operators performed better with the passive gravity compensation system (1.38 vs 0.97 shapes inserted per minute). Based on feedback from study participants, we speculate that the passive system allows more smooth and predictable movements. Study participants mention that the active system requires more force and

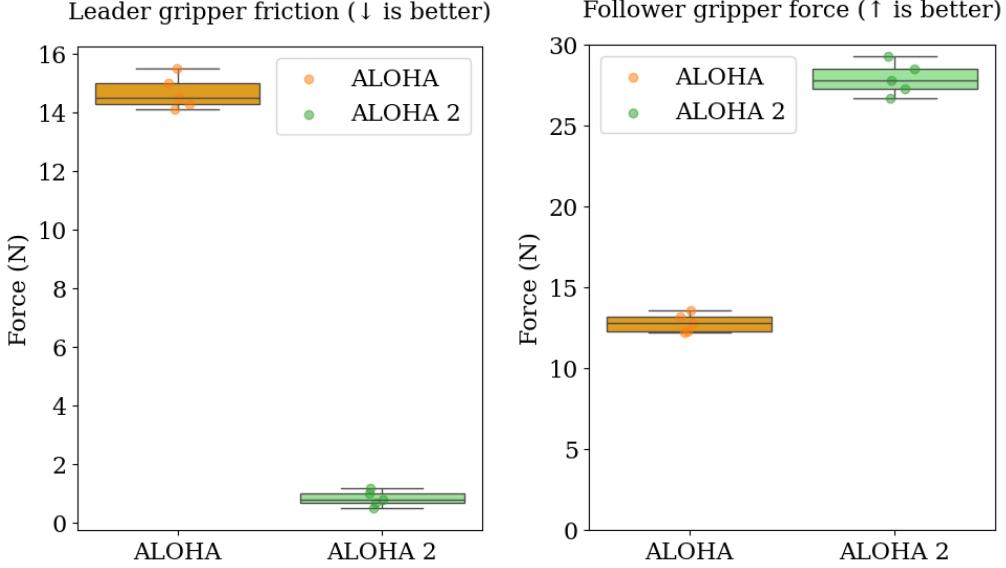


Figure 4 | ALOHA 2 improves the ergonomics and closing force of grippers. **Left:** Force required from the operator to open leader grippers. ALOHA 2 reduces the force from 14.68N to 0.84N, reducing hand fatigue and improving ergonomics. **Right:** Closing force at the tip of follower grippers. ALOHA 2 is capable of exerting more than double the force compare to the old design (27.9N vs. 12.8N)

results in choppier movement, perhaps due to poorly tuned servo gains or slight latency in applying counteracting forces. In addition, we find that the passive system provides two additional advantages. First, it can allow safer teleoperation by fully disabling joint actuation on the leaders, as software bugs or edge cases may cause large or unexpected movements of the leader robots. Second, the force retractors give a natural centering for the arms to keep the wrists on the leader arms from over-rotating, which seemed to be a weakness of the active gravity compensation. Despite choosing the passive system for ALOHA 2, we speculate that an active system can perhaps be developed and tuned to surpass performance of our passive system, and could potentially be extended to allow useful features like providing tactile feedback to the user.

Frame

We redesign the support frame and build it using 20x20mm aluminum extrusions. The frame provides support for the leader robots, gravity compensation system, and provides mount points for the overhead and worms-eye cameras. Compared to ALOHA, we simplify the design to remove the vertical frames on the side of the table opposite to the teleoperator. The added space allows for diverse styles of data collection. For example, a human *collaborator* can more easily stand at the opposite side of the workspace and interact with the robot, allowing collection of human-robot interactive data. Additionally, larger props can be placed in front of the table for the robot to interact with.

Cameras

We upgrade the cameras used in the ALOHA system to 4 RealSense D405 cameras. These cameras enable high resolution RGB and depth in a small form factor, as well as provide global shutter. We note that although depth and global shutter were not necessary for the results demonstrated on the first ALOHA system, they might be considered “nice to haves” for enabling different experimentation and pushing performance. We design new camera mounts for both the wrist cameras, as well as the

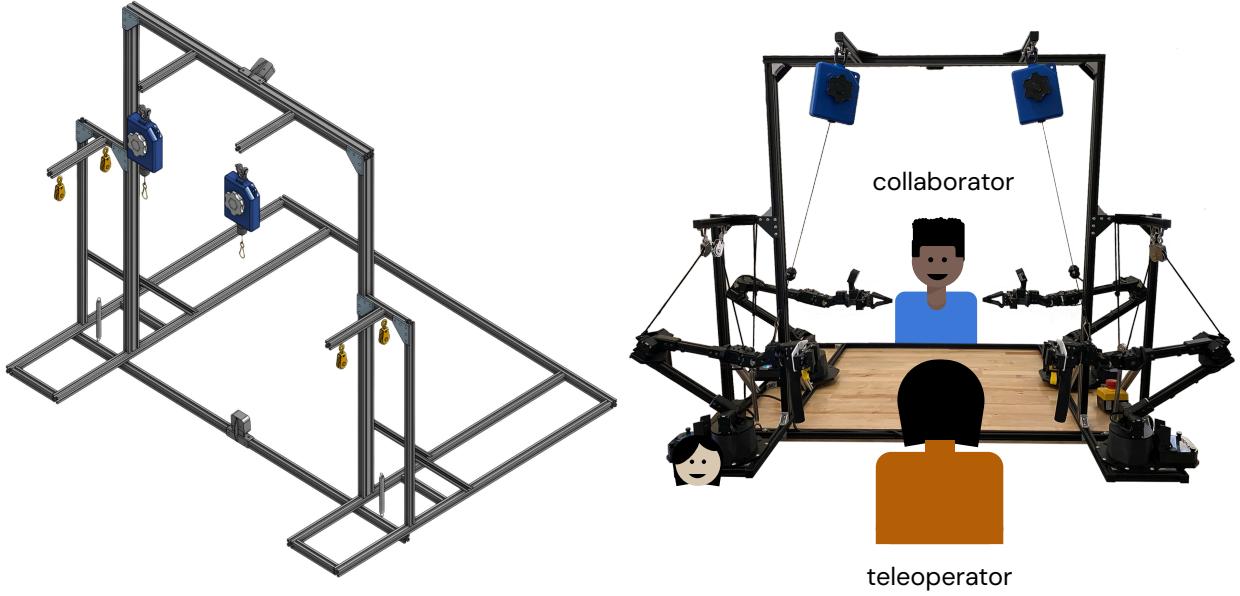


Figure 5 | ***The redesigned frame.*** **Left:** A rendering of the frame, which provides structure for gravity compensation and mount points for the cameras. **Right:** The frame allows space for collection of human-robot collaborative data.

overhead and worms-eye views (See [Figure 6](#)). The lower profile of the cameras on the wrists reduces the number of collision states and improves teleoperation for certain fine grained manipulation tasks, especially those that require close contact between the arms or navigating through tight spaces.

Teleoperation

We run the teleoperation software stack using ROS2 ([Macenski et al., 2022](#)). Upon startup, both leader and follower arms initialize to the home position. Operators can start a data collection *session* by closing the gripper using the finger attachments on either of the leader robots. Operators can save or discard a session using foot pedals positioned underneath the workcell.

During a teleoperation session, we log sensor streams from the robot including images, leader and follower joint positions, and other auxiliary data provided by the ROS2 system. We take several measures at collection time to ensure that downstream pipelines receive complete, high quality data, since this is crucial for robotic learning pipelines. Session statistics such as sensor availability and latency are visible to the operator during collection to ensure data is reliably logged. Sessions are automatically shut down for missing data to ensure downstream learning pipelines always receive complete data. When sessions are logged, we record the operator username, time, and the robot identifier along with the raw sensors streams from the robot. Including the additional data allows filtering data downstream if issues are found for certain robots during a period of time. We record leader and follower joint data at 50Hz, which as the original *ALOHA* system showed, outperforms lower frequency teleoperation.

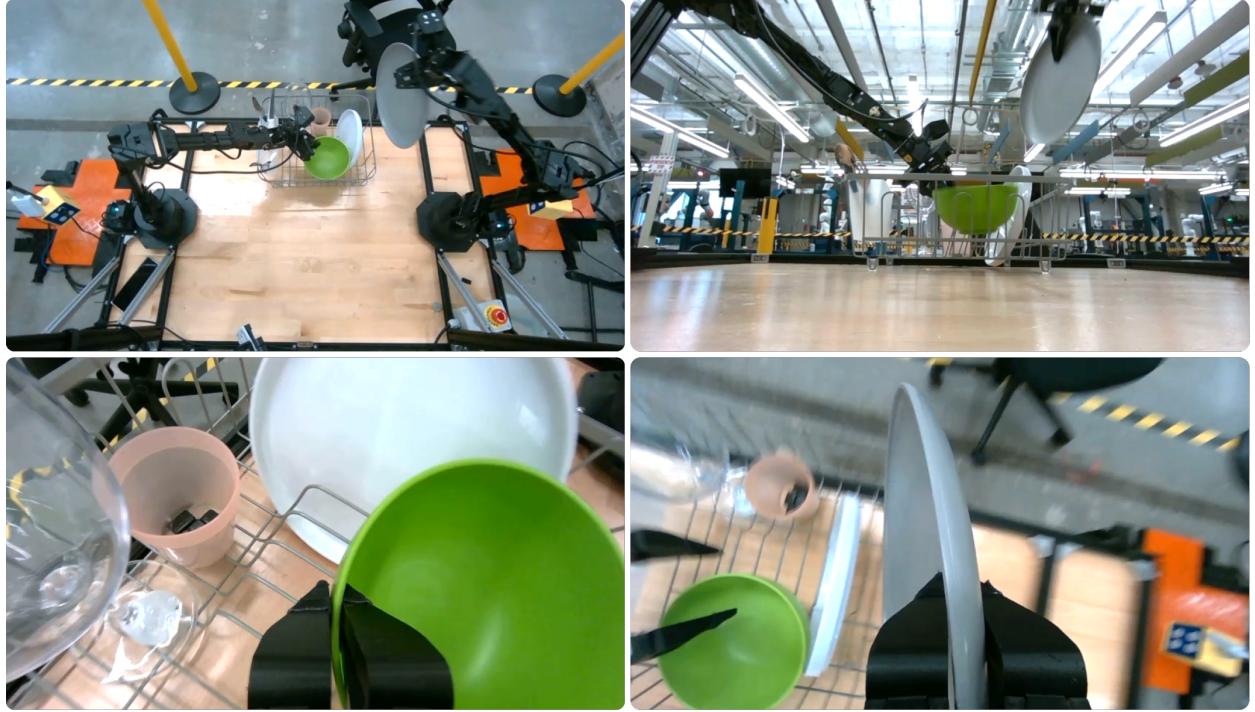


Figure 6 | *The four camera views recorded during teleoperation on a real workcell.* From left to right: overhead camera, worms-eye camera, left wrist camera, right wrist camera. All four cameras record 848 x 480 rgb images.

Ergonomics. Much of the motivation for the improvements highlighted in this report is to ensure comfort for users during teleoperation. We reiterate several of the features, and add several new considerations to maximize ergonomic benefit:

- **Low friction grippers** reduce strain on fingers and wrists during teleoperation.
- **Passive gravity compensation** counteracts the weight of the leader arms, to reduce wear on shoulders and arms during long teleoperation sessions.
- **Swappable finger attachments** of different sizes on the leaders allow customization for users with different hand sizes.
- **Adjustable height chairs** allow users to adjust to their optimal height during teleoperation.
- **Suggested rest intervals** for users to take breaks at least 2 minutes long, and avoid long continuous sessions. We observe that taking breaks at least every 10 minutes minimizes wear from repetitive motion. Users are also instructed to take breaks upon feeling any signs of fatigue, and mix up tasks to teleoperate between breaks to avoid continuous repeated motions for a single task.

Simulation

We release a MuJoCo Menagerie (Todorov et al., 2012) (Zakka et al., 2022) model of the *ALOHA* 2 workcell, useful for teleoperation and learning in simulation. The new model is **more physically accurate** and has **higher visual fidelity** than previously released *ALOHA* models. We perform system identification using logged trajectories from a real *ALOHA* 2 setup to set physics parameters in the MuJoCo model. In particular, we collect 11 trajectories in real using the leader arms and minimize the residuals between real and simulated trajectories using a nonlinear least squares solver. Real

trajectories consist of sinusoidal motions targeting the control limits of motors in the follower arm. The optimization tunes the proportional gain, damping, armature, joint friction, and torque limits of all position controlled actuators. The gripper is modeled using a position controlled linear actuator with an equality constraint between the gripper fingers. For higher visual fidelity, we match camera intrinsics as closely as possible to the real setup (See [Figure 8](#) for the simulated camera views), and we import assets for the table, table extrusions, and follower grippers.

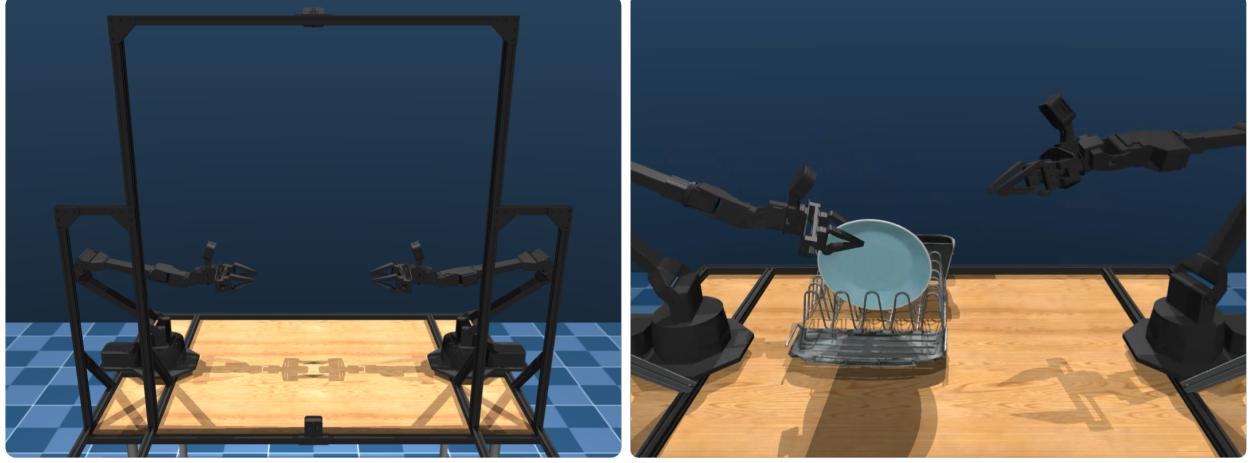


Figure 7 | *Rendering of the MuJoCo model.* The model contains ViperX robots with wrist cameras, mounted on a replica of the aluminum extrusion frame. We precisely model all cameras and robot positions of the ALOHA 2 specification, and perform system identification to ensure similar behavior to real.

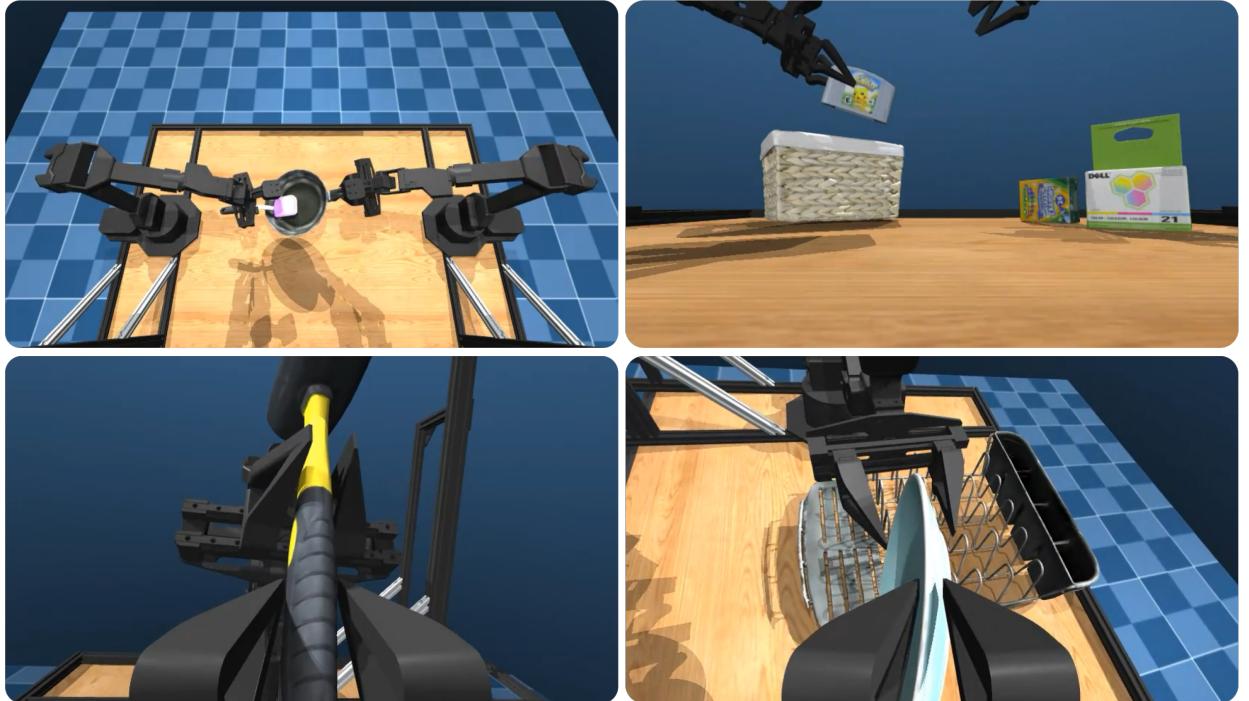


Figure 8 | *Teleoperating tasks in simulation.* We show the 4 camera views recorded during simulation collection runs using objects from the Google Scanned Objects Dataset ([Downs et al., 2022](#)).

The realistic model allows fast, intuitive, and scalable simulation data collection using an *ALOHA* 2 WidowX leader setup. We hope that an open source, high quality model with system identification can enable sharing of teleoperated simulation data across institutions and accelerate research in policy learning in simulation.

Conclusion

We present a low-cost system for bimanual teleoperation, enhancing performance, user-friendliness, and robustness compared to the previous *ALOHA* system. We make concrete improvements to hardware such as grippers, gravity compensation, frame, and cameras, while also providing a high quality simulation model. We hope *ALOHA* 2 will enable large scale data collection for fine-grained bimanual manipulation to advance research in robot learning.

References

- L. Downs, A. Francis, N. Koenig, B. Kinman, R. Hickman, K. Reymann, T. B. McHugh, and V. Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items, 2022. URL <https://arxiv.org/abs/2204.11918>.
- L. Keselman, J. Iselin Woodfill, A. Grunnet-Jepsen, and A. Bhowmik. Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1–10, 2017.
- S. Macenski, T. Foote, B. Gerkey, C. Lalancette, and W. Woodall. Robot operating system 2: Design, architecture, and uses in the wild. *Science Robotics*, 7(66):eabm6074, 2022. doi: 10.1126/scirobotics.abm6074. URL <https://www.science.org/doi/abs/10.1126/scirobotics.abm6074>.
- E. Todorov, T. Erez, and Y. Tassa. Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033. IEEE, 2012. doi: 10.1109/IROS.2012.6386109.
- Trossen Robotics. Viperx 300 robot arm 6dof, a. URL <https://www.trossenrobotics.com/viperx-300-robot-arm-6dof.aspx>. Accessed: 2024-01-24.
- Trossen Robotics. Widowx 250 robot arm 6dof, b. URL <https://www.trossenrobotics.com/widowx-250-robot-arm-6dof.aspx>. Accessed: 2024-01-24.
- K. Zakka, Y. Tassa, and MuJoCo Menagerie Contributors. MuJoCo Menagerie: A collection of high-quality simulation models for MuJoCo, 2022. URL http://github.com/google-deepmind/mujoco_menagerie.
- T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware, 2023.

Author Contributions

Please cite this work as “ALOHA 2 Team (2024)”.

- **Core Team (core contributors; leading the project effort; leading design, implementation, and/or research on the platform):** Travis Armstrong, Chelsea Finn, Pete Florence, Spencer Goodrich, Thinh Nguyen, Jonathan Tompson, Ayzaan Wahid, and Tony Zhao.
- **Hardware (working on hardware design and manufacturing; assembly of the system):** Jorge Aldaco, Kenneth Draper, Pete Florence, Spencer Goodrich, Torr Hage, Thinh Nguyen, Jonathan Tompson, Ayzaan Wahid, and Tony Zhao.
- **Software (software systems to run teleoperation and models; DevOps):** Jeff Bingham, Sankh Chan, Debidatta Dwibedi, Pete Florence, Spencer Goodrich, Wayne Gramlich, Alexander Herzog, Ian Storz, Jonathan Tompson, Sichun Xu, Ayzaan Wahid, Ted Wahrburg, Sergey Yaroshenko, and Tony Zhao.
- **Data (software and infrastructure to collect and process data):** Robert Baruch, Pete Florence, Jonathan Hoech, Ian Storz, Ayzaan Wahid, and Sergey Yaroshenko.
- **Simulation (working on creating and improving the simulation model):** Baruch Tabanpour, Ayzaan Wahid, and Kevin Zakka.
- **HRI / User studies (working on HRI, ergonomics, and conducting user studies):** Travis Armstrong, Spencer Goodrich, Leila Takayama, and Jonathan Tompson.

Acknowledgements

We thank Tom Erez, Matthew Mounteer, Francesco Romano, Stefano Saliceti, and Yuval Tassa for help with the simulation model. We thank Tomas Jackson for photography of the *ALOHA 2* fleet. We thank Corey Lynch for help with software setup and ROS2 integration. We thank Chikezie Ejiasi for creating character illustrations used in figures. We would also like to thank members of the wider Google DeepMind Robotics team for their support.