Science Robotics
AAAS

Supplementary Materials for

Learning contact-rich whole-body manipulation with example-guided reinforcement learning

Jose A. Barreiros *et al.*

Corresponding author: Alex Alspach, alex.alspach@tri.global

Sci. Robot. **10**, eads6790 (2025)
DOI: 10.1126/scirobotics.ads6790

This PDF file includes:

Eigengrasp primitives

We defined five whole-body primitives: enveloping flexion, pinch flexion, torsion, base rotation, and shoulder rotation. For the left arm, these primitives are given by (i) $\Delta\mathbf{q}_{\text{enveloping flexion}} \triangleq [0.5, 0, 0, -1.5, 0, 0, 0]$, (ii) $\Delta\mathbf{q}_{\text{pinch flexion}} \triangleq [0, 0, 0, 0, 0, -1, 0]$, (iii) $\Delta\mathbf{q}_{\text{torsion}} \triangleq [0, -1, -1, 0, -0.5, 0, 0]$, (iv) $\Delta\mathbf{q}_{\text{base rotation}} \triangleq [-0.5, 0, 0, 0, 0, 0, 0]$, and (v) $\Delta\mathbf{q}_{\text{shoulder rotation}} \triangleq [0, 0, -1, 0, 0, 0, 0]$. We use symmetric primitives for the arms, namely, the negative of those for the right arm. The torsion and enveloping flexion for each arm is controlled by the up/down and left/right motions of each thumbstick, respectively. The shoulder rotation can be activated via the "X" and "B" buttons for the right arm and via the left and right arrow buttons for the left arm. Last, a pinch motion can be obtained by using the "LB" and "LT" buttons for the left arm and "RB" and "RT" buttons for the right arm. The activation for each primitive is a binary scalar, and the joint position command is calculated by integrating the primitive activations over time.

Teleoperators

The motion examples were collected by three human operators using a gamepad (Logitecth F310) while watching the real-time simulation on a standard computer screen. The operators had access to as much time as they desired to get familiar with the interface.

Punyo Research Platform Components and Dimensions

The Punyo whole-body manipulation research platform comprises two seven-degree-of-freedom Kinova Gen2 Jaco robot arms (*76*) outfitted with air-filled compliant contact pressure sensing chambers, visuotactile and pressure sensing end effectors, a compliant chest, and a 3D-printed head, built upon an 80/20 frame. The major dimensions of the system, including a shoulder-to-shoulder distance of 540 mm and sensor placement and dimensions, can be seen in fig. S4. The arm and sensor assemblies are outfitted with a fabric cover that offers protection to the sensors and adds a high friction grasping surface over the arms. The whole system is mounted at the origin to the rear-edge of a large table on which manipulation tasks are performed.
Arm sensors: Each of Punyo's two arms has 14 highly compliant air-filled contact pressure sensing modules (fig. S5) which are made of Polyvinyl Chloride (PVC). There are two types of air-filled sensors on the arms, not including a third type found at the end effector which is described in the following subsection. On each upper arm and forearm are six rings that encircle the internal robot arm. Each ring comprises two separate air-filled chambers, each making up half of the sensing ring. The two most proximal rings are divided into "top and bottom" sections, whereas the next four proximal rings are divided into "inner and outer" sections. The sensors are mounted to the robot using VHB tape with lightweight 3D printed structures underneath to provide a more circular underlying cross-section where the robot arm becomes narrow.

Each of these 12 separate air-filled chambers on each arm (24 total) are connected via an air tube (Polyurethane with a 4 mm outer diameter and a 2.4 mm inner diameter, McMaster-Carr 50315K68) to a pressure sensor (Honeywell MPRLS0025PA00001A). To minimize air tubes crossing over robot joints, the two most proximal rings (four chambers) on each upper-arm are connected to a pressure sensor bank on the shoulder, whereas the two distal upper-arm rings (four chambers) on each arm are connected to a pressure sensor bank on the elbow. All other chambers on each arm (forearm, wrist, and paw) are connected to the pressure sensor bank on the elbow.

These bifurcated sensing rings are constructed using a 0.3 mm-thick PVC sheet, molded to form, then heat-sealed at the edges to create airtight chambers. Each chamber has an inflation port similar to those found on commercial inflatables (like pool toys), as well as hook-and-loop fastening features for attaching two half-rings together around the arm. To connect the air tube to the inflation ports, a press-fit barbed tube reducer is used (McMaster-Carr 5116K199).

The sensing chambers at the robot wrists differ from those specified for the upper arms and forearms. Single chambers (*77*) span the wrist joint and move with the two adjacent links when the underlying joint moves. This single chamber is connected via the same tube reducer and air tubing to the pressure sensor bank at the elbow.

For manual service and inflation, all chamber-to-pressure-sensor air tubes include an inline wye-fitting (McMaster-Carr 5225K85) and an accessible push-to-connect shut-off valve (McMaster-Carr 1201N56) so that the chambers can be inflated or deflated without disconnecting any tubing.

For the experiments performed in this paper, most of the air-filled chambers are inflated to atmospheric pressure. This is achieved by slightly pressurizing (above atmospheric pressure) all chambers in contact with one another, e.g., the four on the upper arm, then simultaneously bleeding the air out until atmospheric pressure is reached in all four, after which the chambers are sealed. The exceptions are those chambers nearest the inner side of a joint, e.g., the inner elbow and armpit. For these inner-joints, the adjacent chambers are slightly pressurized, then simultaneously left vented to atmosphere as the joint is manually jogged towards its inner mechanical limit so that excess air is passively released. The chambers are then sealed. This method ensures that these joint-adjacent chambers maintain a maximum amount of air volume for sensing while not impeding the arm's range of motion.

Visuotactile end effectors

Punyo's end effectors, or "paws," each comprise one angled-inward Soft-Bubble module (*69*, *78*), which is a time-of-flight-camera-based visuotactile sensor. The inflated portion of this sensor is made of a 0.6 mm-thick, high-friction latex sheet. This air-filled end-effector chamber (one on each arm) is connected via air tube to the pressure sensor bank at the elbow and is generally inflated to a height of 35 mm. Note that the visuotactile information of the sensor is not used in this study.

Tube and wire routing

As described above, pressure sensor banks exist at both the shoulders and elbows of the Punyo system. This placement reduces overall the number of air tubes crossing the robot joints. However, USB connections to microcontrollers in each pressure sensing bank are still required to route all the way down the arms to a central computer. Further, the time-of-flight sensor, though unused in these experiments, also requires a USB connection. In total, each arm has three USB cables that route down the arms, underneath the pressure sensing chambers. There are various locations along the links where strain relief is employed.

The stock Kinova Gen2 robots have multiple joints that can rotate continuously. To reduce the need for excessive service loops for air tubes and USB cables, the range of motion of these joints was limited using mechanical features to ±172.5 degrees.

Compliant chest

The compliant chest comprises a tall, curved central structure and covers for each of the unsensorized base links of the Kinova Gen2 robot arms. The underlying structural forms are 3D

printed and laser cut, then covered in passively compliant, adhesive-backed ½ inch Polyurethane foam (McMaster-Carr 8643K462).

Fabric coverings

The arms and chest are covered in fabric, both for protection as well as adjustable friction properties. The chest is covered in tailored spandex (*79*) span with features added to reflect our Punyo robot concept artwork (*80*). The chest cover drapes around the front of the robot and is fastened and cinched in the rear.

The arms are covered in the same spandex, so the cover can stretch and comply as the underlying robot moves. Sewn on top of the spandex surrounding each sensing-chamber region is a protective layer of cut-resistant Cut-Tex® PRO (*81*). Again, on top of each sensing region, since these regions are intended to make contact for manipulating objects, a third layer of high-friction silicone banding is incorporated (*82*). This same silicone banding is used on the inside of the cover as well to prevent the cover from sliding and shifting relative to the pressure-sensing chambers underneath.

For servicing the air-filled chambers, particularly inflating or deflating, hook-and-loop fastened access windows are provided at the pressure sensing banks where the push-to-connect inflation valves reside. Also, for servicing the sensors and underlying robot, a zipper (*83*) is installed down the length of the cover.

Head

Currently, the head serves as an aesthetic representation of our concept robot, as well as a placeholde,r preventing manipulation strategies from taking advantage of the typically inaccessible headspace region above the chest. As of yet, there is no sensing or actuation in the head. Punyo's head was printed on a Stratasys Polyjet 3D printer using Vero White and Vero Black materials. The assembly was broken down into parts so that all visible surfaces could be printed with a high-gloss finish, i.e., facing up and not in contact with support material. The head is bolted to the chest frame, and the faceplate is adhered only with magnets for easy access to the head-to-chest mounting fasteners within.

Improvements on the Punyo-1 Hardware Platform

The improved system subdivides the previous Punyo-1 bubble chambers into multiple smaller bubble chambers. The most proximal three sensorized links now contain four bubble chambers where there was originally only one pressure-sensing chamber on each link. The next distal bubble is slightly larger to span flexibility across wrist joints, which is the same as Punyo-1. This system also employs a single TRI Soft-Bubble (*69*) as the end effector for each arm. The sub-divided bubbles were created to increase spatial resolution for contact sensing. The bubble chambers are also split laterally into opposing sides of the arms, allowing us to infer a rough direction of contact when external contacts are made.

Covering the inflated sensors is a protective, high-friction fabric sleeve. The system also uses a similar cut-proof fabric to Punyo-1 on these outer covers, which we call sweaters. We also introduced patches of grippy material that is sewed around the outer surfaces to increase the friction when contact is made. Additionally, this high-friction material is added to the inner surface of the sweaters to minimize shifting of the sweater with respect to the robot sensor links underneath.

Floatie characterization

We performed a compression test of a floatie inflated to nominal pressure (~1 atm) 10 times using a tensile test machine (Mark-10) at steps of approximately 0.06 mm increments. The floatie was placed on a rigid plastic cylinder to mimic the robot link. Measurements were taken using a 500 N (100 LBF) force gauge (Mark-10 model M7-100). A circular metal plate of 5 cm in diameter (area of approximately 20.3 cm^2) was used as the indenter. Fig. S6 shows the results of the experiment.

A second experiment was conducted to evaluate the steady-state pressure drift over a period of 48 hours. For this, the robot was placed with its arms up and under no external disturbances. Fig. S7 shows the pressures after the sensors were zeroed by the first reading and normalized by the room's ambient pressure $p_{ij} = p_{measured_{ij}} - p_{ambient\_room_i} - p_{0j}$, where p_{ij}: pressure of floatie j at time i, p_{0j}: pressure of floatie j at time 0. An average steady-state pressure drift (~ 4.07 hPa) occurs when room temperature change causes thermal expansion of the floatie material, PVC ($\propto = 5 \times 10^{-5} mm/(mm\ ^{\circ}\mathrm{C})$) (*84*).

Comparison between GQDP and teleoperation example motions

To compare example motions generated by GQDP and teleoperation, we collected two demonstrations for the cardboard-box-pivot-and-lift task. We first generated the GQDP demonstration then asked the teleoperator to closely mimic its motion to ensure a meaningful comparison. We further trained policies with these demonstrations. Fig. S8 shows the success rate of the policies during training for each type of example motion averaged across three different random seeds. The GQDP example motion resulted in a higher success rate. We hypothesize that this is due to the GQDP demo including more meaningful state transitions than the teleoperation example. The teleoperation demonstration includes pauses between actions that often occur as the teleoperator figures out the next move, and mimicking these non-essential state transitions could potentially mislead the agent during training. In a sense, GQDP demos contain close to optimal states (from the point of view of the dynamics), whereas teleoperation includes what the teleoperator can or wants to demonstrate, including style. It is noteworthy that due to the disparity of the demonstration lengths (250 timesteps for GQDP and 500 timesteps for teleoperation), no conclusion can be made in terms of sample efficiency.

TABLE S1: Object properties.

Object	Bounding Box Dimensions (m)	Weight (kg)
water jug	0.5 x 0.27 x 0.27	0.80
plastic box	0.57 x 0.33 x 0.40	1.80
cardboard box	0.41 x 0.26 x 0.32	0.54
plastic cube	0.15 x 0.15 x 0.15	0.55

TABLE S2: Observations of the Discriminator, the PPO Policy, and the PPO Critic per task.

Task	Discriminator Observation	PPO Policy Observation	PPO Critic Observation
cube-lift	joint positions	joint positions, object 6D pose, previous action	joint positions, object pose, previous action
plastic-box-paw-lift	joint positions, end effector pose	joint positions, object 6D pose, previous action	joint positions, object pose, previous action
plastic-box-hug-lift	joint positions	joint positions, object 6D pose, previous action	joint positions, object pose, previous action
cardboard-box-pivot-and-lift	joint positions	joint positions, object 6D pose, previous action	joint positions, object pose, previous action
jug-over-shoulder	joint positions	joint positions, object 6D pose, previous action	joint positions, object pose, previous action
jug-rotate	joint positions	joint positions, object 6D pose, previous action	joint positions, object pose, previous action

TABLE S3: Reward function weights.

Weights	Value
w_{kp}	0.1
w_a	-10^{-5}
w_{da}	-10^{-5}
w_τ	-10^{-5}
w_{lin}	-10^{-4}
w_{rot}	-10^{-4}
w_{term}	-1
w_{succ}	1

TABLE S4: Success criteria for each task.

Task	Condition
cube-lift	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.4\ rad/s$
plastic-box-paw-lift	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.3\ rad/s$
plastic-box-hug-lift	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.3\ rad/s$
cardboard-box-pivot-and-lift	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.3\ rad/s$
jug-over-shoulder	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.3\ rad/s$
jug-rotate	$\|d_{trans}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.05\ m$ and $\|d_{rot}(\mathbf{q}^u, \mathbf{q}^u_{goal})\| \leq 0.2\ rad$ and $\|\dot{q}_{trans}\| \leq 0.15\ m/s$ and $\|\dot{q}_{rot}\| \leq 0.3\ rad/s$

TABLE S5: Workspace limits per task.

Task	Workspace Bounds
cube-lift	x: [0.3, 0.9], y: [-0.3, 0.3]
plastic-box-paw-lift	x: [0.3, 0.9], y: [-0.3, 0.3]
plastic-box-hug-lift	x: [0.2, 0.9], y: [-0.3, 0.3]
cardboard-box-pivot-and-lift	x: [0.0, 0.9], y: [-0.4, 0.4]
jug-over-shoulder	x: [0.05, 0.9], y: [-0.4, 0.4]
jug-rotate	x: [0.1, 0.9], y: [-0.3, 0.3]

TABLE S6: Network structure.

Network	Type	Hidden Layers	Activation
policy	MLP	[1024, 512]	Rectified Linear Unit (ReLU)
value function	MLP	[1024, 512]	Rectified Linear Unit (ReLU)
discriminator	MLP	[256, 128, 64]	Rectified Linear Unit (ReLU)

TABLE S7: Training parameters.

Parameter	Value
γ	0.99
τ	0.95
λ	0.7 (0.4 for the jug-rotate task)
σ	0
parallel training environments	4096
sample per update iteration	4096
batch size	512
learning rate	5e-5
KL divergence target	8e-3
clip range	0.2
horizon length	64
discriminator weight decay	1e-4
discriminator gradient penalty	10

TABLE S8: Domain randomization parameters.

	Parameter	Type	Distribution
Object	mass	scaling	$\mathcal{U}(0.9, 1.1)$
	friction	scaling	$\mathcal{U}(0.8, 1.2)$
	scale	scaling	$\mathcal{U}(0.95, 1.05)$
	initial x position	additive	$\mathcal{U}(-0.05, 0.05)$
	initial y position	additive	$\mathcal{U}(-0.05, 0.05)$
	initial yaw	additive	$\mathcal{U}(-0.05, 0.05)$
Simulation	gravity	additive	$\mathcal{N}(0, 0.1)$
Robot	friction	scaling	$\mathcal{U}(0.8, 1.2)$
	joint stiffness	scaling	$\mathcal{U}(0.8, 1.2)$
	joint damping	scaling	$\mathcal{U}(0.8, 1.2)$
	action	additive	$\mathcal{N}(0, 0.02)$

TABLE S9: Initial and goal object poses per task.

Task	Object's Initial Pose (x, y, z, roll, pitch, yaw)	Object's Goal Pose (x, y, z, roll, pitch, yaw)
cube-lift	0.65, 0.0, 0.075, 0.0, 0.0, 0.0	0.65, 0.0, 0.4, 0.0, 0.0, 0.0
plastic-box-paw-lift	0.5, 0.0, 0.17, 0.0, 0.0, $\pi/2$	0.5, 0.0, 0.17, 0.0, 0.0, $\pi/2$
plastic-box-hug-lift	0.5, 0.0, 0.17, 0.0, 0.0, $\pi/2$	0.285, 0.0, 0.3, 0.0, 0.0, $\pi/2$
cardboard-box-pivot-and-lift	0.35, 0.0, 0.131, 0.0, 0.0, 0.0	0.25, 0.0, 0.35, 0.0, $-\pi/2$, 0.0
jug-over-shoulder	0.35, 0.0, 0.25, 0.0, 0.0, 0.0	0.13, 0.3, 0.64, 0.0, $-\pi/2$, 0.0
jug-rotate	0.35, 0.0, 0.25, 0.0, 0.0, 0.0	0.2, 0.0, 0.35, $-\pi$, 0.0, 0.0

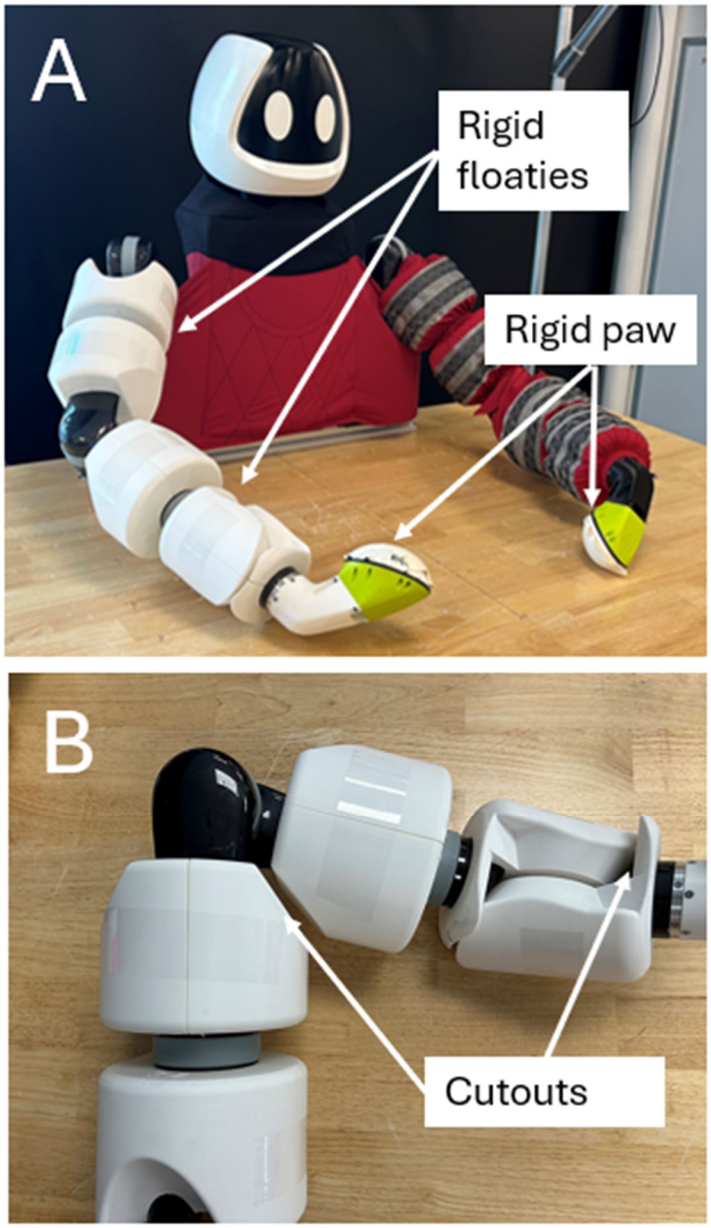


Fig. S1. Pictures showing the (A) rigid air-filled chamber stand-ins and paws used for compliance experiments. (B) close-up picture of the rigid air-filled chamber stand-ins showing the cutouts added to avoid reducing the range of motion due to self-collisions compared to the soft air-filled chambers.

Fig. S2. Picture of a Punyo sweater.

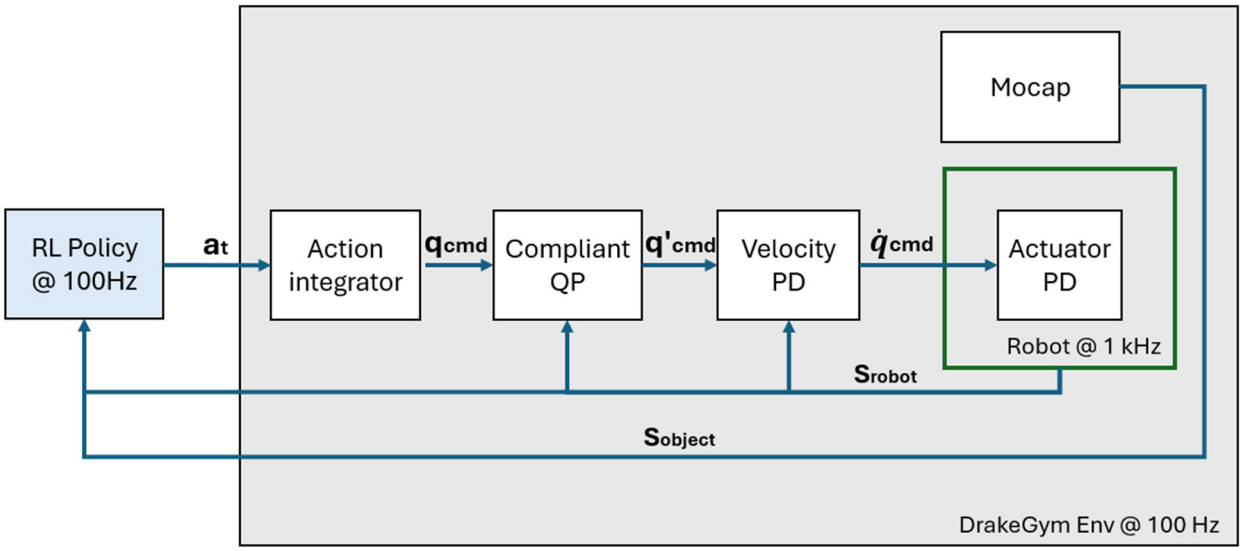


Fig. S3. Diagram showing the control stack.

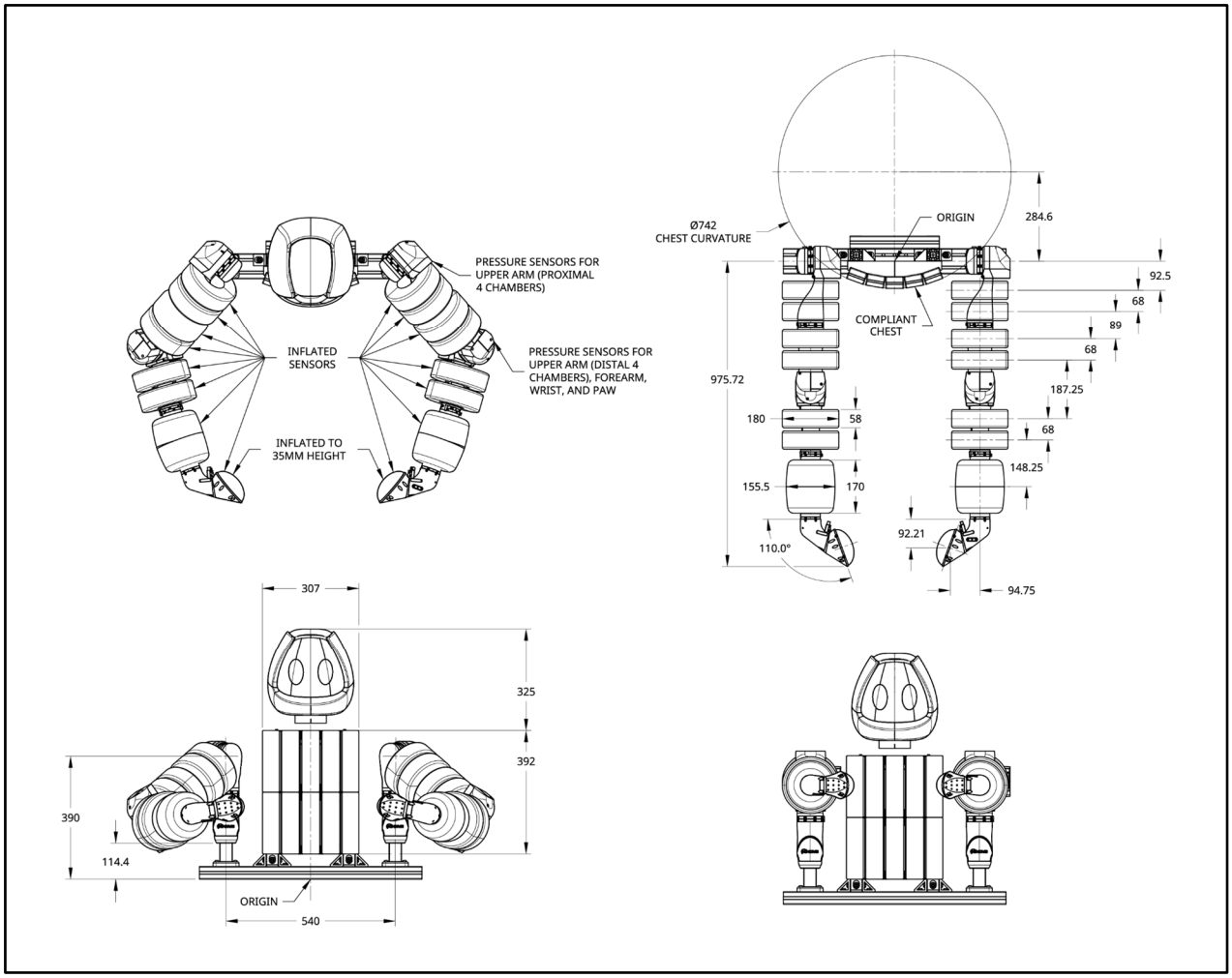


Fig. S4. A diagram of Punyo showing its parts and dimensions.

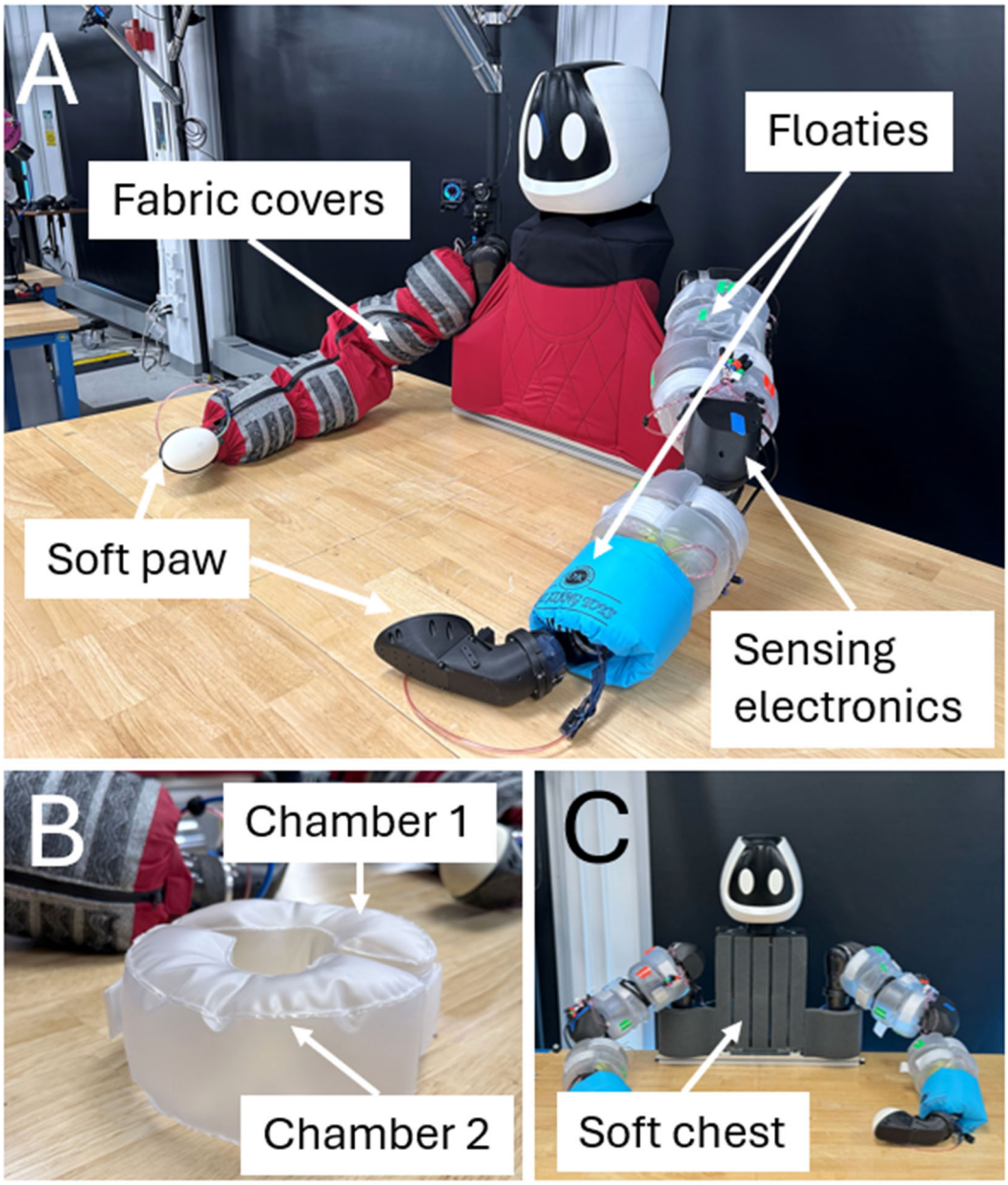


Fig. S5. (A) Picture of the Punyo robot showing the uncovered left arm and covered right arm for comparison. (B) Close-up of a floatie. (C) Uncovered soft chest.

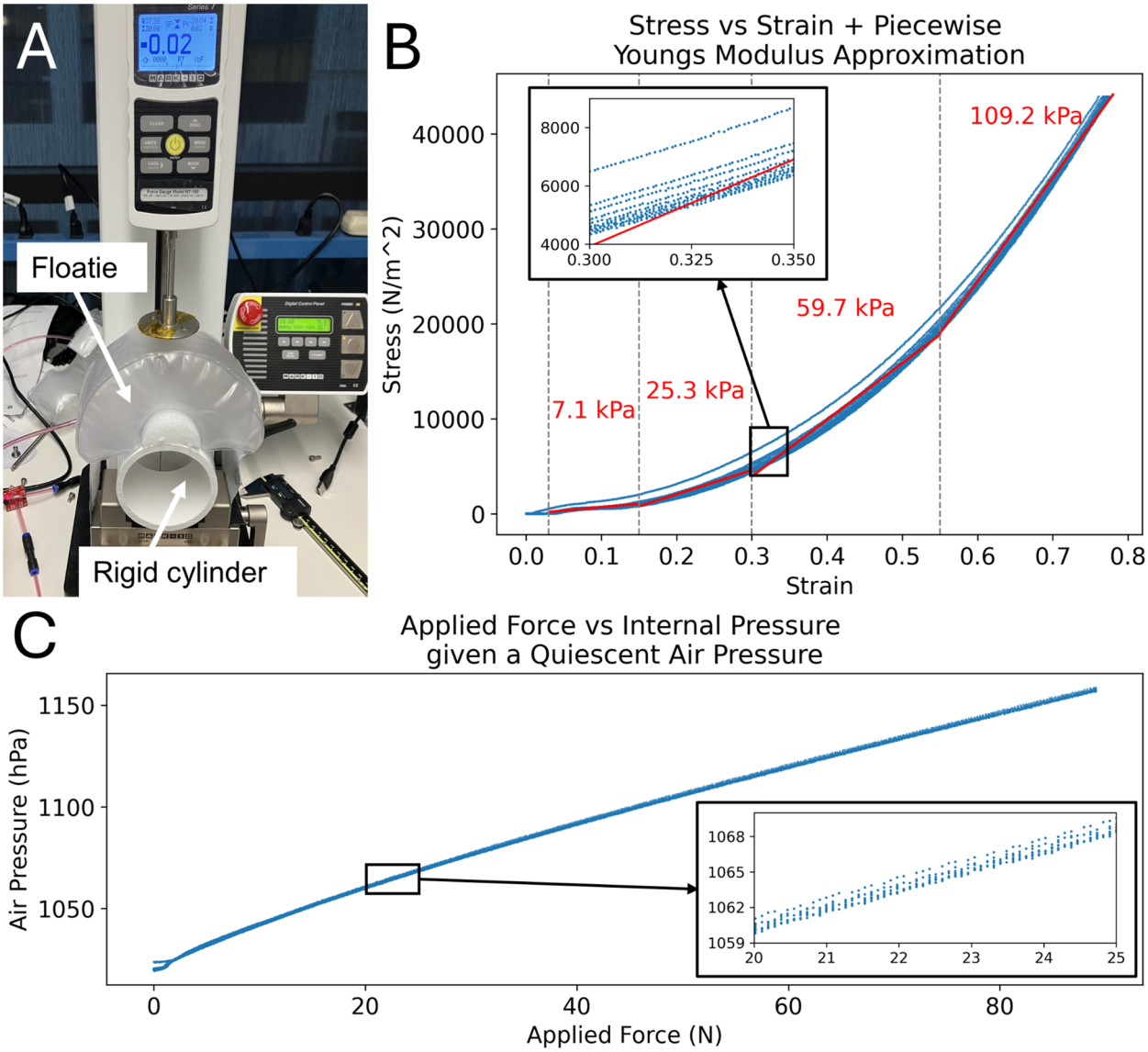
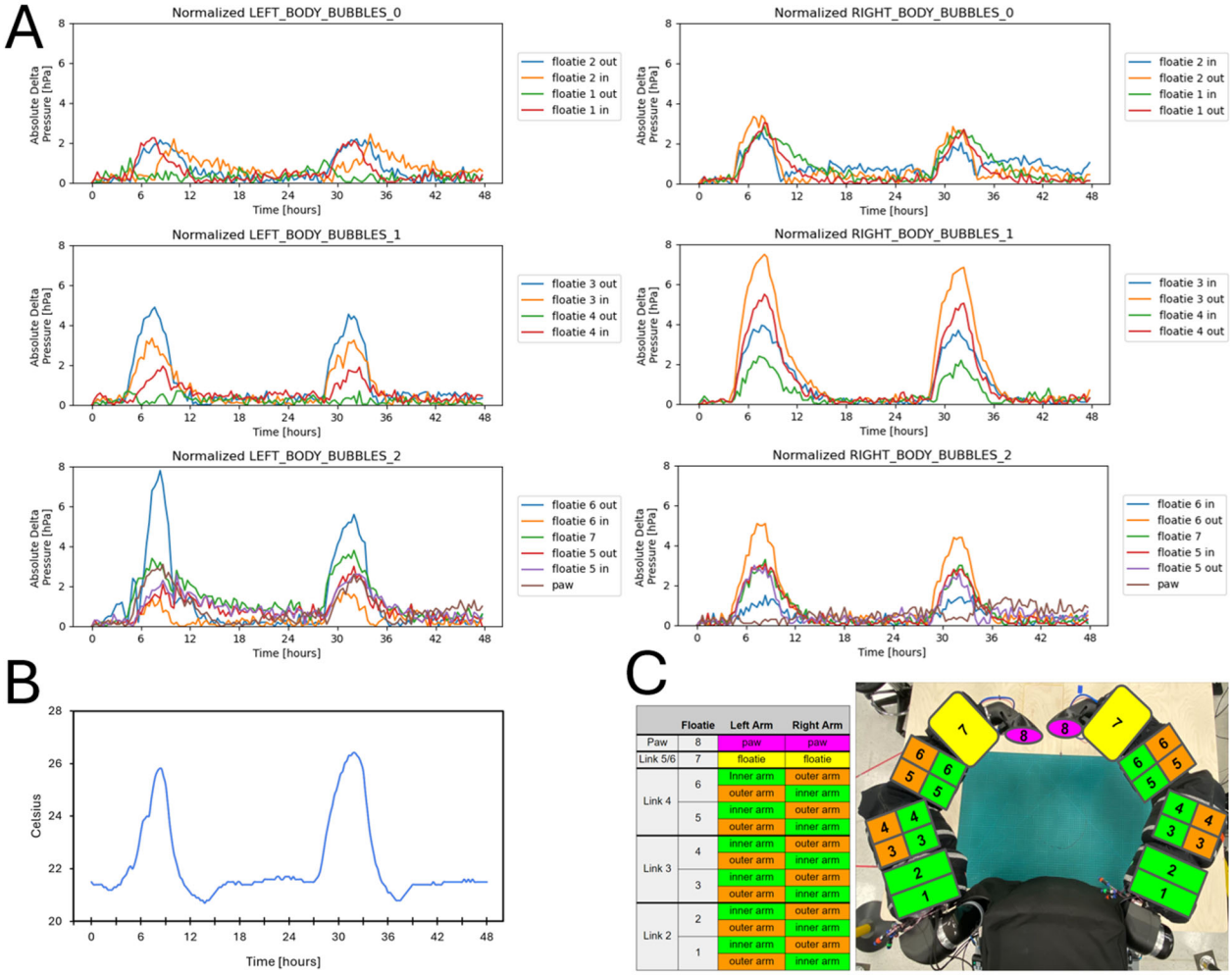


Fig. S6. Compression test conducted on a floatie sensor over ten cycles. (A) Experimental setup, (B) strain-stress material characterization, and (C) pressure-force sensor characterization.



	Floatie	Left Arm	Right Arm
Paw	8	paw	paw
Link 5/6	7	floatie	floatie
Link 4	6	inner arm	outer arm
		outer arm	inner arm
	5	inner arm	outer arm
		outer arm	inner arm
Link 3	4	inner arm	outer arm
		outer arm	inner arm
	3	inner arm	outer arm
		outer arm	inner arm
Link 2	2	inner arm	outer arm
		outer arm	inner arm
	1	inner arm	outer arm
		outer arm	inner arm

Fig. S7. Steady-state characterization. A) Normalized pressure of the floatie sensors under no external disturbances (contact) over a period of 48 hours. B) Room temperature over the same period of time. C) Mappings of sensor names and location of floaties in the robot's body.

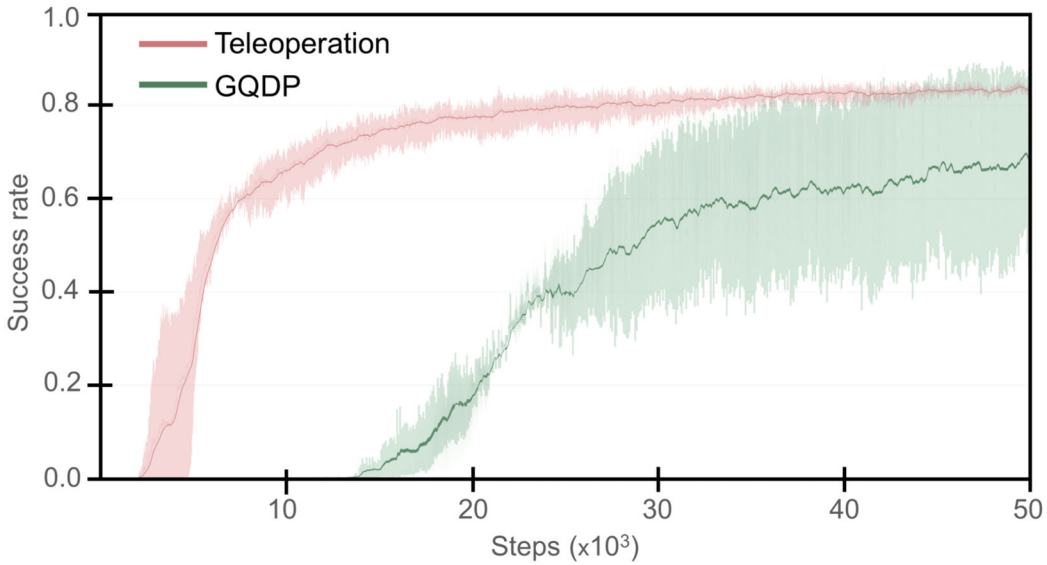


Fig. S8. Training curves, averaged across three seeds, for policies generated using teleoperation (red line) and GQDP (green line) example motions. Shaded regions indicate standard deviation across three seeds (±1 STD, n = 3).