bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Whole-body simulation of realistic fruit fly locomotion with deep reinforcement learning

Roman Vaxenburg,[1] Igor Siwanowicz,[1] Josh Merel,[2,§] Alice A. Robie,[1]
Carmen Morrow,[1] Guido Novati,[3] Zinovia Stefanidi,[1,4]
Gwyneth M. Card,[1,5] Michael B. Reiser,[1] Matthew M. Botvinick,[3,6] Kristin M. Branson,[1]
Yuval Tassa,[3†] Srinivas C. Turaga [1†]

[1]HHMI Janelia Research Campus, Ashburn VA USA
[2]Fauna Robotics, New York City NY USA
[3]Google DeepMind, London UK
[4]Tubingen University, Tubingen, Germany
[5]Columbia University, New York City NY USA
[6]Gatsby Computational Neuroscience Unit, University College London, London UK
[§]Primary contributions while at Google DeepMind
[†]correspondence: tassa@google.com, turagas@janelia.hhmi.org.

March 12, 2024

Abstract

The body of an animal determines how the nervous system produces behavior. Therefore, detailed modeling of the neural control of sensorimotor behavior requires a detailed model of the body. Here we contribute an anatomically-detailed biomechanical whole-body model of the fruit fly *Drosophila melanogaster* in the MuJoCo physics engine. Our model is general-purpose, enabling the simulation of diverse fly behaviors, both on land and in the air. We demonstrate the generality of our model by simulating realistic locomotion, both flight and walking. To support these behaviors, we have extended MuJoCo with phenomenological models of fluid forces and adhesion forces. Through data-driven end-to-end reinforcement learning, we demonstrate that these advances enable the training of neural network controllers capable of realistic locomotion along complex trajectories based on high-level steering control signals. With a visually guided flight task, we demonstrate a neural controller that can use the vision sensors of the body model to control and steer flight. Our project is an open-source platform for modeling neural control of sensorimotor behavior in an embodied context.

Introduction

Animal behavior is the product of sensorimotor feedback control loops involving the animal's brain, its body, and its environment [Merel et al., 2019, Dickinson et al., 2000, Gomez-Marin and Ghazanfar, 2019, Chiel and Beer, 1997]. The body dictates precisely how motor commands from the nervous system are translated into action, and how the consequences of motor actions are reported via sensory feedback. Thus a detailed biomechanical understanding of the body work is essential to understanding the neural control of movement. Here, we provide a new framework for physics simulation of an anatomically detailed biomechanical model of the fruit fly body. We

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

validate the generality of our model by demonstrating realistic simulation of locomotion – both flight and walking – using reinforcement learning. Our general-purpose simulation framework is a platform for future brain-body modeling of diverse fruit fly behaviors.

We developed our fly body model in the open source MuJoCo [Todorov et al., 2012] physics engine. We used high-resolution confocal microscopy to image and construct a model of a female fruit fly (Fig. 1). To enable accurate modeling of fly behaviors such as flight and walking, we added new features to the MuJoCo physics engine. First, we developed a new computationally efficient, phenomenological fluid model to simulate forces resulting from fly wings flapping in air. Second, we developed adhesion actuators to model forces generated by insect feet gripping a surface. We validate the body model and the physics simulation by demonstrating realistic locomotion — both flight (Fig. 2) and walking (Fig. 3) — along complex natural trajectories with closed-loop sensorimotor control. In particular, by using end-to-end reinforcement learning (RL) to train a closed-loop sensorimotor controller to imitate the locomotor patterns of real flies, we validate the physics modeling by showing that we can replicate the body trajectories of real flies with similar leg and wing kinematics. We trained "steerable" low-level closed-loop sensorimotor controllers [Merel et al., 2019] for walking and flight which can drive the fly body to generate complex trajectories using only high-level steering commands. And finally, we demonstrate the use of end-to-end reinforcement learning to train the model fly to perform visually guided flight tasks (Fig. 4).

Our biomechanical model and physics simulation builds on previous work on physics simulation of fruit fly behavior [Dickson et al., 2008, Lobato-Rios et al., 2022, Wang-Chen et al., 2023, Reiser et al., 2004] with qualitative improvements in the realism of the anatomy and biomechanics, physics, and behavior, leading to realistic simulation of both walking and flight in a single detailed general-purpose body model. The pioneering Grand Unified Fly [Dickson et al., 2008] demonstrated sensorimotor closed-loop visually guided flight behavior, with a reduced body model and a hand-designed controller. More recently, NeuroMechFly [Lobato-Rios et al., 2022, Wang-Chen et al., 2023] demonstrated closed-loop visually guided walking, with an anatomically realistic biomechanical body model and a hybrid hand-designed controller with learned components. In this work, we developed a physics environment capable of both flight and walking with a single general-purpose, anatomically realistic body model. We used high-speed video based kinematic tracking [Branson et al., 2009, Pereira et al., 2022, Gosztolai et al., 2021, Karashchuk et al., 2021] to measure natural locomotion [Muijres et al., 2014, Muijres et al., 2015], and trained closed-loop controllers capable of imitating arbitrarily complex natural trajectories using end-to-end reinforcement learning. As we demonstrate through inverse kinematics, our body model is also capable of a rich repertoire of fly behaviors beyond locomotion, including grooming.

Our work lays the foundation for studying the neural basis of sensorimotor behaviors in the adult fruit fly. Our general-purpose and open-source model can be extended in the future to increase the realism of the biomechanics, neural control, and behavior as new data becomes available. For instance, new imaging of musculature and sensory systems can improve biomechanical fidelity of actuation and sensory feedback. The black-box artificial neural network controllers we trained using imitation learning can be made more realistic by introducing connectome constraints [Lappalainen et al., 2023] based on recent measurements of connectivity of the brain [Dorkenwald et al., 2023, Schlegel et al., 2023] and ventral nerve cord [Lesser et al., 2023, Azevedo et al., 2022, Cheong et al., 2023, Marin et al., 2023, Takemura et al., 2023], as well as neural activity [Turner et al., 2022, Pacheco et al., 2021, Aimon et al., 2023]. Thus our present work serves as a platform to enable future studies of embodied cognition [Zador et al., 2023], modeling the neural and biomechanical basis of sensorimotor behavior in unprecedented detail. Our fly model is available at https://github.com/TuragaLab/flybody.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Results

Body model geometry

We used confocal fluorescence microscopy to image the entire adult female fly body with high resolution (Fig. 1a, Methods, Supp. Data). Fluorescence microscopy enabled the staining of chitin to easily segment the shapes of body segments and identify the pivot points of all joints (Fig. 1b). Aberration-free high-resolution imaging of the entire body required disassembling it into smaller parts, chemical elimination of soft tissue, and bleaching of pigmentation (Methods). The dataset also enables the identification of origin and insertion sites of muscles, locations of the proprioceptive hair plates of the neck, coxae, trochanters, wing base and halteres — information that can be incorporated into future versions of the model.

We manually segmented 67 body components using Fiji [Schindelin et al., 2012], then simplified the corresponding meshes in Blender [Community, 2018], reducing the number of vertices by a factor of 1000, to allow efficient computational modeling while preserving morphological features relevant for our biomechanical study (Fig. 1b,c). In Blender, we assembled component meshes into a body by connecting them at 66 identified joint locations (Fig. 1b,d), amounting to 102 degrees of freedom (DoFs) determined in agreement with the literature (Fig. 1k). We modeled each biological joint as either a single hinge joint (1 DoF) or a combination of two or three hinge joints (2 and 3 DoFs, respectively). We estimated the joint angles corresponding to a resting fly pose (Fig. 1d,e) by visual inspection of videography. We note that these model joints are simplified approximations, particularly for the neck joint, wing hinge, and thorax-coxa articulation joint [Strausfeld et al., 1987, Melis et al., 2024, Gorko et al., 2024]. The fly model assembled in Blender defines the fly's geometry and kinematic tree, and is available at https://github.com/TuragaLab/flybody. We describe complementing the geometric Blender model with physics-related components next.

Specifying body model physics

We exported the geometrical Blender fly model to MuJoCo and added the components required for physics simulation. This involved several steps (Supp. Section A). First, we generated primitive "geom" representations of each body part (Fig. 1f,g), which are used for efficient physics simulation and collision detection. Second, we measured the masses of each body part (Methods, Table 1). MuJoCo used these masses, along with an assumption of uniform density within each body part, to estimate the body part moments of inertia. Third, we added actuators to drive all the joints. Torque actuators were used for the wing joints, and position actuators for the rest of the body parts (Methods, Table 3). Joint limits were determined by using inverse kinematics to match a diverse collection of poses from videography data, including grooming postures in which flies showcase their remarkable joint flexibility (Methods). Typically, each actuator applies force to a single DoF. However, in the multi-segmented tarsi and abdomen body parts, several DoFs are coupled by a tendon and actuated by a single actuator to produce coordinated bending (Fig. 1i). Insect tarsi can adhere to surfaces, allowing flies to walk on walls and ceilings [Arzt et al., 2003]. To model this, we introduced adhesion actuators to MuJoCo. The adhesion actuators can simulate both active (controlled) and passive (uncontrolled) forces, and can now be used more generally for other models as a new MuJoCo feature. We added adhesion actuators to the tarsal tips (Fig. 1f,h). We also added adhesion actuators to the labrum to enable modeling of feeding and courtship behaviors [McKellar et al., 2020]. Fourth, we equipped the model with a sensory system, comprising vision, vestibular, proprioceptive, and mechanoreception sensors, see Table 2 for details and Table 20 for correspondence between the sensory systems of our model and the real fly.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

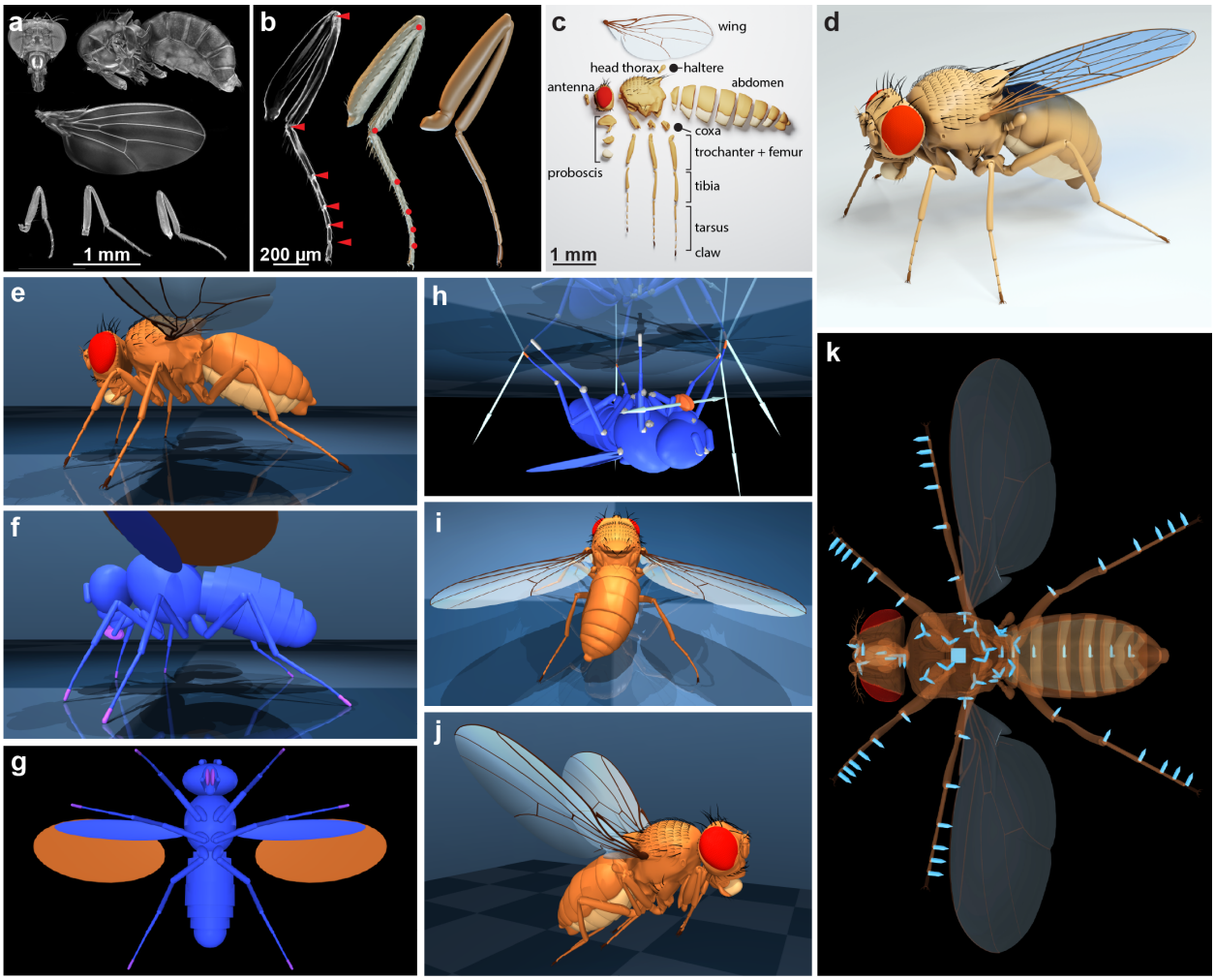


Figure 1: **Constructing a 3D model of a female fruit fly from confocal data.** (**a**) Compilation of 6 datasets representing a single fly; maximum intensity projections of confocal stacks showing head, thorax + abdomen, and legs. (**b**) A partial projection of the mid-leg confocal stack with the joints between femur-tibia and tarsal segments indicated by red triangles (left), a 3D mesh extracted from the stack (middle) and a low-polygon model of the leg (right). (**c**) An exploded simplified model (∼20k faces) showing body segments. (**d**) The complete anatomical model in the rest pose. (**e**) Side view of the body model in MuJoCo. (**f, g**) Side and bottom view of the geometric primitive (geom) approximation of body segments used to efficiently detect collisions and simulate physics. Blue: regular collision geoms. Mauve: Geoms that also have associated adhesion actuators. Ochre: Wing ellipsoids used by the advanced fluid model to enable flight. (**h**) Visualization of actuator forces generated when the model fly hangs upside down. The adhesion actuators of the front-right, middle-left, hind-right legs are actively gripping the ceiling (orange) and the labri (mouth) adhesors are also active; other actuators are inactive (white). The arrows, which visualize the contact forces, are proportional (and opposite) to the applied adhesion forces, minus gravity when applicable. (**i**) Exaggerated posture showing the result of activating the abdominal abduction and back-right tarsal flexion actuators. Abdominal joints and tarsal joints are each coupled with a single actuator (”tendon”) which simultaneously actuates multiple degrees of freedom (DoFs). (**j**) Model in a flight pose with legs retracted. (**k**) Bottom view, translucent visual geometry with light-blue arrows indicating joints. Cube: 6-DoF free joint (required for free center-of-mass motion in the simulator and is not a part of fly’s internal DoFs). Arrows: hinge joints; point in the direction of positive rotation. Groups of three hinge joints effectively form ball joints.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The result of these steps is an anatomically detailed, functional, biomechanical model of the entire fly body. All aspects of this model are programmatically modifiable. For instance, at simulation time, certain DoFs can be frozen (e.g., removing leg DoFs during flight), actuators can be programmatically enabled or disabled, body parts can be rescaled, etc. Similarly, sensors can also be enabled or disabled, and their properties (e.g., filtering in time) modified. The model is also extendable. For instance, the sensory system capabilities can be extended with new custom sensors (e.g., compound eye vision, olfaction).

Simulating body interactions with its environment

In addition to simulating the physics of the body, the MuJoCo physics engine also simulates the interactions of the body with its environment. These interactions include forces due to physical contacts, interactions with a surrounding fluid, e.g. air, as well as gravity. Contacts are modeled both between different body parts of the fly, and between the fly body and its environment. Most important for determining ground reaction forces during walking on a substrate. Fluid interactions consist of forces generated by interactions of the wings with air. Accurate simulation of fluid dynamics is a famously challenging computational problem. We instead developed a stateless quasi-steady state approximation of the fluid dynamics (Methods). Our phenomenological fluid model is a generalization of [Andersen et al., 2005] to three dimensions and it approximates the forces and torques exerted on ellipsoid-shaped bodies moving through an incompressible quiescent fluid. In effect, we approximated forces due to the fluid dynamics phenomena of added mass [Lamb, 1932, Tuckerman, 1925], viscous drag [Duan et al., 2015], viscous resistance [Stokes, 1850], Magnus lift [Seifert, 2012], and Kutta lift [Kutta, 1902]. This model ignores turbulence, vortices and other stateful-fluid phenomena which are known to contribute to insect flight [Sane, 2003]. We used this approximation to compute forces on the flapping wings, which have a slender ellipsoid shape to a good approximation (Fig. 1g). This fluid model is another new MuJoCo feature developed to facilitate our study, and can be used more generally for other models.

Simulation of behavior by the numerical integration of the passive and active dynamics of the fly model and its interactions with the environment is a considerable computational challenge, both in terms of the dimensionality of the system and the small integration time step required for numerically accurate simulation. Our fly body model is of considerable complexity (102 DoFs) and the simulation must model fast behaviors, such as ~ 200 Hz wing flapping during flight, requiring a small simulation time step, e.g. ~ 0.1 ms. On a single core of an Intel Xeon CPU E5-2697 v3, the simulation of 10 ms of full flight and walking environment loops (Fig. 2a, Fig. 3a) took 421.5 ms and 68.65 ms, respectively, which is fast enough to enable reinforcement learning of motor control. The fraction of time used by the MuJoCo simulator alone (i.e., excluding the rest of the environment loop components: policy network and python RL environment overheads) was 55.5 ms and 23.2 ms for flight and walking, respectively (Supp. Section C, Table 21).

Imitation learning of locomotion leads to realistic flight and walking

We used deep reinforcement learning to train our model fruit fly to generate specific behaviors. In particular, we trained an artificial neural network to serve the role of the nervous system, forming a sensorimotor controller driving locomotion in a closed loop. At each time step, MuJoCo simulates signals from the sensory system which are provided as input to the neural network. The neural network then computes control signals corresponding to each actuator in the body, which are then used by MuJoCo to simulate the active forces generated by the body for that time step.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We used imitation learning [Merel et al., 2020a, Hasenclever et al., 2020, Peng et al., 2018], a data-driven approach to train our neural-network controllers, to generate realistic locomotion. Here, the network was trained so that the trajectory of the center of mass (CoM) position and orientation of the body, as well as individual body parts, matched the corresponding trajectory of a real fly during free locomotion measured from video. To train networks which can generalize from training trajectories to locomotion along novel trajectories, we trained "steerable" low-level controllers [Merel et al., 2019]. These neural networks are analogous to the ventral nerve cord (VNC) of the fly central nervous system and are responsible for translating high-level (descending) command signals from the central brain into low-level motor control signals. The high-level commands in our model correspond to the desired change in the fly's (CoM) position and orientation (6 DoFs) at each time step, while the low-level motor control signals drive individual actuators in the body.

We trained two steerable neural network controllers (policy networks): one for flight (Fig. 2) and one for walking (Fig. 3). From training data in the form of real trajectories of the CoM and the body parts of walking and flying fruit flies (Methods), we used the CoM trajectory of the real fly as the high-level steering command to the network, and then constructed a reward function which rewarded the fly for matching both the target CoM as well as the locations of each body part at each time step. The reward is then maximized when the model fly tracks the CoM of the real fly, while moving its body parts in the same manner as the real fly. This form of imitation learning teaches the model fly to move its body along the same path as a real fly, using the same pattern of limb or wing movements as the real fly.

The steerable controller neural networks were feedforward multilayer perceptrons (MLP), which received egocentric vestibular and proprioceptive signals as input, in addition to the high-level steering commands. They were trained to generate motor control signals actuating the body in order to maximize the imitation reward using the DMPO agent, an off-policy model-free RL algorithm [Abdolmaleki et al., 2018b, Abdolmaleki et al., 2018a] with a distributional critic value function [Bellemare et al., 2017]. This is an actor-critic style algorithm which trains two networks: a policy network learns a nonlinear mapping from the sensory observations to the control signals, and a critic network which predicts the expected discounted cumulative reward resulting from applying a given control signal at a given state. We used the implementation of DMPO from the `acme` [Hoffman et al., 2020] reinforcement learning python library. Training until convergence can require the order of $\sim 10^9$ (walking) and $\sim 10^8$ (flight) simulation steps with $\sim 10^8$ and $\sim 10^7$ of policy network weight updates, respectively. To reduce the training time to a practical range of days or hours, we developed a multi-CPU and GPU parallelization scheme [Horgan et al., 2018] using `Ray` [Moritz et al., 2017], a general-purpose distributed computing package (Methods). Our code is general purpose and open source. Other RL algorithms can be easily paired with our model to train controllers in a modular fashion. We describe the specifics of the two controllers next.

Flight

We trained a steerable flight controller (Fig. 2a) with imitation learning using previously collected high-speed videography of freely flying *Drosophila hydei* performing spontaneous saccades [Muijres et al., 2015] and forced evasion maneuvers [Muijres et al., 2014]. These datasets combined contained 272 individual trajectories ($\sim$53 seconds of real time flight) of the CoM along with wing kinematics during turns, speed and altitude changes, flying straight, sideways and backward, and hovering in place (Methods). While *Drosophila hydei* is a somewhat larger fly compared to *Drosophila melanogaster*, we expect that the body and wing kinematics are similar [Dickinson and

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Muijres, 2016], and nevertheless, these are the best available data. A single controller network was trained to imitate all 216 flight trajectories across the (train) dataset. This steerable controller could maintain stable flight and drive the fly to traverse new flight trajectories (Fig. 2).

The design of our flight controller policy network was inspired by the finding that flies use only subtle deviations from a nominal wing beat pattern to control their flight trajectories [Muijres et al., 2014, Muijres et al., 2015]. Our flight controller consisted of two parts: a fixed wing beat pattern generator (WPG) and a trainable fully-connected MLP (Fig. 2a). The WPG produced, by design, with no learning involved, a mirror-symmetric periodic baseline wing beat pattern (Methods), based on wing trajectories recorded from a hovering *Drosophila melanogaster* [Fry et al., 2005, Dickson et al., 2008] (Fig. 2b). The policy neural network could control the frequency of the WPG as well as small wing motion deviations from the WPG baseline pattern, that resulted in the full repertoire of flight behaviors from the training datasets. As the WPG's baseline pattern was already close to the required wing motion solution, in effect the WPG also provided a good initialization which sped up training significantly.

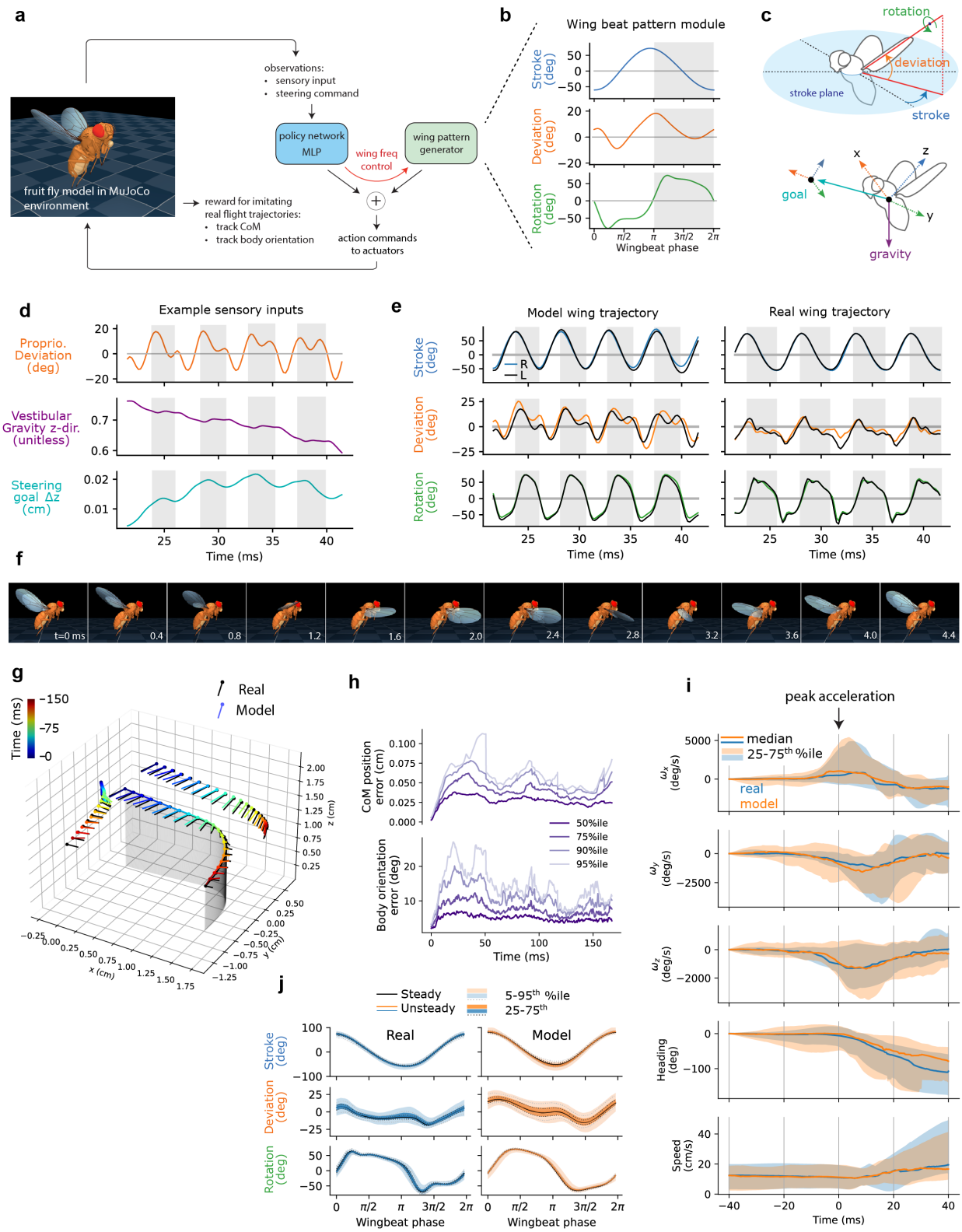
In addition to the steering command, the policy network controller received a 62-dimensional sensory signal from the proprioceptive and vestibular sensory systems (Fig. 2d, Methods, Table 8). The controller output 12-dimensional instantaneous wing torques, head and abdomen angle change, and WPG frequency (Methods, Table 9). The legs were retracted to their typical flight position (Fig. 1j) and their DoFs were frozen to reduce control complexity. The policy network was trained so that the model's body CoM and orientation matched those of the reference trajectories. Note that reference trajectory wing angles were not used to train the model, but were only used for evaluation. The reward function and training details are provided in Methods.

We evaluated the performance of our trained steerable controller on a test set of 56 CoM trajectories. The model fly was able to match the desired CoM trajectory with high accuracy (median CoM position error of 0.25 mm, median body orientation error <5 °, Fig. 2h, example trajectory shown in Fig. 2g). A filmstrip of a single wing beat cycle produced by the trained model fly when commanded to fly straight at 30 cm/s is shown in Fig. 2f. The model fly was trained to match the target CoM trajectories of real flies, but was largely free to achieve this via wing trajectories only constrained (through the DMPO action penalization, Methods) to stay close to a common baseline WPG pattern. Therefore, we compared model and real wing trajectory patterns as a means to characterize the fidelity of our physics simulation of fluid forces as well as the realism of model fly behavior. We found that a good match in CoM trajectory was achieved by qualitatively similar wing trajectories of the model and real flies, but with slight differences in wing beat frequency (Fig. 2e). In general, we found that while the wing beat frequency changed during the maneuvers by up to ~ 40 Hz in the datasets, in the model fly these frequency changes clustered more tightly in the $\sim 0-10$ Hz range. As the model and data flies were different species with average wingbeat frequencies of 218 Hz (*Drosophila melanogaster*) and 192 Hz (*Drosophila hydei*), we did not attempt to quantitatively compare wing beat trajectories.

Similar to real flies, the model fly only generated small differences between the left and right wing trajectories in order to generate large accelerations during saccade turns (Fig. 2e). Since the model fly was trained and evaluated on trajectories involving such extreme maneuvers, we were able to compare changes in wing beat patterns between "steady" flight during periods of minimal acceleration, and "unsteady" flight during periods of large acceleration (Fig. 2j). Our model fly wing trajectories recapitulated previous observations [Fry et al., 2005, Dickson et al., 2008, Muijres et al., 2014, Muijres et al., 2015], that only small differences in the wing stroke between steady and unsteady flight (Fig. 2j) are sufficient to generate large changes in CoM trajectory (Fig. 2i). Further, the match in median angular velocities, heading, and speed, show that our model fly recapitulates many of the stereotypic features of the manuevers used by real flies to turn [Muijres

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

et al., 2015, Dickinson and Muijres, 2016].



bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 2: **Flight imitation.** (**a**) RL task setup. A single policy network is trained to imitate the body CoM and orientation across a data set of 216 trajectories ($\sim$43 seconds total) of flying *Drosophila*. The controller consists of a trainable multi-layer perceptron (MLP) and a fixed wing-beat pattern generator (WPG). The net motor command received by the fly model is the MLP and WPG outputs summed. (**b**) One period of baseline wingbeat pattern (same for both wings) produced by the WPG. Gray stripe indicates wing downstroke. (**c**) Top: illustration of the wing coordinate system and wing angle names. Bottom: illustration of the body coordinate system and some of the sensory information available to the model: the direction to the goal CoM position and the direction of gravity. (**d**) A subset of policy inputs: wing deviation angle (proprioception), z-component of gravity direction (vestibular), z-component of target CoM position (steering command) for an example flight trajectory. (**e**) Wing angles during a saccade produced by the model and the corresponding real fly used for steering. Gray stripes indicate wing downstrokes. (**f**) Filmstrip of one full wing beat cycle of the model flying straight at 30 cm/s, shown with 0.4 ms step between frames. (**g**) Wings produce body movements through a phenomenological fluid model. Plot of the real fly's body pose (black) and the model fly's body pose traversing a test trajectory, colored by time. Projections of these on the xz and yz planes are also shown. Circles: head, lines: tail. (**h**) Percentiles of errors between the model and corresponding real fly's body center-of-mass (top) and orientation (bottom) for 56 test trajectories. (**i**) Median and 25th-75th percentiles of body angular velocity, heading, speed for real and model flies during saccades in test set. The trajectories are aligned to peak acceleration at $t = 0$. Roll, ω_x, and pitch, ω_y, angular velocities are similarly important in model flies' as real flies' turns. A small divergence between the model and real flies occurs after the saccade. (**j**) Percentiles of wing angles during steady (small body acceleration) and unsteady (large body acceleration) wing beats for model and real flies in the test set. Large body accelerations are achieved by similarly small alterations to the median wing beat pattern.

Walking

We trained a steerable closed loop controller for walking (Fig. 3a), also using imitation learning. We performed high-speed top-down-view videography (150 fps) of groups of fruit flies walking and interacting freely in a circular arena. Automated pose tracking was applied to female fruit flies to track the 2D locations of 13 keypoints located at the head, thorax, abdomen and 6 leg tips (Methods). The 3D pose of all the body DoFs cannot be unambiguously inferred from just these 2D keypoints locations, so we used a form of regularized inverse kinematics to infer an approximation of the full 3D fly pose trajectories for all degrees of freedom (Methods). The collected trajectories ($\sim$16k trajectories, 80 minutes total) consisted of flies walking freely at variable speeds ($\sim$0–4 cm/s, Fig. 3e inset), turning, and standing still.

Flies use a variety of gait patterns as a function of walking speed, characterized by differences in limb coordination and the number of feet on the ground at any given time [DeAngelis et al., 2019]. This is in contrast to flight where the full range of flying behaviors is executed by subtle modifications to a common baseline wingbeat pattern [Muijres et al., 2014, Dickinson and Muijres, 2016]. Because of the significant gait variability, we could not utilize a simple pattern generator for walking imitation as we did for flight. Instead, the controller consisted of a single fully-connected MLP (Fig. 3a) and was trained without enforcing any particular structure such as pattern generators. A single MLP policy network was trained on all $\sim$13k train set walking trajectories.

A much larger number of DoFs (59 for walking vs 12 for flight) has to be controlled during walking compared to flight, corresponding to the legs (including adhesion), abdomen, and head. Therefore, the sensory signals provided to the controller network were much higher, 286-dimensional (Methods, Table 16), corresponding largely to an increase in proprioceptive signals. Similarly, a 59-dimensional motor control signal was required from the controller (Methods, Table 17). The wings were folded and their DoFs frozen. During training, the model fly was rewarded for reproducing the detailed leg movements of the real fly, as well as tracking the CoM trajectory, in response to high level steering commands in the form of desired future body CoM position and orientation.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As we only had measurements of the body kinematics and no direct reference measurements of leg adhesion, the reward function did not include a term for imitating leg adhesion. However, we observed that the agent generally learned to activate adhesion when legs were in stance (on the ground). See Methods for walking imitation reward and training details.

We characterized the walking behavior produced by the steerable controller on a test set of 3.2k trajectories. The CoM position tracked the desired position and orientation well (median position error: 0.4 cm, median orientation error: 4°, Fig. 3b). A single walking cycle produced by the model fly, commanded to walk straight at 2 cm/s, is shown in Fig. 3c. We then analyzed the gaits produced by the model fly when commanded to walk straight at a given constant speed. The gait diagrams (Fig. 3d) show the duration that each leg spent in stance or swing phases, defined by whether the leg was stationary on the ground or moving, respectively (see the distribution of walking speeds in data in Fig. 3e, inset). At any time, fruit flies generally have at least three legs on the ground in stance phase, but more legs can be in stance at slower walking speeds [DeAngelis et al., 2019]. The model fly recapitulates this phenomenon with 3.1 legs in stance phase on average at during high speed walking (4 cm/s), and 3.9 at the slower walking speed of 1 cm/s. Across the full range of walking speeds observed in the dataset, we find good agreement for average number of legs on the ground between model and real flies (Fig. 3e). We analyzed the leg coordination patterns of the model and real flies by computing the phase delay during a walking cycle between the swing phase onsets of each leg relative to the left foreleg (L1), and find good agreement between the phase distributions of the model and real fly legs (Fig. 3f).

Fig. 3g shows the learned turning strategy. We analyzed leg tip trajectories produced by the model fly when commanded to either walk straight (black) or turn left (green) or right (red) at a constant walking speed of 1, 2, or 3 cm/s and a fixed turning radius of 1 cm. The model fly decreases stride length of legs on the turning side and increases stride length of legs on the opposite side as seen in the leg tip projections for turning at 2 cm/s (top), and qualified for all speeds (bottom). This agrees with the observed change in step length for the middle- and hind-limbs of freely walking flies [DeAngelis et al., 2019]. In contrast, our model has learned an asymmetric pattern for the forelimbs, modulating the length for the front limbs during left turns but not right turns, a pattern not typically seen in walking measurements of real flies.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

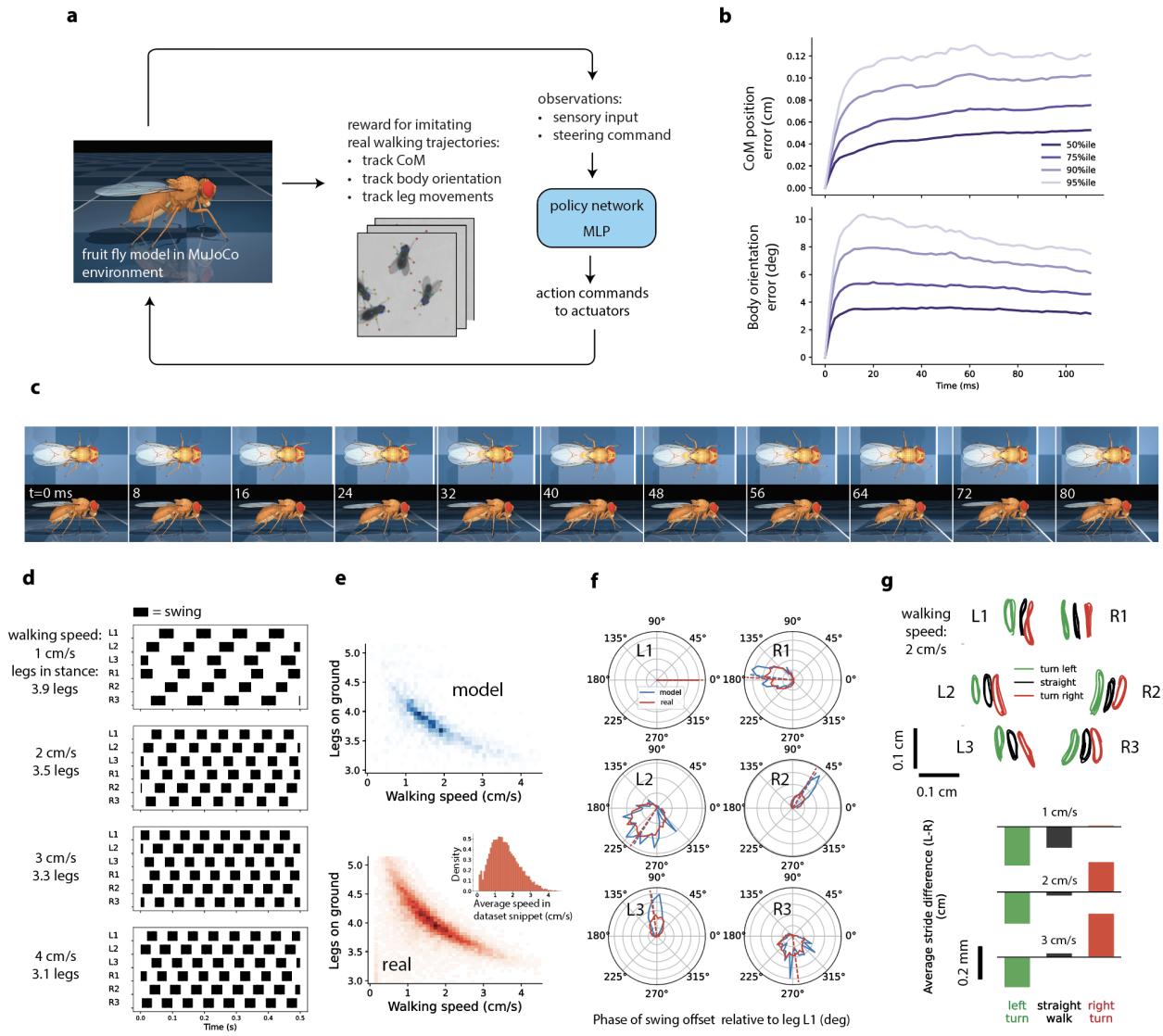


Figure 3: **Walking imitation.** (**a**) RL task setup. A single policy network is trained to imitate a dataset of 13k walking snippets ($\sim$64 minutes total) of real *Drosophila*. Full body movements are imitated, including tracking body CoM position and orientation, and detailed leg movements. (**b**) Percentiles of errors between the model and corresponding real fly's body center-of-mass (top) and orientation (bottom) for 3.2k test walking trajectories. (**c**) Filmstrip of one full leg cycle of the model walking straight at 2 cm/s with 8 ms step between frames. (**d**) Gait diagrams of fly model tracking artificial fixed-speed straight-walking trajectories at four speeds. For each speed, average number of legs simultaneously in stance position (on ground) is indicated. Black stripes indicate swing leg motion. (**e**) Number of legs simultaneously in stance position (on the ground) averaged over snippet vs average snippet walking speed. Top: model tracking all test set trajectories. Bottom: all of the walking dataset. Inset shows the distribution of average walking speeds per snippet in the dataset. (**f**) Distributions of phases of swing onsets of all legs relative to front left leg L1 in walking trajectories with mean speed in range [1.2, 1.7] cm/s. Blue is the model tracking testset trajectories and red is all of the reference data. Dashed lines indicate circular medians. (**g**) Learned turning strategy. Top: xy-projection of leg tip trajectories in egocentric frame for model walking straight, turning left, right, at speed of 2 cm/s. Leg tip trajectories are shifted horizontally for clarity. Bottom: Difference between left and right leg tip swing length averaged over all legs at different walking speeds.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Vision-guided flight

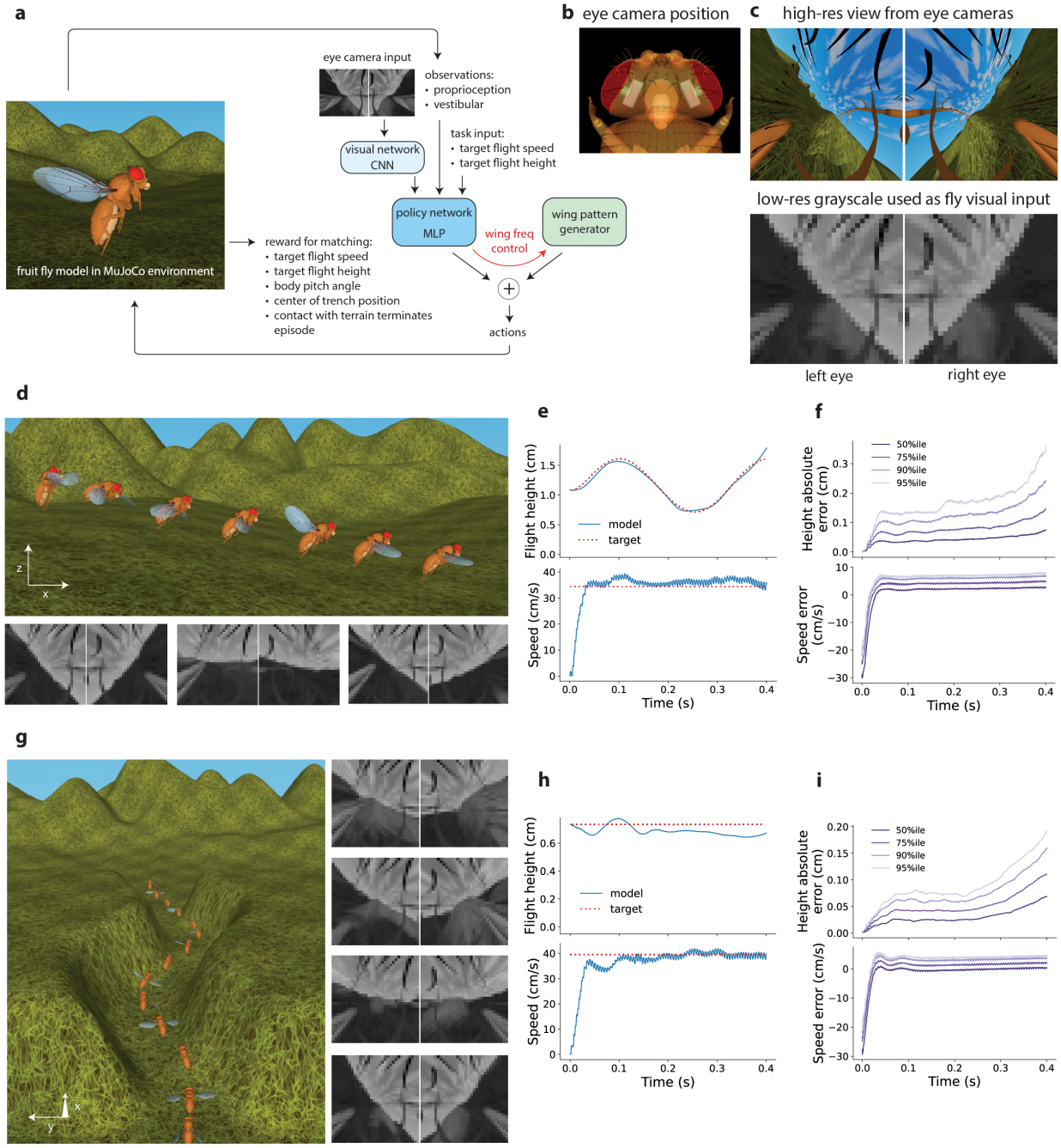
The fruit fly is a highly visual insect, with two compound eyes dominating its head and with two optic lobes which each comprise about a third of the fly's brain [Schlegel et al., 2023, Dorkenwald et al., 2023]. Therefore, in addition to proprioptive and vestibular sensors, we also modeled visual sensors. We model the eyes with MuJoCo camera sensors (Fig. 4b), rendering a square lattice of 32×32 pixels with a field of view of 150 degrees. This roughly corresponds to the resolution of the *Drosophila* compound eyes with an inter-ommatidial ("pixel-to-pixel") visual angle of 4.6 degrees [Zhao et al., 2022] (Fig. 4c).

We demonstrate the use of the visual system by training the model fly to perform two visually guided flight tasks ("bumps" and "trench" tasks) using end-to-end reinforcement learning (Fig. 4a). Both tasks required using visual inputs to navigate. Fig. 4c shows an example of a low-resolution visual input frame and its high-resolution counterpart rendered during flight. The flight controller architecture is similar to that of the flight imitation task described above. As before, policy network and WPG work together to produce wing action commands. In addition, the policy network also received a low-dimensional visual feature representation produced by a convolutional network (CNN) applied to the visual inputs. Thus the controller network (vision CNN + policy network) effectively received a 62-dimensional proprioceptive and vestibular sensory signal, as well as 6,144-dimensional visual signal from the eyes. The controller also received an additional 2-dimensional task-specific input: target flight height and speed (Methods, Table 12). As in the flight imitation task, the controller produces 12-dimensional wing torques, head and abdomen angles, and WPG frequency (Methods, Table 13). The CNN representing the visual system and the MLP producing the motor control signals were jointly trained in an end-to-end fashion to maximize task reward. In both tasks, the precise terrain of the virtual world, as well as the target height and speed were randomly generated in each training and test episode (Methods, Table 14). Contact with terrain results in early episode termination with failure (Methods). The model fly starts at zero speed, requiring that the fly speed up to the target speed at the beginning of the task.

Bumps task. During flight, fruit flies use visual cues to estimate their altitude above the ground, which could be used to maintain a constant height above the ground when flying over uneven terrain [Straw et al., 2010]. We modeled this visually guided altitude control task by constructing a virtual world with a randomly generated sinusoidal height profile of the terrain. The model fly is rewarded for flying straight at a constant target forward velocity while maintaining a constant target height above the ground (Fig. 4d). After training, the model fly successfully learned to use its visual system to match the target altitude (Fig. 4e,f, median errors after the initial speed-up period: height error: 0.045 cm, speed error: 2.2 cm/s).

Trench task. In a second task, we trained the fly to navigate through a trench without bumping into the walls of the trench. A virtual world was constructed with a trench with a sinusoidal curving profile (Fig. 4g) and constant width and depth. The model fly was rewarded for maintaining a constant forward speed and constant height inside the trench, while a collision with the trench walls resulted in an early demise and termination of the episode and a loss of future rewards. Thus maximizing the reward requires that the model fly successfully navigate the trench by turning to avoid collisions with the walls. After training, the model fly successfully learned to use its visual system to detect the walls of the trench and navigate the entirety of the trench while maintaining constant target height and target speed (Fig. 4h,i, median errors after the initial speed-up period: height error: 0.032 cm, median speed error: -0.16 cm/s).

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure 4: **Vision-guided flight tasks: altitude control ("bumps") and obstacle avoidance ("trench").** (**a**) RL task setup. Policy is trained to carry out forward flight at a given target speed and height while avoiding contact with terrain. In each training trajectory, the terrain is procedurally generated and the target speed and altitude are randomly selected. We considered two kinds of terrain. In the "bumps" task, the terrain is a sequence of sine-like bumps perpendicular to the flight direction. To solve the task of flying at given target height above the bumpy terrain, the agent has to learn to constantly adjust the flight altitude. In the "trench" task, the terrain is in the shape of a sine-like trench. To fly through the trench, the agent learns to execute turning maneuvers. To solve either task, the agent has to utilize the visual input from eye cameras to avoid terrain, and maintain target altitude. As in the flight imitation task, the flight controller consists of policy network and wingbeat pattern generator. The visual input is processed by an added convolutional network. The target flight speed and height are provided to the policy as task inputs. Both the policy and the convolutional network are trained end-to-end with RL. (**b**) Semi-transparent top-down view of fly model head showing schematically the position of MuJoCo eye cameras. (**c**) Fly visual input. Top: High-res view from eye cameras for fly's flight position in Panel A. Bottom: Corresponding downsampled grayscale frames are used as fly visual input. (**d**) Top: Time-lapse of flight executed by policy trained in "bumps" task. Bottom: visual input frames captured by eye cameras for the first three time-lapse positions. (**e**) "Bumps" task performance: model and target flight height and speed for a single representative trajectory. (**f**) Percentiles of height and speed errors for 1000 trajectories. (**g**) Left: Time-lapse of flight generated by policy trained in "trench" task. Right: visual input frames for the first four time-lapse positions. (**h**) "Trench" task performance: model and target flight height and speed for a single representative trajectory. (**i**) Percentiles of height and speed errors for 1000 trajectories.

Discussion

The behavior of an animal emerges from the interplay between its nervous system, its body, and its environment. Here, we demonstrated realistic complex locomotion behavior, both walking and flight, in an anatomically detailed whole body model of the fruit fly. This advance was enabled by improved simulation of the physics of the body and its interactions with the world, and by using deep reinforcement learning to approximate the nervous system with an artificial neural network trained to mimic the behavior of real flies. We constructed an anatomically detailed model of the fruit fly with 67 rigid body parts and 102 degrees of freedom, actuated via torques at the joints. Using the MuJoCo physics engine, we simulated the body and its interactions in its environment through rigid body collisions and through fluid interactions with air. Finally, we have used deep RL and imitation learning approaches to construct a closed-loop low-level neural controller that induces the body to generate realistic patterns of body movement, both for walking and flight modes of locomotion, along arbitrary trajectories. All three components are released as open source software along with pre-trained controllers, and can be accessed at https://github.com/TuragaLab/flybody/.

We intend our open source project to serve as a platform for realistic physics simulation of fruit fly biomechanics and motor control. The present work integrates measurements across multiple spatial and temporal scales, bringing together microscopy of static anatomy with high speed videography of dynamic locomotor behavior. Our work can be built upon in several ways. Confocal imaging (Extended Data Fig. 1), as well as micro CT [Lobato-Rios et al., 2022] and XNH [Kuan et al., 2020] imaging of the body support detailed musculo-skeletal, adding further realism to the directly actuated joints in the present model. This includes modeling of the neck [Gorko et al., 2024], wing hinge [Melis et al., 2024], and coxa [Kuan et al., 2020, Mamiya et al., 2023] joints. On the sensory side, the idealized sensory systems can be improved by incorporating recent mapping of proprioceptive organs in the leg and wing [Kuan et al., 2020, Mamiya et al., 2023], along with recently developed eye-maps which can be used to inform the positions of ommatidia in visual

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

space [Zhao et al., 2022]. Further, model-based pose-tracking algorithms could be used to extract improved kinematic detail from existing high-speed single- and multi-view videography [Bolaños et al., 2021, Jiang and Ostadabbas, 2023, Plum et al., 2023, Sun et al., 2023]. Finally, newly available connectome datasets can be used to better model the neural control of behavior.

It has recently been proposed that understanding the brain requires embodied models of the nervous system capable of generating an ethologically relevant repertoire of behavior [Zador et al., 2023]. This NeuroAI approach is enabled with unprecedented spatial and temporal detail by our body model and imitation learning framework. Future work can leverage the connectomic mapping of the entire fruit fly nervous system [Dorkenwald et al., 2023, Schlegel et al., 2023, Lesser et al., 2023, Azevedo et al., 2022, Cheong et al., 2023, Marin et al., 2023, Takemura et al., 2023] to more realistically model neural circuits underlying sensorimotor behavior. In the short term, our body model and imitation learning framework can enable the model-based investigation of the neural underpinnings of sensory-motor behaviors such as escape invoked by looming stimuli [Card, 2012], gaze-stabilization [Cruz and Chiappe, 2023], the control of movement by the ventral nerve cord [Lesser et al., 2023, Azevedo et al., 2022, Cheong et al., 2023, Marin et al., 2023, Takemura et al., 2023]. In the long term, our whole body model, in concert with connectome-constrained deep mechanistic neural network models[Lappalainen et al., 2023, Mi et al., 2022] of the whole nervous system could eventually be used to construct whole animal models of both the entire nervous system and body of the adult fruit fly.

Data availability

Confocal imaging stack, flight and walking imitation datasets, base wingbeat pattern, and trained controller networks are available at https://doi.org/10.25378/janelia.25309105

Code availability

Fly model and code are publicly available at https://github.com/TuragaLab/flybody

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Methods

Sample preparation for anatomical measurements

The flies were anesthetized on ice, briefly washed with ethanol and dissected under PBS-T (PBS + 0.1% Triton X-100). Disassembling the fly into manageable elements allowed us to use high magnification, high numerical aperture (NA) objectives that have – in relation to the size of a fly's body – short working distances but have the benefit of higher axial resolution than the lower magnification/NA objectives. Heads, wings, thoraces with abdomens, fore-, mid- and hind legs were transferred to individual tubes. All body parts except the wings were incubated with 0.25 mg/ml trypsin in PBS-T for 48 hours at 37° C to remove the soft tissues. The cuticle was then bleached in 20% H_2O_2 for 24h, and the exoskeleton. The tendons were stained overnight with Congo Red (0.5 mg/mL, Sigma-Aldrich #C676-25G), a bright and comparatively photo-stable chitin-binding dye stains both soft, membranous as well as hard, sclerotized cuticle. It also shows affinity to tendons and fine tendrils, which is very convenient for identification of muscles' origins and insertion sites, even in the absence of soft tissues. The samples were dehydrated in ethanol and mounted in methyl salicylate (Sigma-Aldrich #M6752), which has refractive index very close to that of glass, facilitated imaging throughout the relatively thick/bulky samples without degradation of the signal. Serial optical sections were obtained on a Zeiss 880 confocal microscope at 2 µm with a Plan-Apochromat 10x/0.45 NA objective, 1 µm intervals with a LD-LCI 25x/0.8 NA objective, or 0.3 µm with a Plan-Apochromat 40x/1.3 NA objective. The 560 nm laser line was used to excite Congo Red.

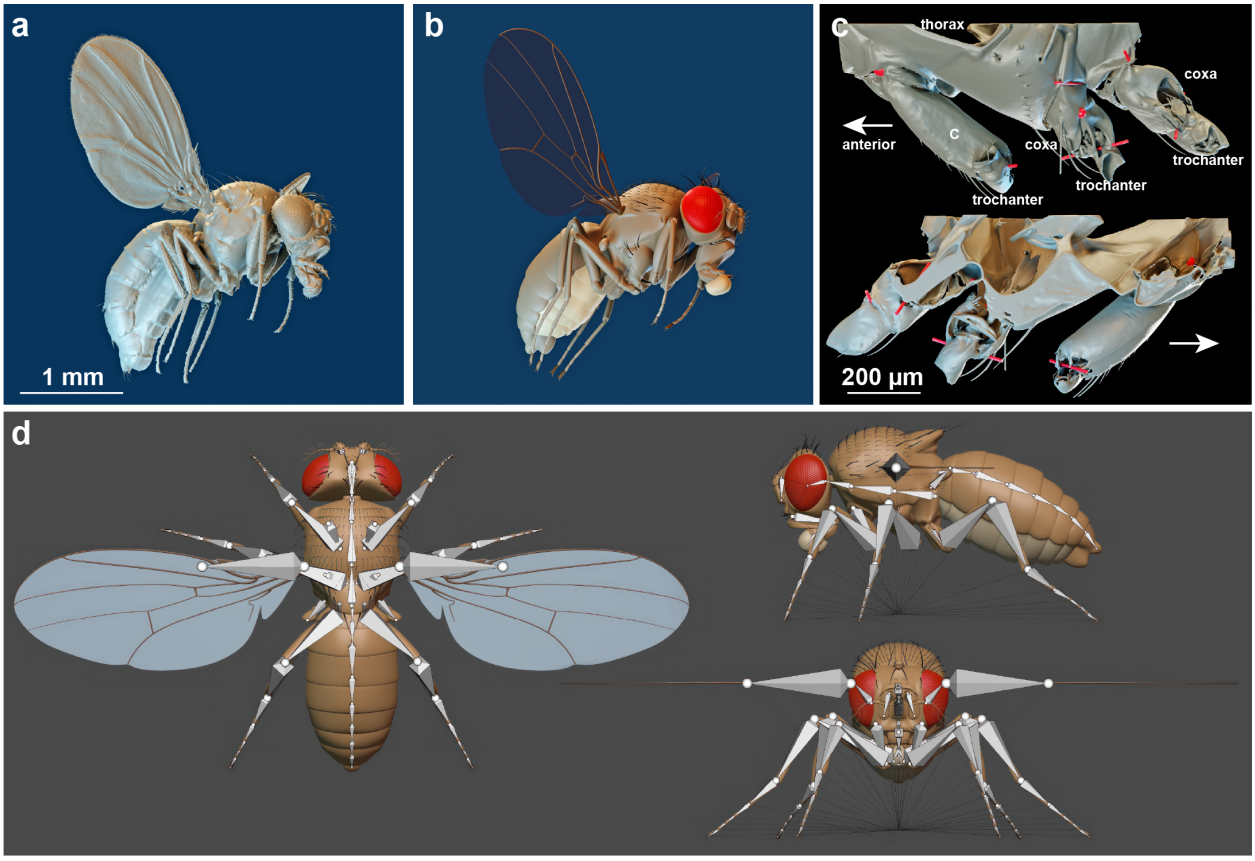
Blender model of body geometry

3D meshes were extracted from the confocal stacks using Fiji's 3D viewer plugin [Schindelin et al., 2012] and imported into Blender [Community, 2018]. A 3D model was constructed from meshes representing the head, thorax+abdomen, wing and fore-, mid- and hind-leg of a single female fly. Appendage meshes were mirrored across the body's medial plane (Extended Data Fig. 1a). This model was used as the reference for creating a lower polygon count, simplified one, in which the total number of vertices was reduced from 22.6M to 20K (Extended Data Fig. 1b). This simplified model consisted of 67 articulated body segments (Extended Data Fig. 1d): 9 body axis segments (head, thorax, 7 abdominal segments), proboscis (4 segments), antennae, wings, halteres (6 total) and legs (coxa, femur, tibia, 4 tarsal segments and tarsal claws, 6×8 segments). The exact positions of joints, articulations, and axes of joints' rotation were determined with high confidence from confocal microscopy data (Fig. 1b, Extended Data Fig. 1c). The model was posed in the rest position and rigged in Blender by creating constraints defining movement of the body segments with respect to each other. Each of the 67 body segments was assigned (parented to) a control element called "bone", forming a hierarchical kinmatic tree system resembling a skeleton called "armature" (Extended Data Fig. 1d).

MuJoCo model of body physics

The Blender model was then exported to MuJoCo using the dm_control exporter. The components representing head, thorax, abdomen, wings and legs were assigned densities based on average values from weighing 2 groups of 30 and 22 disassembled wild type female flies: head - 0.15 mg, thorax - 0.34 mg, abdomen - 0.38 mg, legs (each) - 0.0162 mg, wings (each) - 0.008 mg. This

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 1: Constructing a 3d model of a female fruit fly from confocal data. (**a**) Hi-polygon (∼22.6 M faces) model of the fly reconstructed in Blender. (**b**) Simplified model (∼20 K faces). (**c**) 3D model of the left-side coxae based on the confocal data. Red bars represent hinge joints, spheres represent ball joints. Arrows indicate anterior. (**d**) Dorsal, sagittal and frontal views of the rigged model in the rest position, the elements of the armature called "bones" shown as elongated octahedrons.

corresponds to total fly mass of 0.983 mg. The full body length of the model is 0.297 cm, wing span 0.604 cm.

Joint limits were initially determined using Blender's inverse kinematics tool. We started with fairly tight joint limits and then used reference images of extreme articulated postures (mostly from grooming behaviors) to increase joint limits as required, until all reference poses could be achieved. We then refined the leg joint limits using automated inverse kinematics fitting of the model to 392 frames from manually annotated grooming behavior videos, see below. The sensory system details in the model's default configuration is shown in Table 2. Degrees of freedom were actuated using torque or position actuators, with certain DoFs (abdomen, tarsi) coupled by tendons (Table 3). For position actuators, control ranges were set to be equal to the corresponding joint ranges. For more detail on building the fly MuJoCo model, see Supplement. For task-specific model modifications, see the Methods sections on Flight and Walking below.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1: Empirical masses of the fly body parts. Averaged over 52 female flies.

Body part	Mass (mg)
head	0.15
thorax	0.34
abdomen	0.38
leg (each)	0.0162
wing (each)	0.008
fly total	0.983

Table 2: Sensory system of the fly model in its default configuration. Task-specific model modifications (e.g., disabling legs for flight) may include disabling some of the sensors and altering the dimensions of the proprioception observables. All observables are calculated in fly's egocentric reference frame. The actuator activation state units are the same as the corresponding actuator control units.

Category	**Sensor**	**Units**	**Array shape**
vision	right eye camera	unitless RGB	(32, 32, 3)
	left eye camera		(32, 32, 3)
vestibular	velocimeter	$\mathrm{cm/s}$	3
	accelerometer	$\mathrm{cm/s^2}$	3
	gyro (rotational velocity)	$\mathrm{rad/s}$	3
	gravity direction	cm	3
proprioception	joint angles	rad	102
	joint velocities	$\mathrm{rad/s}$	102
	actuator activation state	same as control	78
	end-effector positions	cm	(7, 3)
mechanoreception	leg force sensors	dyn	(6, 3)
	leg touch sensors	dyn	6

Table 3: Actuators of the fly model in its default configuration. Task-specific model modifications (e.g., disabling legs for flight) may include removing some of the actuators. By default, each actuator actuates one DoF. In the "position coupled" category, one actuator actuates several DoFs coupled by a MuJoCo tendon.

Actuator category	**Located at**	**# of actuators**
torque	wings	6
position	neck	3
	proboscis	5
	antennae	6
	legs (excluding tarsi)	42
position coupled	tarsi	6
(several DoFs per actuator)	abdomen	2
adhesion	leg tips	6
	labrum	2
total		78

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Analysis of leg degrees of freedom

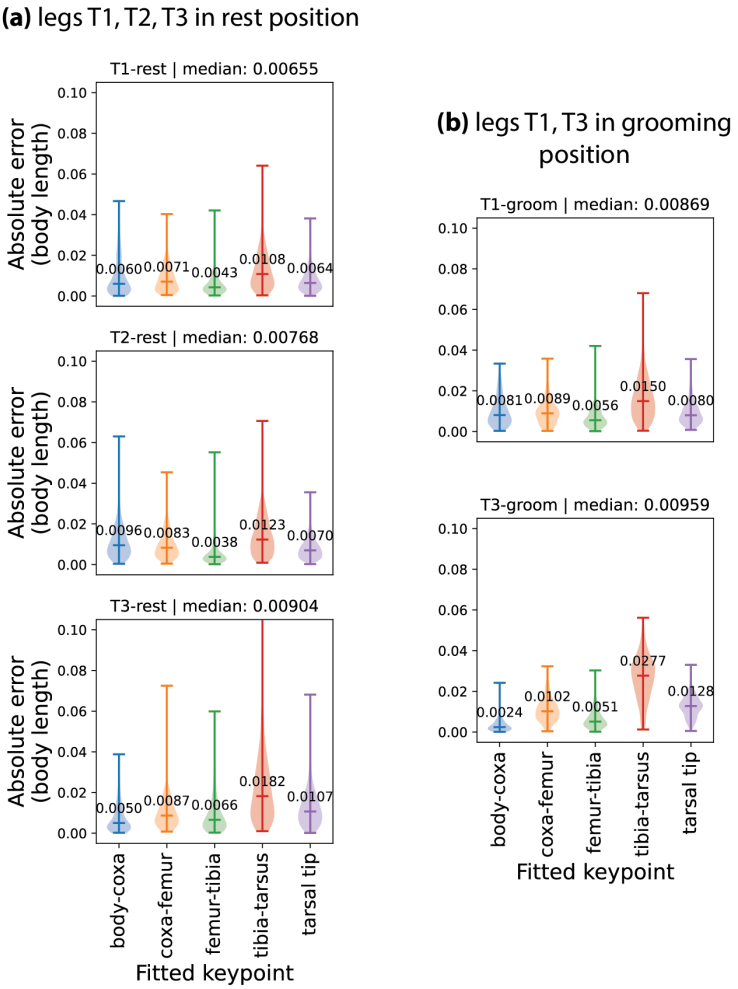
In order to verify our approximation of the leg degrees of freedom and leg joint ranges we applied the following procedure. We recorded 2-camera videos [Williamson et al., 2018] of multiple free *Drosophila* during grooming behavior. We then uniformly sampled and annotated individual frames of the fly postures during grooming, giving us 3D coordinates of five keypoints for each leg: the four leg joints (body-coxa, coxa-femur, femur-tibia, tibia-tarsus), plus the tarsal tip. We annotated all six legs per frame regardless which legs were actively involved in grooming in the frame. This provided us with data for legs both in grooming positions and in rest (standing) positions. We only observed grooming with T1 and T3 legs and we collected a total of 392 frame annotations. Then we performed inverse kinematics fitting of the model legs to the annotated frames as follows (for details on the inverse kinematics fitting procedure, see also the walking reference data preparation below.) In order to decouple the effect of fly-to-fly size/proportion variability and the actual degrees-of-freedom mismatch, in each frame we rescaled the model's leg segments to match data. We then fitted simultaneously all five keypoints per leg, separately for each leg, and computed the absolute fitting error (distance) for each of the five keypoints for each leg. Extended Data Fig. 2 shows the distributions of the inverse kinematics fitting errors for each keypoint and each leg. Extended Data Fig. 2a shows the errors for leg fits in rest position, Extended Data Fig. 2b shows errors in grooming positions. The median errors per leg are generally small, below 1% of the fly body length, and there is no significant difference between the rest position and grooming position fits. There appears to be a slight systematic increase in the tibia-tarsus keypoint error, more noticeable in the grooming fits in Extended Data Fig. 2b, which is not surprising as grooming leg positions tend to be more intricate than the rest position. We also used the fitted poses to verify and adjust the joint limits of the fly model.

Distributed Reinforcement Learning

For each locomotion task, we trained a policy network using a distributed RL setup [Nair et al., 2015, Horgan et al., 2018] powered by `Ray`, an open-source general-purpose distributed computing package [Moritz et al., 2017]. The distributed training configuration is shown in Extended Data Fig. 3. Multiple CPU-based actors run in parallel in separate MuJoCo environment instances, generate experiences, and log them into a replay buffer. A single GPU-based learner samples training batches from the replay buffer and updates the policy and critic network weights. The critic network is a part of the training process only and is not used by the fly model directly. Each actor explores the environment and generate experiences using its own copy of the policy network whose weights are periodically synchronized with the current learner policy. For learner policy updates, we used the off-policy actor-critic DMPO agent, a distributional extension [Bellemare et al., 2017] of the MPO agent [Abdolmaleki et al., 2018b, Abdolmaleki et al., 2018a]. We used `dm_control` [Tunyasuvunakool et al., 2020] to set up the RL environments and for MuJoCo python bindings. We used the DMPO agent implemented in `acme` [Hoffman et al., 2020] and the replay buffer implemented in `reverb` [Cassirer et al., 2021], and the `Adam` optimizer [Kingma and Ba, 2017]. The hyperparameters of the distributed setup and of the DMPO agent are shown in Table 4 and in Table 5.

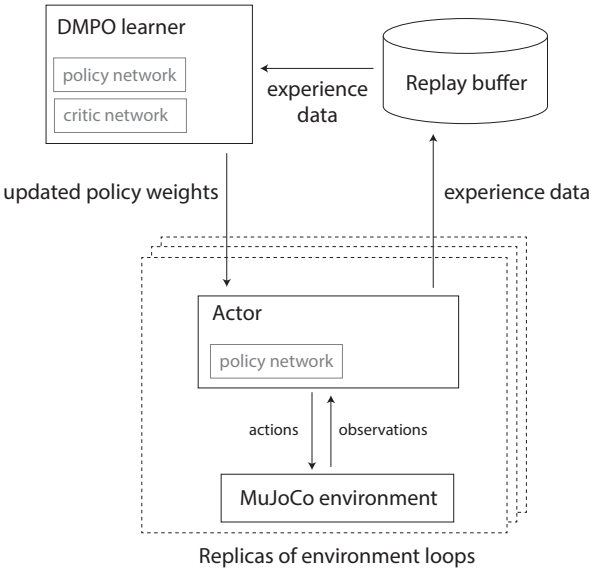
To guarantee stability, we ran MuJoCo physics simulations at $\times$4–10 smaller time steps than we sampled action commands from the policies [Tunyasuvunakool et al., 2020], see the physics and control time-step values in Tables 10, 14, 18. The policies were stochastic during training (outputting distribution over actions) to facilitate exploration. The distributions were Gaussian, independent for each action dimension and parameterized by mean and standard deviation. The

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 2: **Leg degrees of freedom analysis.** Inverse kinematics fits of the fly model legs to real *Drosophila* grooming poses (392 poses in total.) To separate the effect of degrees-of-freedom from fly-to-fly size variability, model legs were rescaled to match these in each individual reference pose frame. Individual model legs were fitted separately by simultaneously matching five leg keypoints located at four leg joints and leg tip. Absolute fitting errors for each leg keypoint are shown for (**a**) all leg pairs in rest position, (**b**) leg pairs T1 and T3 in grooming positions. Horizontal bars and corresponding values are median errors for each keypoint. Median errors across all keypoints in each leg pair are also indicated.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 3: **Distributed training architecture.** Multiple replicas of actors in MuJoCo environments collect experiences and feed them to a single replay buffer. The DMPO learner samples experiences from the replay buffer, updates the policy and critic network weights, and sends the updated weights to the actors' copies of the policy.

policy network architectures for each task are given in Tables 11, 15, 19. At test time, the policies reverted to being deterministic by using the means of the predicted action distributions. The actions output by the policy networks were in the canonical range $[-1, 1]$. The actions were then rescaled to match their corresponding proper ranges in the fly model, e.g. joint limits for position actuators or force limits for force actuators. All observables (policy inputs) were strictly egocentric, e.g. calculated in the local reference frame of the fly model. We only used feedforward policy and critic networks and did not extensively sweep the network architectures.

We trained the model in episodes of finite number of time steps. An episode ends either when (i) the episode time limit is successfully reached, or (ii) when an early termination condition, signifying failure, is met. Termination condition details are explained in the corresponding RL task sections below. In the first case, the agent estimates the remaining infinite-horizon future rewards (beyond episode's final step) by bootstrapping from the value of the state at the end of the episode. In the second case, the reward sequence is truncated by setting the future reward to zero. The loss of the future infinite-horizon rewards is an unfavorable outcome and the agent will try to learn to avoid events triggering early episode termination.

Phenomenological model of the fluid forces

This section describes the fluid force model we introduced to MuJoCo to facilitate the simulation of fly's flight. The fluid model computes forces exerted on moving rigid bodies whose shape can be approximated by ellipsoids. The model provides fine-grained control of the different types of fluid forces via five dimensionless coefficients, the `fluidcoef` attribute in MuJoCo, see the MuJoCo documentation for more detail [1]. We used the fluid model to compute forces on fly's flapping wings whose shape closely matches slender ellipsoids (Fig. 1g). Elements of the fluid force model are a generalization of [Andersen et al., 2005] to three dimensions. The force $\mathbf{f}$ and torque $\mathbf{g}$ that the surrounding fluid exerts onto the translating and/or rotating body are approximated as a sum of

[1] https://mujoco.readthedocs.io/

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 4: Distributed training hyperparameters. See implementation for more detail.

Hyperparameter	Value (walking)	Value (flight)
num_actors	32	32
batch_size	256	256
prefetch_size	4	4
min_replay_size	10^4	10^4
max_replay_size	4×10^6	4×10^6
samples_per_insert	15	15
n_step	10	5
num_samples	20	20
policy_optimizer_lr	10^{-4}	10^{-4}
critic_optimizer_lr	10^{-4}	10^{-4}
dual_optimizer_lr	10^{-3}	10^{-3}
target_critic_update_period	200	107
target_policy_update_period	200	101

Table 5: DMPO agent hyperparameters. See implementation for more detail.

Hyperparameter	Value (walking)	Value (flight)
epsilon	0.05	0.1
epsilon_mean	5×10^{-5}	2.5×10^{-3}
epsilon_stddev	10^{-7}	10^{-7}
action_penalization	False	True
epsilon_penalty	N/A	0.1
init_log_temperature	10	10
init_log_alpha_mean	10	10
init_log_alpha_stddev	10^3	10^3
per_dim_constraining	True	True

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

effects:

$$\mathbf{f} = \mathbf{f}_A + \mathbf{f}_D + \mathbf{f}_M + \mathbf{f}_K + \mathbf{f}_V,$$
$$\mathbf{g} = \mathbf{g}_A + \mathbf{g}_D + \mathbf{g}_V.$$

The subscripts A, D, M, K and V denote Added mass, viscous Drag, Magnus lift, Kutta lift, and Viscous resistance, respectively. The forces imposed by the fluid onto each ellipsoid are computed independently by approximating the effect of an incompressible quiescent fluid of density ρ and (dynamic) viscosity ν. The problem is described in a reference frame aligned with the principal axes of the ellipsoid and moving with it. The ellipsoid has semi-axes $\mathbf{r} = \{r_x, r_y, r_z\}$, volume $V = (4\pi/3)r_x r_y r_z$, velocity $\mathbf{v} = \{v_x, v_y, v_z\}$, and angular velocity $\boldsymbol{\omega} = \{\omega_x, \omega_y, \omega_z\}$. We will also use $r_{\text{max}} = \max\{r_x, r_y, r_z\}$, $r_{\text{min}} = \min\{r_x, r_y, r_z\}$, and $r_{\text{mid}} = r_x + r_y + r_z - r_{\text{max}} - r_{\text{min}}$. The area projected by the ellipsoid onto the plane normal to $\mathbf{v}$ is

$$A_{\mathbf{v}}^{\text{proj}} = \pi \sqrt{\frac{r_y^4 r_z^4 v_x^2 + r_z^4 r_x^4 v_y^2 + r_x^4 r_y^4 v_z^2}{r_y^2 r_z^2 v_x^2 + r_z^2 r_x^2 v_y^2 + r_x^2 r_y^2 v_z^2}}. \tag{1}$$

The circulation Γ is the line integral of the fluid velocity field $\mathbf{v}_f$ around a closed curve $\Gamma = \oint \mathbf{v}_f \cdot \mathrm{d}\mathbf{l}$. The individual force and torque components are described below.

Added mass Added mass measures the inertia of the fluid that is put into motion by the body's motion. In the case of a body with three planes of symmetry, the forces $\mathbf{f}_A$ and torques $\mathbf{g}_A$ can be written as [Lamb, 1932]:

$$\mathbf{f}_A = -\mathbf{m}_A \odot \dot{\mathbf{v}} + (\mathbf{m}_A \odot \mathbf{v}) \times \boldsymbol{\omega} \tag{2}$$
$$\mathbf{g}_A = -\mathbf{I}_A \odot \dot{\boldsymbol{\omega}} + (\mathbf{m}_A \odot \mathbf{v}) \times \mathbf{v} + (\mathbf{I}_A \odot \boldsymbol{\omega}) \times \boldsymbol{\omega} \tag{3}$$

Here, $\odot$ denotes an element-wise product, $\dot{\mathbf{v}}$ is the linear acceleration and $\dot{\boldsymbol{\omega}}$ is the angular acceleration. $\mathbf{m}_A \odot \mathbf{v}$ and $\mathbf{I}_A \odot \boldsymbol{\omega}$ are the virtual linear and angular momentum respectively. The added-mass vector $\mathbf{m}_A = \{m_{A,x}, m_{A,y}, m_{A,z}\}$ and added-moment of inertia vector $\mathbf{I}_A = \{I_{A,x}, I_{A,y}, I_{A,z}\}$ measure the inertia of the fluid displaced by the motion of the body in the corresponding direction and can be derived from potential flow theory for certain simple geometries.

For an ellipsoid, the virtual mass $m_{A,i}$ for a motion along axis i and the virtual moment of inertia $I_{A,i}$ for a rotation along axis i are:

$$m_{A,i} = \rho V \frac{\kappa_i}{2 - \kappa_i}, \tag{4}$$
$$I_{A,i} = \frac{\rho V}{5} \frac{(r_j^2 - r_k^2)^2 (\kappa_k - \kappa_j)}{2(r_j^2 - r_k^2) + (r_j^2 + r_k^2)(\kappa_j - \kappa_k)} \tag{5}$$

where the dimensionless virtual inertia coefficients are [Tuckerman, 1925]:

$$\kappa_i = \int_0^\infty \frac{r_i r_j r_k \mathrm{d}\lambda}{\sqrt{(r_i^2 + \lambda)^3 (r_j^2 + \lambda)(r_k^2 + \lambda)}} \tag{6}$$

which we compute by 15-point Gauss–Kronrod quadrature. The indices i, j, k are cyclic permutations of the axes (x, y, z), (y, z, x), (z, x, y).

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Viscous drag The drag force acts to oppose the motion of the body relative to the surrounding flow. For high Reynolds numbers, the viscous drag can be approximated with the drag equations, first proposed by Newton, $f_{\mathsf{D}} = -C_D \rho s_D v^2$ and $g_{\mathsf{D}} = -C_D \rho I_D \omega^2$. Here C_D is a drag coefficient, s_D is a reference surface area (e.g., a measure of the area projected on the plane normal to the flow), and I_D is a reference moment of inertia. These quantities depend on the properties of the fluid, the shape of the body and its velocity [Duan et al., 2015]. We derive a correlation for the viscous drag based on two surfaces: the projection surface $A_{\mathbf{v}}^{\text{proj}}$ of the ellipsoid onto a plane normal to $\mathbf{v}$ and the maximum projected surface $A_{\mathsf{max}} = \pi r_{max} r_{mid}$, such that $A_{\mathbf{v}}^{\text{proj}} \leq A_{\mathsf{max}}$:

$$\mathbf{f}_{\mathsf{D}} = -\rho \left[C_D^{\mathsf{blunt}} \, A_{\mathbf{v}}^{\text{proj}} \; + C_D^{\mathsf{slender}} \left(A_{\mathsf{max}} - A_{\mathbf{v}}^{\text{proj}} \right) \right] \, \|\mathbf{v}\| \, \mathbf{v}$$

We propose an analogous model for the angular drag. For each Cartesian axis we consider the moment of inertia of the maximum swept ellipsoid obtained by the rotation of the body around the axis. The resulting components of the moment of inertia are:

$$I_{D,i} = \frac{8\pi}{15} \, r_i \; \max\{r_j, \; r_k\}^4 \tag{7}$$

where, as before, the indices i, j, k are cyclic permutations of the axes (x, y, z), (y, z, x), (z, x, y). The angular drag torque is computed as:

$$\mathbf{g}_{\mathsf{D}} = -\rho \left(\left[C_D^{\mathsf{angular}} \, \mathbf{I}_D \; + C_D^{\mathsf{slender}} \left(\mathbf{I}_{\mathsf{max}} - \mathbf{I}_D \right) \right] \cdot \boldsymbol{\omega} \right) \boldsymbol{\omega} \tag{8}$$

Here $\mathbf{I}_{\mathsf{max}}$ is a vector with each entry equal to the maximal component of $\mathbf{I}_D$.

Viscous resistance The viscous resistance terms, also known as linear drag, well approximate the fluid forces for low Reynolds numbers. These are computed for the equivalent sphere with Stokes' law [Stokes, 1850]:

$$\mathbf{f}_{\mathsf{V}} = -6\pi r_V \nu \mathbf{v} \tag{9}$$

$$\mathbf{g}_{\mathsf{V}} = -8\pi r_V^3 \nu \boldsymbol{\omega} \tag{10}$$

where $r_V = (r_x + r_y + r_z)/3$ is the radius of the equivalent sphere.

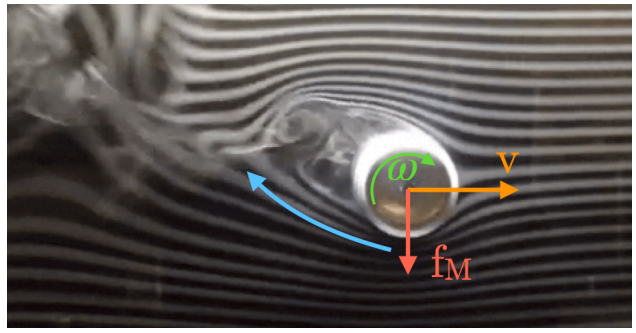
Magnus force The Magnus effect describes the motion of a rotating object moving through a viscous fluid. A spinning body induces rotation in the surrounding fluid which deflects the trajectory of the fluid flow past the body, and the body receives an equal an opposite reaction. For a cylinder, the Magnus force per unit length of the cylinder can be computed as $F_{\mathsf{M}}/L = \rho v \Gamma$, where Γ is the circulation of the flow caused by the rotation and $v = \|\mathbf{v}\|$ the speed of the object. We estimate this force for an arbitrary body as:

$$\mathbf{f}_{\mathsf{M}} = C_M \rho V \boldsymbol{\omega} \times \mathbf{v} \tag{11}$$

where V is the volume of the body and C_M is a coefficient for the force, which we typically set to 1.

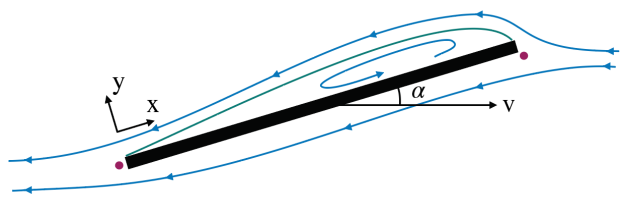
Kutta lift The Kutta condition describes an effect that is valid also for potential (i.e. inviscid) flows. For a body moving in a potential flow there are two stagnation points (a location in the flow field where the velocity is zero): in the front, where the stream-lines separate to either sides of the body, and in the rear, where they reconnect. The Kutta condition is the observation that a moving body

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 4: Schematic of the Magnus force. Smoke flow past a rotating cylinder. Due to viscosity, the rotating cylinder deflects the incoming flow upward and receives a downwards force (red arrow). Adapted from WikiMedia Commons, CC BY-SA 4.0.

with a sharp rear edge will generate in the surrounding flow a circulation of sufficient strength to hold the rear stagnation point at the trailing edge. In two-dimensional potential flow, the circulation due to the Kutta condition for a slender body can be estimated as $\Gamma_\mathsf{K} = C_K\ r_x\ \|\mathbf{v}\|\ \sin 2\alpha$. Here C_K is a Kutta lift coefficient and α is the angle between the velocity vector and its projection onto the surface. The lift force per unit length can be computed with the Kutta–Joukowski theorem as $\mathbf{f}_K/\ell = \rho\Gamma_\mathsf{K} \times \mathbf{v}$.



Extended Data Figure 5: Schematic of the Kutta condition. Blue lines are streamlines and the two magenta points are the stagnation points. The dividing streamline, which connects the two stagnation points, is marked in green. The dividing streamline and the body inscribe an area where the flow is said to be "separated" and recirculates within. This circulation produces an upward force acting on the plate.

In order to extend the lift force equation to three-dimensional motions, we consider the normal $\mathbf{n}_{A,\mathbf{v}} = \{\frac{r_y r_z}{r_x} v_x, \frac{r_z r_x}{r_y} v_y, \frac{r_x r_y}{r_z} v_z\}$ to the cross-section of the body which generates the body's projection $A_\mathbf{v}^{\text{proj}}$ onto a plane normal to the velocity. We use this direction to decompose $\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp$ with $\mathbf{v}_\perp = (\mathbf{v} \cdot \hat{\mathbf{n}}_{A,\mathbf{v}})\, \hat{\mathbf{n}}_{A,\mathbf{v}}$ ($\hat{\mathbf{n}}_{A,\mathbf{v}}$ is the unit vector). We write the lift force as:

$$\mathbf{f}_\mathsf{K} = \frac{C_K\ \rho\ A_\mathbf{v}^{\text{proj}}}{\|\mathbf{v}\|} \left(\mathbf{v} \times \mathbf{v}_\parallel\right) \times \mathbf{v} \tag{12}$$

Note that the direction of $\hat{\mathbf{n}}_{A,\mathbf{v}}$ differs from $\mathbf{v}$ only on the planes where the semi-axes of the body are unequal. So for example, for spherical bodies $\hat{\mathbf{n}}_{A,\mathbf{v}} \equiv \hat{\mathbf{v}}$ and by construction $\mathbf{f}_\mathsf{K} = 0$.

Modeling flight behavior

Wing actuator gain, wing joint damping, MuJoCo fluid model parameters

We used the following procedure to fit the flight physics parameters. We started with a wing motion trajectory recorded previously from a hovering *Drosophila Melanogaster* [Dickson et al.,

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

2008] [2]. We placed the model in a hovering position and actuated the wings to reproduce the real wing trajectory by using the real wing angles as target angles for the wing actuators. We then iteratively adjusted (increased) the wing actuator gain to a point where the mean absolute error between the reference wing angles and the trajectory traversed by the model's wings was below 5% of the wing angle amplitude. At each iteration, we also fitted the wing joint damping coefficient to avoid underdamping and ensuing wing oscillations. We used the same gain value for all three wing actuators (yaw, roll, pitch). Our final values for the gain and damping pair were `gainprm = [18, 18, 18]`, `damping = 0.007769`.

Having found suitable wing actuator gain and wing joint damping, we adjusted the MuJoCo fluid model coefficients (Table 6) which scale the drag and lift forces produced by the flapping wings. These (dimensionless) fluid model coefficients are defined in the fluid model Methods section and stored in the `fluidcoef` MuJoCo attribute. We placed the model in a flight position and again drove the wings with the real reference angles as target angles for the wing actuators. We then iteratively found a set of fluid parameters such that the net lift approximately balanced the fly model weight during several wingbeat cycles, `fluidcoef = [1.0, 0.5, 1.5, 1.7, 1.0]`. The flight physics parameters are summarized in Table 7.

Table 6: Fluid force model coefficients. See Section for definition. The coefficients are stored in the `fluidcoef` MuJoCo attribute, in this order.

Index	Coefficient	Description	Value
0	C_D^{blunt}	Blunt drag	1.0
1	C_D^{slender}	Slender drag	0.5
2	C_D^{angular}	Angular drag	1.5
3	C_K	Kutta lift	1.7
4	C_M	Magnus lift	1.0

Table 7: Flight physics parameters.

Parameter	Value	Units
`fluidcoef`	[1.0, 0.5, 1.5, 1.7, 1.0]	unitless
`gainprm`	[18, 18, 18]	dyn cm
`damping`	0.007769	dyn s
`body_pitch_angle`	47.5	deg
`stroke_plane_angle`	0	deg
`base_freq`	218	Hz

Wingbeat Pattern Generator

All our flight tasks utilize a wingbeat pattern generator (WPG) that produces, in an open-loop fashion, a fixed mirror-symmetric cyclic baseline wing trajectory. The WPG generates the baseline pattern by design with no learning involved. The baseline trajectory closely follows a previously recorded wing pattern of a hovering *Drosophila Melanogaster* [Dickson et al., 2008]. The baseline pattern is available at Supp. Data. At each simulation time-step, the WPG retrieves and outputs the six wing angles (three per wing, Fig. 2) of the baseline wing pattern for the current wing beat cycle

[2] https://github.com/willdickson/fmech

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

step. These baseline wing angles get converted to torque action commands for the wing actuators. While already producing a realistically looking wing motion, the fixed baseline wing trajectory alone is not sufficient to support a stable hovering due to the lack of feedback loop, approximations in the MuJoCo fluid model, and the sim-to-real gap. It is the role of the policy network to provide these missing components. To achieve this, the WPG torque action is combined (additively) with the policy output to produce the final action vector to be sent to the wing actuators. In this way, the policy modulates the fixed baseline pattern and produces flight required by the task at hand, e.g. stabilize flight, hover, speed up, turn, etc.

The wing motion produced by the combination of the WPG and the trained policy stays close to the initial WPG baseline pattern. This is achieved by penalizing the magnitude of the policy actions during training with the DMPO agent [Abdolmaleki et al., 2018b, Abdolmaleki et al., 2018a]. In this way, the agent is encouraged to discover a physically viable wing motion pattern by using only minimal policy actions without deviating significantly from the baseline pattern. In addition, the WPG can vary the frequency of the output baseline pattern within a pre-defined range. A single scalar out of the policy action vector is used by the WPG to control the baseline wingbeat frequency. In our setting, the wingbeat frequency was allowed to vary within 10% range centered at 218 Hz, *Drosophila melanogaster* average frequency [Fry et al., 2005]. The WPG is implemented as a lookup table containing the single fixed baseline wing pattern resampled at wingbeat frequencies within the 10% frequency range. When a frequency change is requested by the policy, the WPG will smoothly connect the patterns at old and new frequencies.

Flight imitation task configuration

As the flight reference data, we used previously recorded trajectories of freely flying *Drosophila hydei*. The trajectories contain a fly's Cartesian center-of-mass position and body orientation represented as a quaternion. The trajectories were recorded at 7500 fps. We started with 44 trajectories of spontaneous turns (saccades) [Muijres et al., 2015] and 92 trajectories of evasion maneuvers [Muijres et al., 2014] in response to visual looming stimuli. Each reference trajectory started with the fly first flying normally and then performing a maneuver. During and after the maneuver, the fly could fly straight, sideways, and backwards. The flies could also ascend and descend. We linearly interpolated the raw trajectories to the flight simulation control step of 0.2 ms. Then we augmented (doubled) the dataset by mirroring the trajectories in a vertical plane, taking proper quaternion reflection into account. This resulted in a dataset of 272 flight trajectories, equivalent to $\sim$53 seconds of real time flight. The dataset is available at Supp. Data. We used 80% of the trajectories for training and the rest for testing. Due to the small size of the dataset, to maintain balance between left and right turns in the training data, we split the dataset such that if a trajectory is in the training set, so is its mirrored counterpart. We simulated flight at 0.05 ms physics time-steps and 0.2 ms control time-steps, Table 10.

The reinforcement learning task is set as follows. In each episode, the fly model is required to track a reference trajectory selected from the flight dataset at random. The episode begins from a random step within the selected reference trajectory, excluding the last 50 steps. The model's initial position, orientation, linear and rotational velocities are set equal to the reference. The initial phase of the wing cycle is randomized. The episode ends either when the end of trajectory is successfully reached, or terminates early if the model hits the ground or is displaced from the reference CoM position by more than 2 cm.

The reward is calculated based on how closely the fly model tracks the reference trajectory at each timestep. The reward is a product of two terms measuring the quality of CoM tracking and

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

body orientation tracking. At every simulation timestep, the reward $R \in [0, 1]$ is computed as:

$$R = \underbrace{\max(0,\ 1 - \tfrac{1}{\delta_{\text{com}}}\|\mathbf{r} - \mathbf{r}^\star\|)}_{\text{CoM displacement}} \times \underbrace{\max(0,\ 1 - \tfrac{1}{\delta_{\text{quat}}}\|\log(q \circ\ q^{\star -1})\|)}_{\text{body orientation}} \tag{13}$$

The first term measures the current displacement between the model and reference CoM positions, $\mathbf{r}$ and $\mathbf{r}^\star$ respectively. The second term is the body orientation mismatch, calculated as the norm of the quaternion "minus operator" between the model and reference quaternions, q and $q^\star$, respectively [Sola, 2017]. δ_{com} and δ_{quat} are the reward tightness hyperparameters. For instance, the reward is zero when the CoM displacement in the first term $\|\mathbf{r} - \mathbf{r}^\star\| > \delta_{\text{com}}$, while it is one when $\|\mathbf{r} - \mathbf{r}^\star\| = 0$, and similarly for the second term. See Table 10 for the hyperparameter values used.

In the flight imitation task, we placed the fly legs in a retracted flight position and disabled the leg DoFs and actuators. The retracted leg configuration is stored in the `springref` leg parameters, which were fitted by matching the model's legs to images of flying *Drosophila*. We also removed the antennae and proboscis actuators and excluded their joints from the observations. This reduced the number of observable joint angles and joint velocities to 25. It also reduced the total action dimension to 12. We didn't use vision in this task. The observables (policy inputs) are listed in Table 8 and the actions (policy outputs) are shown in Table 9. In addition to the standard set of egocentric vestibular and proprioception observables, the policy receives task-specific inputs: the Cartesian CoM displacement and the orientation (quaternion) displacement of the reference trajectory with respect to the model at the current timestep plus 5 timesteps into the future. The displacements are calculated with respect to the current fly model position and orientation and are expressed in the egocentric reference frame of the model. The policy network architecture is shown in Table 11. The trained flight imitation policy can be used as a low-level flight controller. In this scenario, the two task-specific inputs (CoM and quaternion displacements) serve as high-level steering control commands.

At training time, we engage action penalization, a DMPO agent feature. It encourages the agent to learn to utilize actions of small magnitude thus keeping the wing motion close to the baseline wing pattern produced by the WPG. The task and agent hyperparameters, including DMPO action penalization, are summarized in Table 5 and Table 10.

Table 8: Observations in the flight imitation task. All observables are calculated in fly's egocentric reference frame.

Category	**Observables (policy inputs)**	**Units**	**Array shape**
vestibular	velocimeter	cm/s	3
	accelerometer	cm/s^2	3
	gyro	rad/s	3
	gravity direction	cm	3
proprioception	joint angles	rad	25
	joint velocities	rad/s	25
task inputs (steering commands)	reference CoM displacement	cm	(6, 3)
	reference orientation displacement	unitless quaternion	(6, 4)
total obs. dimension			104

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 9: Actions in the flight imitation task.

Actions (policy outputs)	**Array shape**
head	3
wings	6
abdomen	2
WPG frequency control	1
total action dimension	12

Table 10: Flight imitation task hyperparameters. See implementation for more details.

Parameter	**Value**	**Units**
`physics_timestep`	5×10^{-5}	s
`control_timestep`	2×10^{-4}	s
`joint_filter`	0	s
`adhesion_filter`	N/A	s
`future_steps`	5	
`time_limit`	0.6	s
`terminal_com_dist`	2.0	cm
reward parameters:		
δ_{com}	0.4	cm
δ_{quat}	3.14	rad

Table 11: Flight policy network architecture. Layer sizes indicated in parentheses.

input: observations	
Linear (256) LayerNorm tanh Linear (256) ELU Linear (256) ELU	
Linear (12)	Linear (12) softplus
output: action mean	*output*: action std

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Vision-guided flight task configuration

We considered two vision-guided reinforcement learning tasks, which were set as follows. The fly model is required to fly generally in the positive x-direction, to maintain a given target speed and height above terrain, and to avoid collisions with the terrain. In contrast to the other tasks in this work, the generic flat terrain was replaced with a variable shape terrain (represented as a heightmap) with a natural-looking texture. As the policy has no direct access to the flight height and terrain shape, avoiding collisions and assessing the current height requires utilization of the visual input from the two eye cameras. In each episode, the terrain shape is procedurally regenerated, and a new target speed and target height are randomly selected (Table 14). The episodes end normally when the time limit is reached, or terminate early when the fly collides with terrain, which results in the loss of all future infinite-horizon rewards.

The task is set in a square 40×40 cm arena. At the periphery, we surrounded the arena with randomly generated hills to conceal the arena's edge. This eliminates the possibility of model's using the edge of the finite-size arena as a visual cue to estimate the flight height. Also, we removed overhead light sources to eliminate the fly's shadows as the shadow size could be exploited to infer the flight height. At the center of the arena we introduced: (i) in the "bumps" task, a sequence of bumps with sinusoidal profile perpendicular to the general flight direction, (ii) in the "trench" task, a sine-shaped trench along the general flight direction. In both tasks, the amplitude, phase, period, height of the bumps and trench were randomly selected in each episode (see Table 14 for the terrain randomization details). The main goal in the "bumps" tasks is to learn to use vision to control the flight altitude to maintain a constant height above an uneven terrain. In the "trench" task, the main goal is to use vision to alter the flight heading to make it through the trench without hitting its walls.

In the "bumps" tasks, the (multiplicative) reward consists of several factors as follows:

$$R_{\text{bumps}} = R_{\text{height}} \times R_{\text{speed}} \times R_{\text{dir}} \times R_{\text{heading}} \times R_{\text{gravity}} \tag{14}$$

where the individual reward terms are:

$$\begin{aligned} R_{\text{height}} &= \max(0, 1 - \tfrac{1}{\delta_h}|h - h^\star|) \\ R_{\text{speed}} &= \max(0, 1 - \tfrac{1}{s^\star}\big||\mathbf{v}| - s^\star\big|) \\ R_{\text{dir}} &= \max(0, 1 - \tfrac{1}{s^\star}|v_x - s^\star|) \\ R_{\text{heading}} &= \max(0, 1 - \tfrac{1}{\delta_y}v_y^\dagger) \\ R_{\text{gravity}} &= \max(0, 1 - \tfrac{1}{\pi}\theta_{\text{g}}) \end{aligned} \tag{15}$$

Here, R_{height} is the reward factor measuring how close the current flight height h is to the target height $h^\star$. As in the other tasks, δ_h is the reward tightness hyperparameter (see Table 14). Similarly, R_{speed} compares the current flight speed $|\mathbf{v}|$ to the target speed $s^\star$ ($\mathbf{v}$ is the fly velocity vector). R_{dir} prescribes a preferred general flight direction by rewarding fly's propagation in the positive x-direction. v_x is the x-component of fly velocity computed in the arena reference frame. R_{heading} requires the fly to keep the body heading parallel to the current velocity direction (e.g., avoid flying sideways). This is achieved by minimizing the lateral velocity component $v_y^\dagger$ computed in fly's egocentric reference frame. R_{gravity} sets a preferred body orientation with respect to the gravity direction (the vertical z-axis) by minimizing θ_{g}, the angle between the current gravity direction vector $\mathbf{g}$ as measured by fly's gravity sensor and the preferred egocentric gravity direction $\mathbf{g}^\star$. This angle is computed as $\theta_{\text{g}} = \cos^{-1}(\mathbf{g} \cdot \mathbf{g}^\star)$. The preferred gravity direction is expressed in fly's

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

egocentric reference frame as $\mathbf{g}^\star = (\sin(\alpha_{\text{pitch}}), 0, \cos(\alpha_{\text{pitch}}))$, where $\alpha_{\text{pitch}} = 47.5^\circ$ is the default body pitch angle during stable flight in *Drosophila* [Muijres et al., 2014], see Table 7.

In the "trench" task, the reward is as in the "bumps" task with an additional term:

$$R_{\text{trench}} = R_{\text{bumps}} \times R_{\text{mid}} \tag{16}$$

The additional term R_{mid} requires the fly to stay close to the midline of the trench and it is written as:

$$R_{\text{mid}} = \begin{cases} \max(0, 1 - \frac{1}{\delta_m}|y - y^\star|) & \text{fly inside trench} \\ 1 & \text{fly before or after trench} \end{cases} \tag{17}$$

This term compares the current fly's y-coordinate (lateral, perpendicular to the general trench direction x) and the current midpoint of the trench, $y^\star$. In the "trench" task, the target height is always smaller than the trench wall height. Also, the trench width and turn amplitude are sampled such that the "trivial" solution of making it through the trench by simply flying straight wouldn't be possible.

In both vision-guided flight tasks, we used a similar flight controller architecture as in the flight imitation task, consisting of a combination of the WPG and a policy network with the same action dimension of 12, see Table 13. The fly legs were retracted and their DoFs and actuators disabled. The observations, on the other hand, were altered in two ways. First, we added visual input from the eye cameras as two 32×32 RGB frames. Second, we provided a task-input consisting of the target speed and height, which were randomly sampled in the beginning of each episode. The total observation dimension was 6208, see Table 12 for more detail. The RGB visual input was first converted to grayscale, then processed by a convolutional visual module, and then fed into the policy network as a vector of dimension 8, along with the rest of the observables. We used a generic convolutional network mostly similar to a single block of ResNet[He et al., 2015]. The policy network architecture is shown in Table 15. We used the same DMPO agent configuration, including action penalization, as in the flight imitation task, see Table 5. We trained the vision flight controller, including the convolutional visual module, end-to-end with reinforcement learning. The vision task hyperparameters are summarized in Table 14.

Table 12: Observations in the vision-guided flight task. All observables are calculated in fly's egocentric reference frame.

Category	**Observable (policy input)**	**Units**	**Array shape**
vision	right eye	RGB (byte)	(32, 32, 3)
	left eye		(32, 32, 3)
vestibular	velocimeter	cm/s	3
	accelerometer	cm/s^2	3
	gyro (rotational velocity)	rad/s	3
	gravity direction	cm	3
proprioception	joint angles	rad	25
	joint velocities	rad/s	25
task input	target flight speed	cm/s	1
	target flight height	cm	1
total observation dimension			6208

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 13: Actions in the vision-guided flight task.

Action (policy output)	**Array shape**
head	3
wings	6
abdomen	2
WPG frequency control	1
total action dimension	12

Table 14: Vision-guided flight task parameters and ranges of sine bump and trench terrain randomization.

Parameter	**Value**	**Units**
`eye_camera_fovy`	150	deg
`eye_camera_size`	32	pixels
`vis_output_dim`	8	
`time_limit`	0.4	s
reward parameters:		
δ_h	0.15	cm
$s^\star$	$1.1\times$`target_speed`	cm/s
δ_y	10	cm/s
δ_m	0.15	cm
flight speed range	[20, 40]	cm/s
flight height range	[0.5, 0.8]	cm
bump terrain parameters:		
period range	[10, 15]	cm
height range	[0.5, 1]	cm
phase range	[0, 2π]	rad
trench terrain parameters:		
period range	[5, 8]	cm
amplitude range	[0.35, 0.6]	cm
phase range	[0, 2π]	rad
width range	$2\times$ amplitude + [0.3, 0.6]	cm
height	1.3	cm
length range	[4, 10]	cm
initial fly distance from trench	[0, 2]	cm

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 15: Policy network architecture for vision-guided flight. Visual input is pre-processed by the convolutional visual module and concatenated with the rest of the observations. Layer sizes indicated in parentheses.

input: observations	
Linear (256) LayerNorm tanh Linear (256) ELU Linear (256) ELU	
Linear (12)	Linear (12) softplus
output: action mean	*output*: action std

Modeling walking behavior

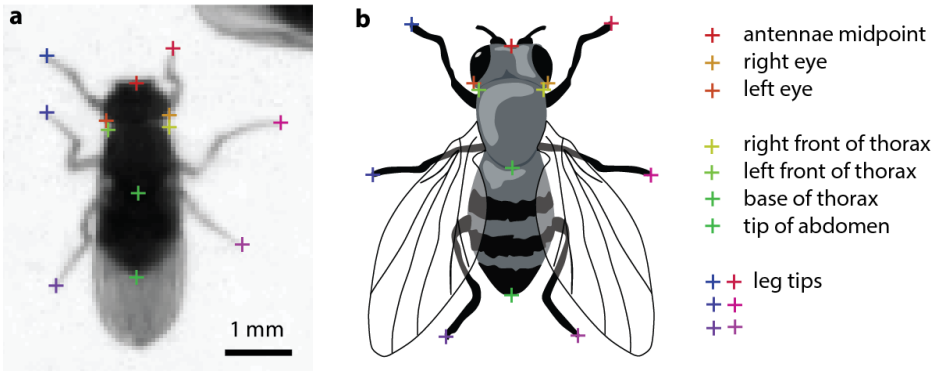
Reference walking data preparation

We obtained single-camera top-down view videos of multiple freely behaving *Drosophila* with 2D keypoint tracking from Robie et al., (Extended Data Fig. 6A). Briefly, groups of 10 walking flies were recorded at 150 fps in a shallow, flat-bottomed, 50 mm diameter arena. The 2D positions of 17 keypoints were predicated with the Animal Part Tracker (APT) [3]. From 9 such videos, we prepared the walking reference dataset. We used 13 of the 17 keypoints: three on the head, three on the thorax, one at the tip of abdomen and the 6 leg tips, as shown in Extended Data Fig. 6. We selected female flies and isolated walking trajectory segments based on the following criteria. At each frame, we required: (i) the distance to the other flies in the arena is larger than one body length, (ii) the velocity component parallel to the fly body is larger than perpendicular to the body. Then we required (iii) snippet duration of at least 20 frames (133 ms, roughly corresponds to one fly step), and (iv) ratio of mean leg tip speed to mean center-of-mass speed smaller than 1.5. This produced a set of trajectory snippets with flies walking at different speeds, turning, and standing still. The average walking speeds per snippet are distributed approximately in the range [0, 4] cm/s (Fig. 3e, inset). We then linearly interpolated the walking snippets from 6.7 ms timesteps to the walking simulation control of 2 ms timesteps.

The 13 keypoints tracked in two dimensions, however, are not sufficient for the RL task reward calculation (described below), which requires complete specification of the model's pose, position, and orientation. Specifically, this full-body represention should include all joint angles, joint axis orientations, joint velocities, body position and orientation, and leg tip positions. Obtaining the full-body data for the model from the experimental data required, first, to lift the 2D walking snippets by complementing the horizontal x, y-coordinates with the third vertical z-dimension. Based on a separate side-view video of a walking fly [Akitake et al., 2015] and on our fly model's default standing position, we approximated the body height and pitch angle during walking by a single fixed value. From this video, we also estimated the amplitudes of the arcs traversed by leg tips during swing motion. The amplitudes were $A = 0.086, 0.047, 0.051$ cm for the T1, T2, T3 legs, respectively. We approximated the z-coordinate of the leg-tip swing arcs by the sine function as $z = A\sin(x)$, with x going from 0 to π for each single leg swing. Using the 2D coordinates

[3] https://github.com/kristinbranson/APT

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



Extended Data Figure 6: **2D keypoints used to track fly walking.** (**a**) A top-down view video frame of a walking fly with keypoints inferred with Animal Part Tracker (APT). (**b**) Keypoint definition.

of the leg tip keypoints, we separated leg swings from stances based on the leg tip horizontal velocities in fly's egocentric reference frame. Then we added the approximate sine-arcs to the swing segments of the leg tip trajectories, while we kept $z = 0$ for the stances. This procedure produced 3D coordinates for the 13 keypoints in the walking snippets selected earlier.

As a final step, we computed the full-body reference poses for each frame in all the snippets. We added to the fly model equivalent 13 keypoint sites and performed inverse kinematics fitting of the whole model body to the fly poses in the 3D walking snippets. For each snippet, we rescaled the reference keypoints to match the size of the fly model. In each frame, we fit simultaneously all 13 keypoints by minimizing the following objective with respect to the model joint angles, $\mathbf{q} = (q_1, q_2, ...)$:

$$\min_{\mathbf{q}} \left[\sum_{i=1}^{13} ||\mathbf{s}_i(\mathbf{q}) - \mathbf{s}_i^\star||^2 + \lambda ||\mathbf{q} - \mathbf{q}_0||^2 \right]$$

where $\mathbf{s}_i(\mathbf{q})$ and $\mathbf{s}_i^\star$ are the 3D Cartesian coordinates of the 13 keypoints of the model and the fitting target pose, respectively. We used gradient descent to minimize this objective. To utilize the time-continuity across frames, we used the final pose fitted for the previous frame to initialize the fitting procedure for each subsequent frame. As the 13 keypoints (only 6 of which specify the leg tips) do not fully define the leg postures in space, we added a small regularization term to encourage fitting poses that are closer to the default standing pose of the fly model. The default pose is specified by a vector of model joint angles, $\mathbf{q}_0$, a vector of zeros in our case. The regularization strength is $\lambda = 1 \times 10^{-4}$ cm^2/rad^2. Having found the joint angles $\mathbf{q}$ for the reference poses in each frame, we also computed joint velocities, $d\mathbf{q}/dt$, using finite differences.

This procedure resulted in a complete full-body representation (joint angles, positions, orientations, velocities) of the reference walking trajectories. This is our reference data for the walking imitation task. In total, the walking dataset is comprised of ∼16,000 walking snippets, amounting to ∼80 minutes of fly walking behavior. The dataset is available at Supp. Data.

Walking imitation task configuration

The reinforcement learning task is set as follows. In each episode, a reference trajectory is randomly selected from the walking dataset described above. The model is required to track the CoM position and orientation of the reference fly body, as well as the detailed motion of the legs. The initial model posture and the CoM position and velocity are set equal to the reference in the first step of the selected trajectory. The episode ends when the end of the trajectory is successfully

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

reached or terminates early if the model CoM is displaced from the reference by more than 0.3 cm (one body length.)

We used a multiplicative version of the imitation reward mostly similar to [Peng et al., 2018, Merel et al., 2020b]. At every timestep, the reward measures the similarity between the current model and reference pose, position, and orientation. The reward function is formulated as a product of (not normalized) Gaussian functions, one Gaussian for each quantity compared, e.g. a joint angle, the CoM position, etc. As the Gaussians are not normalized, each of them is in the range $[0, 1]$, and so is the total multiplicative reward. There is no unique way to construct this imitation reward function and some of the terms could be redundant/overlapping. At each time step, the reward R is calculated as:

$$R = \exp\big(- E_{\text{com}} - E_{\text{quat}} - E_{\text{qvel}} - E_{\text{ee}}\big) \tag{18}$$

where the four terms expressing different aspects of the pose tracking objective are:

$$\begin{aligned}
E_{\text{com}} &= \frac{1}{2\sigma_{\text{com}}^2}||\mathbf{r}_{\text{com}} - \mathbf{r}_{\text{com}}^{\star}||^2 \quad \text{center-of-mass position} \\
E_{\text{quat}} &= \frac{1}{2\sigma_{\text{quat}}^2}\sum_{i=1}^{N_{\text{joints}}} ||\log(q_i \circ q_i^{\star -1})||^2 \quad \text{joint quaternions} \\
E_{\text{qvel}} &= \frac{1}{2\sigma_{\text{qvel}}^2}\sum_{i=1}^{N_{\text{joints}}} (v_i - v_i^{\star})^2 \quad \text{joint velocities} \\
E_{\text{ee}} &= \frac{1}{2\sigma_{\text{ee}}^2}\sum_{i=1}^{6} ||\mathbf{r}_i - \mathbf{r}_i^{\star}||^2 \quad \text{end-effector positions}
\end{aligned} \tag{19}$$

E_{com} measures the similarity between the current model and reference Cartesian CoM positions, $\mathbf{r}_{\text{com}}$ and $\mathbf{r}_{\text{com}}^{\star}$. E_{quat} measures the similarity between the model and reference joint quaternions q_i and $q_i^{\star}$ for all joints, including the root joint. Joint quaternions encode both the joint angle and the direction of the joint rotation axis for hinge joints (or body orientation for the root joint.) E_{qvel} expresses the similarity between the model and reference joint velocities, v_i and $v_i^{\star}$. E_{ee} compares the Cartesian positions of the end-effectors (six leg tips), $\mathbf{r}_i$ and $\mathbf{r}_i^{\star}$. Each one of the four error E terms has its own reward tightness (scale) hyperparameter, expressed as a standard deviation σ. See Table 18 for the hyperparameter values used.

In the walking imitation task, we set the wings in the retracted position. We removed the actuators of the wings, proboscis, antennae. We retained their DoFs in the model but excluded these DoFs from the policy observations. We didn't use vision in this task. The observation and action dimensions were 741 and 59, respectively, see Table 16 and Table 17 for detail. The policy network architecture is shown in Table 19. We attached MuJoCo adhesion actuators to leg tips to simulate fly's adhesive pads. The control semantics for the adhesion actuators is the required adhesion force, ranging between zero to one fly body weight for each leg. The rest of the actuators were position actuators receiving target joint angles as control from the policy. We didn't use any data related to adhesion during walking and didn't constrain the model as to how to use the adhesion actuators (e.g., there are no adhesion terms in the reward.) We observed, however, that the agent generally preferred to activate the adhesion while the legs were in stance to increase friction with the ground. In this task, we did not use the DMPO action penalization because the full-body imitation reward automatically constrains the policy outputs (target joint angles) to the proper ranges.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

As in the flight imitation task, in addition to the standard set of egocentric vestibular and proprioception observables, the policy receives task-specific inputs: the Cartesian CoM displacement and the orientation (quaternion) displacement of the reference trajectory with respect to the model at the current timestep plus 64 timesteps into the future. In contrast to the flight tasks, the actions output by the policy are not fed to the actuators directly but first undergo filtering (averaging) in time with time constants `joint_filter` and `adhesion_filter` for joint and adhesion actuators, respectively (see Table 18). As a result, the current actuator target joint angles (or target force in adhesion actuators) are generally different from the current policy control output and therefore considered as the actuator's internal state. This internal state is added as an additional observable in this task, see Table 16. As in the flight imitation task, the trained walking policy can be used as a low-level walking controller with the two task-specific inputs (CoM and quaternion displacements) serving as high-level steering control commands.

Table 16: Observations in the walking imitation task. All observables are calculated in fly's egocentric reference frame. The actuator activation state units are the same as the corresponding actuator control units.

Category	**Observables (policy inputs)**	**Units**	**Array shape**
vestibular	velocimeter	cm/s	3
	accelerometer	cm/s^2	3
	gyro (rotational velocity)	rad/s	3
	gravity direction	cm	3
proprioception	joint angles	rad	85
	joint velocities	rad/s	85
	actuator activation state	same as control	59
	end-effector positions	cm	(7, 3)
mechanoreception	leg force sensors	dyn	(6, 3)
	leg touch sensors	dyn	6
task input (steering commands)	reference CoM displacement	cm	(65, 3)
	reference orientation displacement	unitless quaternion	(65, 4)
total obs. dimension			741

Table 17: Actions in the walking imitation task.

Actions (policy outputs)	**Array shape**
head	3
abdomen	2
legs	(6, 8)
leg adhesion	6
total action dimension	59

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 18: Walking imitation task hyperparameters.

Parameter	Value	Units
`physics_timestep`	2×10^{-4}	s
`control_timestep`	2×10^{-3}	s
`joint_filter`	0.01	s
`adhesion_filter`	0.007	s
`claw_friction`	1.0	
`future_steps`	64	
`time_limit`	10.0	s
`terminal_com_dist`	0.3	cm
reward parameters:		
σ_{com}	0.0785	cm
σ_{qvel}	53.780	rad/s
σ_{ee}	0.0735	cm
σ_{quat}	1.225	rad

Table 19: Walking policy network architecture. Layer sizes indicated in parentheses.

input: observations	
Linear (512) LayerNorm tanh Linear (512) ELU Linear (512) ELU Linear (512) ELU	
Linear (59)	Linear (59) softplus
output: action mean	*output*: action std

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Acknowledgments

The authors thank Piotr Trochim for writing the Blender-to-MuJoCo export plug-in, Ben Moran for open-sourcing the plug-in, Nimrod Gileadi for reviewing code submissions to dm_control. Thomson Rymer, Emily Tenshaw (Janelia Project Technical Resources) and Tyler Paterson (Janelia Project Pipeline Support) for performing APT annotations. Janne Lappalainen, Gert-Jan Both, Diptodip Deb, and the members of the John Tuthill lab and the Bing Brunton lab for critical reading of the manuscript and stimulating discussions. ZS was supported by the German Research Foundation (DFG) through SPP 2041, Germany's Excellence Strategy (EXC-Number 2064/1, Project number 390727645) and the German Federal Ministry of Education and Research (BMBF; Tübingen AI Center, FKZ: 01IS18039A). ZS is a member of the International Max Planck Research School for Intelligent Systems (IMPRS-IS).

Declarations

The authors declare no competing interests.

References

[Abdolmaleki et al., 2018a] Abdolmaleki, A., Springenberg, J. T., Degrave, J., Bohez, S., Tassa, Y., Belov, D., Heess, N., and Riedmiller, M. (2018a). Relative entropy regularized policy iteration.

[Abdolmaleki et al., 2018b] Abdolmaleki, A., Springenberg, J. T., Tassa, Y., Munos, R., Heess, N., and Riedmiller, M. (2018b). Maximum a posteriori policy optimisation.

[Aimon et al., 2023] Aimon, S., Cheng, K. Y., Gjorgjieva, J., and Grunwald Kadow, I. C. (2023). Global change in brain state during spontaneous and forced walk in *Drosophila* is composed of combined activity patterns of different neuron classes. *eLife*, 12:e85202.

[Akitake et al., 2015] Akitake, B., Ren, Q., Boiko, N., Ni, J., Sokabe, T., Stockand, J. D., Eaton, B. A., and Montell, C. (2015). Coordination and fine motor control depend on drosophila trp. *Nature Communications*, 6(1):7288.

[Andersen et al., 2005] Andersen, A., Pesavento, U., and Wang, Z. J. (2005). Analysis of transitions between fluttering, tumbling and steady descent of falling cards. *Journal of Fluid Mechanics*, 541:91–104.

[Arzt et al., 2003] Arzt, E., Gorb, S., and Spolenak, R. (2003). From micro to nano contacts in biological attachment devices. *Proceedings of the National Academy of Sciences*, 100(19):10603–10606.

[Azevedo et al., 2022] Azevedo, A., Lesser, E., Mark, B., Phelps, J., Elabbady, L., Kuroda, S., Sustar, A., Moussa, A., Kandelwal, A., Dallmann, C. J., Agrawal, S., Lee, S.-Y. J., Pratt, B., Cook, A., Skutt-Kakaria, K., Gerhard, S., Lu, R., Kemnitz, N., Lee, K., Halageri, A., Castro, M., Ih, D., Gager, J., Tammam, M., Dorkenwald, S., Collman, F., Schneider-Mizell, C., Brittain, D., Jordan, C. S., Dickinson, M., Pacureanu, A., Seung, H. S., Macrina, T., Lee, W.-C. A., and Tuthill, J. C. (2022). Tools for comprehensive reconstruction and analysis of drosophila motor circuits. *bioRxiv*.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[Bellemare et al., 2017] Bellemare, M. G., Dabney, W., and Munos, R. (2017). A distributional perspective on reinforcement learning.

[Bolaños et al., 2021] Bolaños, L. A., Xiao, D., Ford, N. L., LeDue, J. M., Gupta, P. K., Doebeli, C., Hu, H., Rhodin, H., and Murphy, T. H. (2021). A three-dimensional virtual mouse generates synthetic training data for behavioral analysis. *Nature Methods*, 18(4):378–381.

[Branson et al., 2009] Branson, K., Robie, A. A., Bender, J., Perona, P., and Dickinson, M. H. (2009). High-throughput ethomics in large groups of drosophila. *Nature Methods*, 6(6):451–457.

[Card, 2012] Card, G. M. (2012). Escape behaviors in insects. *Current Opinion in Neurobiology*, 22(2):180–186.

[Cassirer et al., 2021] Cassirer, A., Barth-Maron, G., Brevdo, E., Ramos, S., Boyd, T., Sottiaux, T., and Kroiss, M. (2021). Reverb: A framework for experience replay.

[Cheong et al., 2023] Cheong, H. S. J., Eichler, K., Stuerner, T., Asinof, S. K., Champion, A. S., Marin, E. C., Oram, T. B., Sumathipala, M., Venkatasubramanian, L., Namiki, S., Siwanowicz, I., Costa, M., Berg, S., Team, J. F. P., Jefferis, G. S. X. E., and Card, G. M. (2023). Transforming descending input into behavior: The organization of premotor circuits in the drosophila male adult nerve cord connectome. *bioRxiv*.

[Chiel and Beer, 1997] Chiel, H. J. and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends in Neurosciences*, 20(12):553–557.

[Cole and Palka, 1982] Cole, E. S. and Palka, J. (1982). The pattern of campaniform sensilla on the wing and haltere of Drosophila melanogaster and several of its homeotic mutants. *Development*, 71(1):41–61.

[Community, 2018] Community, B. O. (2018). *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam.

[Cruz and Chiappe, 2023] Cruz, T. L. and Chiappe, M. E. (2023). Multilevel visuomotor control of locomotion in drosophila. *Current Opinion in Neurobiology*, 82:102774.

[DeAngelis et al., 2019] DeAngelis, B. D., Zavatone-Veth, J. A., and Clark, D. A. (2019). The manifold structure of limb coordination in walking *Drosophila*. *eLife*, 8:e46409.

[Dickinson and Palka, 1987] Dickinson, M. and Palka, J. (1987). Physiological properties, time of development, and central projection are correlated in the wing mechanoreceptors of drosophila. *Journal of Neuroscience*, 7(12):4201–4208.

[Dickinson et al., 2000] Dickinson, M. H., Farley, C. T., Full, R. J., Koehl, M. A. R., Kram, R., and Lehman, S. (2000). How animals move: An integrative view. *Science*, 288(5463):100–106.

[Dickinson and Muijres, 2016] Dickinson, M. H. and Muijres, F. T. (2016). The aerodynamics and control of free flight manoeuvres in drosophila. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1704):20150388.

[Dickson et al., 2008] Dickson, W. B., Straw, A. D., and Dickinson, M. H. (2008). Integrative model of drosophila flight. *AIAA journal*, 46(9):2150–2164.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[Dinges et al., 2021] Dinges, G. F., Chockley, A. S., Bockemühl, T., Ito, K., Blanke, A., and Büschges, A. (2021). Location and arrangement of campaniform sensilla in drosophila melanogaster. *Journal of Comparative Neurology*, 529(4):905–925.

[Dorkenwald et al., 2023] Dorkenwald, S., Matsliah, A., Sterling, A. R., Schlegel, P., chieh Yu, S., McKellar, C. E., Lin, A., Costa, M., Eichler, K., Yin, Y., Silversmith, W., Schneider-Mizell, C., Jordan, C. S., Brittain, D., Halageri, A., Kuehner, K., Ogedengbe, O., Morey, R., Gager, J., Kruk, K., Perlman, E., Yang, R., Deutsch, D., Bland, D., Sorek, M., Lu, R., Macrina, T., Lee, K., Bae, J. A., Mu, S., Nehoran, B., Mitchell, E., Popovych, S., Wu, J., Jia, Z., Castro, M., Kemnitz, N., Ih, D., Bates, A. S., Eckstein, N., Funke, J., Collman, F., Bock, D. D., Jefferis, G. S., Seung, H. S., Murthy, M., and the FlyWire Consortium (2023). Neuronal wiring diagram of an adult brain. *bioRxiv*.

[Duan et al., 2015] Duan, Z., He, B., and Duan, Y. (2015). Sphere drag and heat transfer. *Scientific reports*, 5(1):1–7.

[Fayyazuddin and Dickinson, 1996] Fayyazuddin, A. and Dickinson, M. H. (1996). Haltere afferents provide direct, electrotonic input to a steering motor neuron in the blowfly, *Calliphora*. *The Journal of Neuroscience*, 16(16):5225.

[Foelix et al., 1989] Foelix, R. F., Stocker, R. F., and Steinbrecht, R. A. (1989). Fine structure of a sensory organ in the arista of drosophila melanogaster and some other dipterans. *Cell and Tissue Research*, 258(2):277–287.

[Fry et al., 2005] Fry, S. N., Sayaman, R., and Dickinson, M. H. (2005). The aerodynamics of hovering flight in drosophila. *Journal of Experimental Biology*, 208(12):2303–2318.

[Fudalewicz-Niemczyk, 1963] Fudalewicz-Niemczyk, W. (1963). L'innervation et les organes sensoriels des ailes des dipteres et comparaison avec l'innervation des ailes d'insectes d'autres ordres. *Acta Zoologica Cracoviensia*, 8:351–462.

[Gomez-Marin and Ghazanfar, 2019] Gomez-Marin, A. and Ghazanfar, A. A. (2019). The life of behavior. *Neuron*, 104(1):25–36.

[Gorko et al., 2024] Gorko, B., Siwanowicz, I., Close, K., Christoforou, C., Hibbard, K. L., Kabra, M., Lee, A., Park, J. Y., LI, s. y., Chen, A. B., Namaki, S., Chen, C., Tuthill, J. C., Rouault, H., Branson, K., Ihrke, G., and Huston, S. J. (2024). Motor neurons generate pose-targeted movements via proprioceptive sculpting. *Nature*. accepted.

[Gosztolai et al., 2021] Gosztolai, A., Günel, S., Lobato-Ríos, V., Pietro Abrate, M., Morales, D., Rhodin, H., Fua, P., and Ramdya, P. (2021). Liftpose3d, a deep learning-based approach for transforming two-dimensional to three-dimensional poses in laboratory animals. *Nature Methods*, 18(8):975–981.

[Hasenclever et al., 2020] Hasenclever, L., Pardo, F., Hadsell, R., Heess, N., and Merel, J. (2020). Comic: complementary task learning & mimicry for reusable skills. In *Proceedings of the 37th International Conference on Machine Learning*, ICML'20. JMLR.org.

[He et al., 2015] He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[Hengstenberg, 1993] Hengstenberg, R. (1993). Multisensory control in insect oculomotor systems, in visual motion and its role in the stabilization of gaze. *Reviews of Oculomotor Research*, 5(2):285–298.

[Hoffman et al., 2020] Hoffman, M. W., Shahriari, B., Aslanides, J., Barth-Maron, G., Momchev, N., Sinopalnikov, D., Stańczyk, P., Ramos, S., Raichuk, A., Vincent, D., Hussenot, L., Dadashi, R., Dulac-Arnold, G., Orsini, M., Jacq, A., Ferret, J., Vieillard, N., Ghasemipour, S. K. S., Girgin, S., Pietquin, O., Behbahani, F., Norman, T., Abdolmaleki, A., Cassirer, A., Yang, F., Baumli, K., Henderson, S., Friesen, A., Haroun, R., Novikov, A., Colmenarejo, S. G., Cabi, S., Gulcehre, C., Paine, T. L., Srinivasan, S., Cowie, A., Wang, Z., Piot, B., and de Freitas, N. (2020). Acme: A research framework for distributed reinforcement learning. *arXiv preprint arXiv:2006.00979*.

[Horgan et al., 2018] Horgan, D., Quan, J., Budden, D., Barth-Maron, G., Hessel, M., van Hasselt, H., and Silver, D. (2018). Distributed prioritized experience replay.

[Houot et al., 2017] Houot, B., Gigot, V., Robichon, A., and Ferveur, J.-F. (2017). Free flight odor tracking in drosophila: Effect of wing chemosensors, sex and pheromonal gene regulation. *Scientific Reports*, 7(1):40221.

[Jiang and Ostadabbas, 2023] Jiang, L. and Ostadabbas, S. (2023). Spac-net: Synthetic pose-aware animal controlnet for enhanced pose estimation.

[Kamikouchi et al., 2009] Kamikouchi, A., Inagaki, H. K., Effertz, T., Hendrich, O., Fiala, A., Göpfert, M. C., and Ito, K. (2009). The neural basis of drosophila gravity-sensing and hearing. *Nature*, 458(7235):165–171.

[Karashchuk et al., 2021] Karashchuk, P., Rupp, K. L., Dickinson, E. S., Walling-Bell, S., Sanders, E., Azim, E., Brunton, B. W., and Tuthill, J. C. (2021). Anipose: A toolkit for robust markerless 3d pose estimation. *Cell Reports*, 36(13):109730.

[Kingma and Ba, 2017] Kingma, D. P. and Ba, J. (2017). Adam: A method for stochastic optimization.

[Kuan et al., 2020] Kuan, A. T., Phelps, J. S., Thomas, L. A., Nguyen, T. M., Han, J., Chen, C.-L., Azevedo, A. W., Tuthill, J. C., Funke, J., Cloetens, P., Pacureanu, A., and Lee, W.-C. A. (2020). Dense neuronal reconstruction through x-ray holographic nano-tomography. *Nature Neuroscience*, 23(12):1637–1643.

[Kutta, 1902] Kutta, W. (1902). Lift forces in fluid flow. *Illustrierte Aeronautische Mitteilungen*, page 133.

[Lamb, 1932] Lamb, H. (1932). *Hydrodynamics. Sixth edition.* Cambridge University Press.

[Lappalainen et al., 2023] Lappalainen, J. K., Tschopp, F. D., Prakhya, S., McGill, M., Nern, A., Shinomiya, K., ya Takemura, S., Gruntman, E., Macke, J. H., and Turaga, S. C. (2023). Connectome-constrained deep mechanistic networks predict neural responses across the fly visual system at single-neuron resolution. *bioRxiv*.

[Lesser et al., 2023] Lesser, E., Azevedo, A. W., Phelps, J. S., Elabbady, L., Cook, A., Mark, B., Kuroda, S., Sustar, A., Moussa, A., Dallmann, C. J., Agrawal, S., Lee, S.-Y. J., Pratt, B., Skutt-Kakaria, K., Gerhard, S., Lu, R., Kemnitz, N., Lee, K., Halageri, A., Castro, M., Ih, D., Gager, J., Tammam, M., Dorkenwald, S., Collman, F., Schneider-Mizell, C., Brittain, D., Jordan, C. S.,

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Seung, H. S., Macrina, T., Dickinson, M., Lee, W.-C. A., and Tuthill, J. C. (2023). Synaptic architecture of leg and wing motor control networks in drosophila. *bioRxiv*.

[Lobato-Rios et al., 2022] Lobato-Rios, V., Ramalingasetty, S. T., Özdil, P. G., Arreguit, J., Ijspeert, A. J., and Ramdya, P. (2022). Neuromechfly, a neuromechanical model of adult drosophila melanogaster. *Nature Methods*, 19(5):620–627.

[Mamiya and Dickinson, 2015] Mamiya, A. and Dickinson, M. H. (2015). Antennal mechanosensory neurons mediate wing motor reflexes in flying drosophila. *The Journal of Neuroscience*, 35(20):7977.

[Mamiya et al., 2023] Mamiya, A., Sustar, A., Siwanowicz, I., Qi, Y., Lu, T.-C., Gurung, P., Chen, C., Phelps, J. S., Kuan, A. T., Pacureanu, A., Lee, W.-C. A., Li, H., Mhatre, N., and Tuthill, J. C. (2023). Biomechanical origins of proprioceptor feature selectivity and topographic maps in the drosophila leg. *Neuron*.

[Marin et al., 2023] Marin, E. C., Morris, B. J., Stuerner, T., Champion, A. S., Krzeminski, D., Badalamente, G., Gkantia, M., Dunne, C. R., Eichler, K., ya Takemura, S., Tamimi, I. F. M., Fang, S., Moon, S. S., Cheong, H. S. J., Li, F., Schlegel, P., Berg, S., Team, F. P., Card, G. M., Costa, M., Shepherd, D., and Jefferis, G. S. (2023). Systematic annotation of a complete adult male drosophila nerve cord connectome reveals principles of functional organisation. *bioRxiv*.

[McKellar et al., 2020] McKellar, C. E., Siwanowicz, I., Dickson, B. J., and Simpson, J. H. (2020). Controlling motor neurons of every muscle for fly proboscis reaching. *eLife*, 9:e54978.

[Melis et al., 2024] Melis, J. M., Siwanowicz, I., and Dickinson, M. H. (2024). Machine learning reveals the control mechanics of an insect wing hinge. *bioRxiv*.

[Merel et al., 2019] Merel, J., Botvinick, M., and Wayne, G. (2019). Hierarchical motor control in mammals and machines. *Nature Communications*, 10(1):5489.

[Merel et al., 2020a] Merel, J., Tunyasuvunakool, S., Ahuja, A., Tassa, Y., Hasenclever, L., Pham, V., Erez, T., Wayne, G., and Heess, N. (2020a). Catch & carry: reusable neural controllers for vision-guided whole-body tasks. *ACM Transactions on Graphics (TOG)*, 39(4):39–1.

[Merel et al., 2020b] Merel, J., Tunyasuvunakool, S., Ahuja, A., Tassa, Y., Hasenclever, L., Pham, V., Erez, T., Wayne, G., and Heess, N. (2020b). Catch & carry: Reusable neural controllers for vision-guided whole-body tasks. *ACM Trans. Graph.*, 39(4).

[Merritt and Murphey, 1992] Merritt, D. J. and Murphey, R. K. (1992). Projections of leg proprioceptors within the cns of the fly phormia in relation to the generalized insect ganglion. *Journal of Comparative Neurology*, 322(1):16–34.

[Mi et al., 2022] Mi, L., Xu, R., Prakhya, S., Lin, A., Shavit, N., Samuel, A., and Turaga, S. C. (2022). Connectome-constrained latent variable model of whole-brain neural activity. In *International Conference on Learning Representations*.

[Mimura et al., 1970] Mimura, K., Tateda, H., Morita, H., and Kuwabara, M. (1970). Convergence of antennal and ocellar inputs in the insect brain. *Zeitschrift für vergleichende Physiologie*, 68(3):301–310.

[Montell, 2009] Montell, C. (2009). A taste of the drosophila gustatory receptors. *Current Opinion in Neurobiology*, 19(4):345–353. Sensory systems.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[Moritz et al., 2017] Moritz, P., Nishihara, R., Wang, S., Tumanov, A., Liaw, R., Liang, E., Paul, W., Jordan, M. I., and Stoica, I. (2017). Ray: A distributed framework for emerging AI applications. *CoRR*, abs/1712.05889.

[Muijres et al., 2015] Muijres, F. T., Elzinga, M. J., Iwasaki, N. A., and Dickinson, M. H. (2015). Body saccades of Drosophila consist of stereotyped banked turns. *Journal of Experimental Biology*, 218(6):864–875.

[Muijres et al., 2014] Muijres, F. T., Elzinga, M. J., Melis, J. M., and Dickinson, M. H. (2014). Flies evade looming targets by executing rapid visually directed banked turns. *Science*, 344(6180):172–177.

[Murphey et al., 1989] Murphey, R. K., Possidente, D., Pollack, G., and Merritt, D. J. (1989). Modality-specific axonal projections in the cns of the flies phormia and drosophila. *Journal of Comparative Neurology*, 290(2):185–200.

[Nair et al., 2015] Nair, A., Srinivasan, P., Blackwell, S., Alcicek, C., Fearon, R., De Maria, A., Panneershelvam, V., Suleyman, M., Beattie, C., Petersen, S., Legg, S., Mnih, V., Kavukcuoglu, K., and Silver, D. (2015). Massively parallel methods for deep reinforcement learning.

[Pacheco et al., 2021] Pacheco, D. A., Thiberge, S. Y., Pnevmatikakis, E., and Murthy, M. (2021). Auditory activity is diverse and widespread throughout the central brain of drosophila. *Nature Neuroscience*, 24(1):93–104.

[Peng et al., 2018] Peng, X. B., Abbeel, P., Levine, S., and Van de Panne, M. (2018). Deepmimic: Example-guided deep reinforcement learning of physics-based character skills. *ACM Transactions On Graphics (TOG)*, 37(4):1–14.

[Pereira et al., 2022] Pereira, T. D., Tabris, N., Matsliah, A., Turner, D. M., Li, J., Ravindranath, S., Papadoyannis, E. S., Normand, E., Deutsch, D. S., Wang, Z. Y., McKenzie-Smith, G. C., Mitelut, C. C., Castro, M. D., D'Uva, J., Kislin, M., Sanes, D. H., Kocher, S. D., Wang, S. S. H., Falkner, A. L., Shaevitz, J. W., and Murthy, M. (2022). Sleap: A deep learning system for multi-animal pose tracking. *Nature Methods*, 19(4):486–495.

[Plum et al., 2023] Plum, F., Bulla, R., Beck, H., Imirzian, N., and Labonte, D. (2023). replicant - generating annotated images of animals in complex environments with unreal engine. *bioRxiv*.

[Reiser et al., 2004] Reiser, M., Humbert, J., Dunlop, M., Del Vecchio, D., Murray, R., and Dickinson, M. (2004). Vision as a compensatory mechanism for disturbance rejection in upwind flight. In *Proceedings of the 2004 American Control Conference*, volume 1, pages 311–316 vol.1.

[Sane, 2003] Sane, S. P. (2003). The aerodynamics of insect flight. *Journal of Experimental Biology*, 206(23):4191–4208.

[Schindelin et al., 2012] Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682.

[Schlegel et al., 2023] Schlegel, P., Yin, Y., Bates, A. S., Dorkenwald, S., Eichler, K., Brooks, P., Han, D. S., Gkantia, M., dos Santos, M., Munnelly, E. J., Badalamente, G., Capdevila, L. S., Sane, V. A., Pleijzier, M. W., Tamimi, I. F., Dunne, C. R., Salgarella, I., Javier, A., Fang, S.,

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Perlman, E., Kazimiers, T., Jagannathan, S. R., Matsliah, A., Sterling, A. R., chieh Yu, S., McKellar, C. E., Consortium, F., Costa, M., Seung, H. S., Murthy, M., Hartenstein, V., Bock, D. D., and Jefferis, G. S. (2023). Whole-brain annotation and multi-connectome cell typing quantifies circuit stereotypy in drosophila. *bioRxiv*.

[Seifert, 2012] Seifert, J. (2012). A review of the magnus effect in aeronautics. *Progress in Aerospace Sciences*, 55:17–45.

[Sola, 2017] Sola, J. (2017). Quaternion kinematics for the error-state kalman filter. *arXiv preprint arXiv:1711.02508*.

[Stokes, 1850] Stokes, G. (1850). On the effect of internal friction of fluids on the motion of pendulums. *Trans. Camb. phi1. S0c*, 9(8):106.

[Strausfeld and Seyan, 1985] Strausfeld, N. J. and Seyan, H. S. (1985). Convergence of visual, haltere, and prosternai inputs at neck motor neurons of calliphora erythrocephala. *Cell and Tissue Research*, 240(3):601–615.

[Strausfeld et al., 1987] Strausfeld, N. J., Seyan, H. S., and Milde, J. J. (1987). The neck motor system of the fly calliphora erythrocephala. *Journal of Comparative Physiology A*, 160(2):205–224.

[Straw et al., 2010] Straw, A. D., Lee, S., and Dickinson, M. H. (2010). Visual control of altitude in flying drosophila. *Current Biology*, 20(17):1550–1556.

[Sun et al., 2023] Sun, J. J., Karashchuk, L., Dravid, A., Ryou, S., Fereidooni, S., Tuthill, J. C., Katsaggelos, A., Brunton, B. W., Gkioxari, G., Kennedy, A., Yue, Y., and Perona, P. (2023). Bkind-3d: Self-supervised 3d keypoint discovery from multi-view videos. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9001–9010.

[Takemura et al., 2023] Takemura, S., Hayworth, K. J., Huang, G. B., Januszewski, M., Lu, Z., Marin, E. C., Preibisch, S., Xu, C. S., Bogovic, J., Champion, A. S., Cheong, H. S., Costa, M., Eichler, K., Katz, W., Knecht, C., Li, F., Morris, B. J., Ordish, C., Rivlin, P. K., Schlegel, P., Shinomiya, K., Stürner, T., Zhao, T., Badalamente, G., Bailey, D., Brooks, P., Canino, B. S., Clements, J., Cook, M., Duclos, O., Dunne, C. R., Fairbanks, K., Fang, S., Finley-May, S., Francis, A., George, R., Gkantia, M., Harrington, K., Hopkins, G. P., Hsu, J., Hubbard, P. M., Javier, A., Kainmueller, D., Korff, W., Kovalyak, J., Krzemiński, D., Lauchie, S. A., Lohff, A., Maldonado, C., Manley, E. A., Mooney, C., Neace, E., Nichols, M., Ogundeyi, O., Okeoma, N., Paterson, T., Phillips, E., Phillips, E. M., Ribeiro, C., Ryan, S. M., Rymer, J. T., Scott, A. K., Scott, A. L., Shepherd, D., Shinomiya, A., Smith, C., Smith, N., Suleiman, A., Takemura, S., Talebi, I., Tamimi, I. F., Trautman, E. T., Umayam, L., Walsh, J. J., Yang, T., Rubin, G. M., Scheffer, L. K., Funke, J., Saalfeld, S., Hess, H. F., Plaza, S. M., Card, G. M., Jefferis, G. S., and Berg, S. (2023). A connectome of the male drosophila ventral nerve cord. *bioRxiv*.

[Todorov et al., 2012] Todorov, E., Erez, T., and Tassa, Y. (2012). Mujoco: A physics engine for model-based control. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5026–5033.

[Tuckerman, 1925] Tuckerman, L. (1925). *Inertia factors of ellipsoids for use in airship design*. US Government Printing Office.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

[Tunyasuvunakool et al., 2020] Tunyasuvunakool, S., Muldal, A., Doron, Y., Liu, S., Bohez, S., Merel, J., Erez, T., Lillicrap, T., Heess, N., and Tassa, Y. (2020). dm_control: software and tasks for continuous control. *Software Impacts*, 6:100022.

[Turner et al., 2022] Turner, M. H., Krieger, A., Pang, M. M., and Clandinin, T. R. (2022). Visual and motor signatures of locomotion dynamically shape a population code for feature detection in *Drosophila*. *eLife*, 11:e82587.

[Wang-Chen et al., 2023] Wang-Chen, S., Stimpfling, V. A., Özdil, P. G., Genoud, L., Hurtak, F., and Ramdya, P. (2023). Neuromechfly 2.0, a framework for simulating embodied sensorimotor control in adult drosophila. *bioRxiv*.

[Williamson et al., 2018] Williamson, W. R., Peek, M. Y., Breads, P., Coop, B., and Card, G. M. (2018). Tools for rapid high-resolution behavioral phenotyping of automatically isolated drosophila. *Cell Reports*, 25(6):1636–1649.e5.

[Zador et al., 2023] Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., Botvinick, M., Chklovskii, D., Churchland, A., Clopath, C., DiCarlo, J., Ganguli, S., Hawkins, J., Körding, K., Koulakov, A., LeCun, Y., Lillicrap, T., Marblestone, A., Olshausen, B., Pouget, A., Savin, C., Sejnowski, T., Simoncelli, E., Solla, S., Sussillo, D., Tolias, A. S., and Tsao, D. (2023). Catalyzing next-generation artificial intelligence through neuroai. *Nature Communications*, 14(1):1597.

[Zhao et al., 2022] Zhao, A., Gruntman, E., Nern, A., Iyer, N. A., Rogers, E. M., Koskela, S., Siwanowicz, I., Dreher, M., Flynn, M. A., Laughland, C. W., Ludwig, H. D., Thomson, A. G., Moran, C. P., Gezahegn, B., Bock, D. D., and Reiser, M. B. (2022). Eye structure shapes neuron function in drosophila motion vision. *bioRxiv*.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Supplement

A Constructing the MuJoCo physics model

Here we describe the steps taken in order to create the physically-simulatable MuJoCo fly model, given the geometrical Blender model and measured masses. The resulting MuJoCo model is available at https://github.com/TuragaLab/flybody.

Initial conversion The Blender-to-MuJoCo export plug-in, available at https://github.com/google-deepmind/dm_control, was used to export a raw MuJoCo model containing only geometrical information: body meshes and a kinematic tree with joint axes and limits.

Model building script The raw MuJoCo model was then loaded and manipulated with a Python script using PyMJCF, a Python library for model manipulation which is part of Google DeepMind's dm_control suite [Tunyasuvunakool et al., 2020]. The following steps were taken.

1. Enforced consistent naming everywhere using a `part_sternum_side` convention e.g., "`coxa_T3_right`". Consistent naming allows for conveniently readable loops like

```
thorax = mjcf_model.find('body', 'thorax')
legs = [body for body in thorax.find_all('body') if 'coxa' in body.name]
links = ['coxa', 'femur', 'tibia', 'tarsus', 'tarsus2', 'tarsus3', 'tarsus4']
sternums = ['T1', 'T2', 'T3']
sides = ['right', 'left']
for sternum in sternums:
  for side in sides:
    for leg in legs:
      if sternum in leg.name and side in leg.name:
        for link in links:
          # Do something with each link in each leg.
```

2. Made use of MuJoCo's cascading defaults mechanism, avoiding repeated values. Common properties like angle ranges, damping and actuator properties for all joints of the same type or masses of links of the same type are given by a single number in the model, defined in a default class, which is then inherited by elements in the kinematic tree.
3. Model units were chosen to be CGS, for better numerical precision. Note that in MKS, some values are extremely small, for example the inertia of a tarsus link is on the order of 1×10^{-20} kg m^2, while in CGS is 1×10^{-13} g cm^2. The difference in accuracy is significant for double-precision floating-point arithmetic.
4. Body moments-of-inertia were computed by MuJoCo given mesh geometries and empirical masses, assuming uniform density for each body part.
5. Kinematic symmetry was enforced to numerical precision, transverse of the sagittal plane.
6. Joint axis orientations were reflected through the saggital plane, ensuring identical semantics on both sides i.e., joint rotation in the positive (negative) direction always corresponds to extension (flexion) and abduction (adduction), respectively.
7. Joint angle reference values were chosen so that 0.0 corresponds to the base pose, see Fig. 1.
8. A layer of primitive collision geoms was created, initially by letting MuJoCo fit primitives to meshes by matching ineritas, and then by manual fine-tuning.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

9. Manually excluded pairs of bodies that cannot collide, both to avoid spurious collisions and to increase simulation speed. Note that certain collisions that are possible in real flies were also excluded. In particular since the modeled wings are rigid and while real wings are flexible, wing-wing and wing-body collisions cannot be well modeled.
10. Added tendons to the abdomen and the tarsi. These "fixed" tendons are a simplification of spatially routed tendons, whose length corresponds to a linear combinations of joint angles, allowing a single actuator to act on multiple joints e.g., to flex the multiple links of the tarsus or abduct the entire abdomen. See examples in Fig. 1.
11. Set the default body pitch angle to $47.5°$, following [Muijres et al., 2014]. Re-orient the wing joint axes such that when the body is at $47.5°$, the wing yaw axes are strictly vertical and the stroke plane is horizontal.
12. Added a total of 78 actuators:
    - 8 actuators in each leg (coxa: 3, femur: 2, tibia: 1, tarsus: 2), using desired angle (position) semantics, with gains chosen so that a force of approximately one body weight can be applied at the end-effector at the base pose.
    - Wings actuators (yaw, roll, pitch) have torque semantics with gain values of $18.0\ \mathrm{dyn\ cm}$, see Methods.
    - 16 additional actuators for proximal joints: head and rostrum (4), haustelli (2), labri (2), antennae (6) and abdomen (2).
    - 6 adhesion actuators at the claws which can apply a force up to $1\times$ body weight.
    - 2 adhesion actuators in the mouth (labrum).
13. Added egocentric sensors (also see Table 20):
    - An ideal accelerometer and gyro to the thorax, corresponding to processed vestibular sensor information.
    - An ideal velocimeter, corresponding to processed information from air motion sensors in the hair follicles.
    - Added force and touch sensors at the end effectors. The former report the force passing through the first tarsus joints, while the latter report pressure applied to the claw.
    - Two eye cameras, at the geometric center of the eyes. These are standard openGL cameras with a wide $140°$ field-of-view, see main text for how these were used in visually guided experiments.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

B Sensory system components

Table 20: Correspondence between the sensory system components in *Drosophila* and in our fly model.

Drosophila sense organ, receptor	**Sensory modality**	**Purpose, detection, sensing**	**Corresponding fly model sensor**	**References**
Compound eyes	Light, photoreception	Self-movement (optic flow, optomotor response), looming stimuli	Eye cameras, velocimeter, gravity direction	[Hengstenberg, 1993, Strausfeld and Seyan, 1985]
Ocelli	Light, photoreception	Dorsal light response, body orientation, postural correction	ditto	[Hengstenberg, 1993, Strausfeld and Seyan, 1985, Mimura et al., 1970]
Halteres (modified 2nd wing pair)	mechanoreception	Coriolis forces during flight; balance/equilibrium organ	Gyro, accelerometer	[Strausfeld and Seyan, 1985, Fayyazuddin and Dickinson, 1996]
Leg campaniform sensilla	mechanoreception	Load/gravity	Force, touch, gravity direction	[Dinges et al., 2021]
Johnston's organ (2nd antennal segment)	Mechanoreception (exteroception)	Gravity, air flow, sound (courtship song)	Gravity direction, velocimiter, sound not implemented	[Kamikouchi et al., 2009, Mimura et al., 1970, Mamiya and Dickinson, 2015]
Neck proprioceptors (prosternal organ, CO)	Mechanoreception (proprioception)	head position; postural correction	Joint angles, joint velocities	[Strausfeld and Seyan, 1985]
Wing base tegula	Mechanoreception via a hair plate and campaniform sensilla	Wing position (proprioception)	ditto	[Fudalewicz-Niemczyk, 1963]
Femoral chordotonal organ (FeCO)	mechanoreception	Flexion, extension of the tibia (proprioception); vibration (exteroception)	ditto	[Mamiya et al., 2023]
Coxa and trochanter hair plates	mechanoreception	Registering extreme positions of the legs (proprioception)	ditto	[Murphey et al., 1989, Merritt and Murphey, 1992]
Wing base campaniform sensilla	mechanoreception	Wing twist, wing load	*Not implemented*	[Cole and Palka, 1982, Dickinson and Palka, 1987]
Wing veins campaniform sensilla	mechanoreception	Aeroelastic deformations of wing	*Not implemented*	[Cole and Palka, 1982, Dickinson and Palka, 1987]
Wing margin chemosensory bristles	chemoreception	Airborne molecules (odor plumes, pheromones)	*Not implemented*	[Houot et al., 2017]
Arista (4th-6th antennal segments)	Hygro-, chemo-, thermoreception	Odor, pheromone plumes, temperature gradients	*Not implemented*	[Foelix et al., 1989]
Labrum and tarsal gustatory receptors	Hygro- and chemoreception	Chemical compounds in the substrate	*Not implemented*	[Montell, 2009]

bioRxiv preprint doi: https://doi.org/10.1101/2024.03.11.584515; this version posted March 14, 2024. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

C Simulation speed

The timing of different components within one control timestep of the simulation is shown in Table 21. We simulated the different fly model behaviors on a single core of Intel Xeon CPU E5-2697 v3 @ 2.60GHz. The total simulation control step time is broken down into its components as *Total step time* = *Policy* + *RL env* + *MuJoCo control*, where *Policy* is the forward pass through the policy network, *RL env* is the dm_control python RL environment, and *MuJoCo control* is the total of MuJoCo physics timesteps per one RL environment control timestep [Tunyasuvunakool et al., 2020]. The policy forward pass was run on CPU, same as in actors during training. % real time is shown for both the total simulation timestep and for the MuJoCo component alone. The averages are calculated from 10,000 simulation steps with std-to-mean ratios being on the order of $\sim$0.01. For vision-guided flight, "bumps" task times are shown; "bumps" and "trench" tasks perform similarly; visual input was rendered on an NVIDIA Titan Xp GPU.

Table 21: Break down of average times of one control timestep for different fly model behaviors simulated on a single core of Intel Xeon CPU E5-2697 v3 @ 2.60GHz. All times are in units of ms.

Component:	Policy	RL env	MuJoCo		Total	Real time	% real time	
Behavior:			control	physics	step time	simulated	total	MuJoCo
Flight	3.64	3.67	1.11	0.278	8.43	0.2	2.37%	18.0%
Walking	4.31	4.49	4.64	0.464	13.73	2.0	14.6%	43.1%
Vision-guided flight	4.51	9.83	1.03	0.257	15.36	0.2	1.30%	19.5%

In addition, note that the simulation speed can be improved further:

- When not flying, a much larger time-step can be used.
- When flying, contacts can (sometimes) be ignored.
- MuJoCo simulations can be easily duplicated across threads.
- As of MuJoCo 3.0, simulation is supported on GPU and TPU accelerators which can achieve much higher throughput, especially for RL training, as both the physics and network are on the same processor. We have not yet attempted to run our environments on accelerators as this would require rewriting the environment logic (standard Python is bound to CPU).